A Gaussian Mixture Model with Gaussian Weight Learning Rate and Foreground Detection using Neighbourhood Correlation

Deepak Kumar Panda
Dept. of Electronics and Communication Engg.
National Institute of Technology Rourkela
Rourkela 769008, India
Email: deepakkumar.panda@gmail.com

Sukadev Meher
Dept. of Electronics and Communication Engg.
National Institute of Technology Rourkela
Rourkela 769008, India
Email: sukadevmeher@gmail.com

***Abstract*—Moving object detection is the first and foremost step in many computer vision applications such as automated visual surveillance, human-machine interface, tracking, traffic surveillance, etc. Background subtraction is widely used for classifying image pixels into either foreground or background in presence of stationary cameras. A Gaussian Mixture Model (GMM) model is one such popular method used for background subtraction due to a good compromise between robustness to various practical environments and real-time constraints. In this paper we assume background pixel follows Gaussian distribution spatially as well as temporally. The proposed research uses Gaussian weight learning rate over a neighbourhood to update the parameters of GMM. The background pixel can be dynamic especially in outdoor environment, so in this paper we have exploited neighborhood correlation of pixels in foreground detection. We compare our method with other state-of-the-art modeling techniques and report experimental results. The performance of the proposed algorithm is evaluated using both qualitative and quantitative measures. Quantitative accuracy measurement is obtained from PCC. Experimental results are demonstrated on publicly available videos sequences containing complex dynamic backgrounds. The proposed method is quiet effective enough to provide accurate silhouette of the moving object for real-time surveillance.**

I. Introduction

Visual surveillance uses video cameras to monitor the activities of targets (humans, vehicles, etc.) in a scene. In order to classify, track or analyze activities of interested objects, it is necessary to extract the foreground object from the scene. In general, moving object detection is the first step in many computer vision applications, making it a important part of the system. Background subtraction is a very popular method for real-time segmentation of moving objects in image sequences taken from static cameras. It finds application in many computer vision task including automated visual surveillance, human-machine interface, tracking, traffic surveillance, etc. Background subtraction involves modeling the background from the current frame, subtracting each new frame from this model and thresholding the resulting difference. This results in a binary segmented image, which highlights the regions of non-stationary objects. The threshold is wisely chosen so as to minimize the number of false negative and false positives. If a small value of threshold is chosen, then a lot of irrelevant pixels are detected as foreground and it results in false positives. If a large value is selected then there is a rise in false negatives. In brief, the background subtraction method can be outlined as background initialization, background modelling, background maintenance and finally foreground detection. Background subtraction methods have to deal with practical problems, such as fast changing illumination variation, relocation of background objects, shadows, initialization with moving objects, and complex dynamic backgrounds such as swaying of trees, ripples in water, etc. Even when the background is static, camera jitters and signal noise present in the image makes it difficult for video object segmentation. In addition, to this implementation of algorithm in real-time is a critical problem.

The paper is organized as follows. In Section II, we present an overview of the existing approaches adopted for background subtraction. Section III gives a detailed description of GMM method for background modeling. Our proposed method is well illustrated in Section IV. In Section V, we present results obtained with the implementation of the proposed approach in terms of visual as well as quantitative measures, in comparison with those obtained by some of the other state of the art techniques. Section VI includes conclusions and future research directions.

II. Related Work

In the literature, a large amount of work has been done on background subtraction technique. The survey papers on background subtraction technique can be found in [1]. A simple technique for background subtraction is to subtract each new video frame from the first frame of the video and then threshold the result. $\Sigma - \Delta$ (SDE) [2] models background by incrementing the value by one if background pixel is smaller than image pixel, or decrementing by one if background is greater than the current image pixel. The background estimated by this method is an approximation of the median of I_t. SDE is not able to model background correctly for fast changing background scene. The W^4 system [3] proposed by Haritaoglu et al. models the background scene by intensity

of minimum, maximum , and the maximum difference between consecutive frames calculated from the training stage. However, the background model fails in the presence of unstationary background pixels, as algorithm relies on median of maximum absolute difference between pixels of consecutive frames for thresholding. Wren et al. [4] modeled background by a single Gaussian distribution. It works well in indoor environment, can deal with small or gradual changes in the background and illumination variation. It fails in the outdoor scene, when the background scene involves multi-modal distributions. To overcome the problem of multi-modal background, Stauffer and Grimson [5], [6] modeled each pixel intensity by a mixture of K adaptive Gaussian distributions. Numerous improvements to the original GMM, have been proposed and a good survey of the related field and an classification of these improvements can be found in [7]. Kim et al. [8] proposed codebook model, from history of observation sequences. It assigns each pixel into a set of codewords based on color difference and brightness bound. All pixels in the image will not have the same number of codewords. Kaewtrakulpong and Bowden [9] proposed updating formulations of the Gaussian mixture to improve the slow learning at the beginning of the updating process and switching to recursive filter learning after sufficient samples were observed. They improved GMM model with shadow detection. Lee [10] presented a new adaptive learning rate of the background model to improve the updating convergence rate without compromising model stability. Lin et al. [11] proposed GMM with different learning rate for pixel of background, shadow, static foreground and moving foreground, respectively, to help the trade-off between robustness to background changes and sensitivity to foreground abnormalities.

III. Mixture of Gaussian Model

Staufer and Grimson [5], [6] have proposed an adaptive parametric GMM to lessen the effect of small repetitive motions like trees and bushes as well as illumination variation. A pixel I at position x and time t is modeled as a mixture of K Gaussian distributions. The current pixel value follows the probability distribution given by

$$P(I_{t,x}) = \sum_{i=1}^{K} w_{t-1,x,i} * \eta(I_{t,x}, \mu_{t-1,x,i}, \sigma^2_{t-1,x,i}) \quad (1)$$

where η is the Gaussian probability density function

$$\eta(I;\, \mu,\, \sigma^2) \;=\; \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(I-\mu)^2}{2\sigma^2}\right) \quad (2)$$

and $w_{t-1,x,i}$, $\mu_{t-1,x,i}$, and $\sigma^2_{t-1,x,i}$ are the weight, mean value and variance of the i_{th} Gaussian in the mixture at time $t-1$. For maintaining the Gaussian mixture model, the parameters $\mu_{t-1,x,i}$, $\sigma^2_{t-1,x,i}$ and $w_{t-1,x,i}$ needs to be updated based on the new pixel $I_{t,x}$. The parameter initialization of the weight, the mean and the covariance matrix is done using an k-means algorithm for practical consideration. The RGB color components are assumed to be independent for computational reasons. So, the covariance matrix is given by:

$$\Sigma_{t,x,i} = \sigma^2_{t,x,i} I \quad (3)$$

If the red, green and blue color channels have identical variances then a single scalar $\sigma^2_{t,i}$ can be assumed.

A pixel is said to be matched, if $I_{t,x}$ lies within T_σ standard deviations of a Gaussian. In our case, T_σ lies between 1 and 5.

$$|I_{t,x} \;-\; \mu_{t-1,\,x,\,i}| \;\leq\; T_\sigma \;\sigma_{t-1,\,x,\,i} \quad (4)$$

If one of the K Gaussian is matched, the matched Gaussian is updated as follows:

$$\mu_{t,x,i} = (1-\rho)\mu_{t-1,x,i} + \rho\,(I_{t,x}) \quad (5)$$

$$\sigma^2_{t,x,i} = (1-\rho)\sigma^2_{t-1,x,i} + \rho\,(I_{t,x} - \mu_{t,x,i})^T\,(I_{t,x} - \mu_{t,x,i}) \quad (6)$$

where $\rho = \alpha\eta\,(I_{t,x} \mid \mu_{t-1,x,i}, \sigma_{t-1,x,i})$ is a learning rate that controls how fast μ and σ^2 converges to new observations.

The weight of the K Gaussian is adjusted as follows:

$$w_{t,x,i} = (1-\alpha)w_{t-1,x,i} + \alpha\,(M_{t,i}) \quad (7)$$

where $M_{t,i} = 1$ is set for the matched Gaussian and $M_{t,i} = 0$ for the others. The learning rate α is used to update the weight and its value ranges between 0 and 1.

If none of the K Gaussian component matches the current pixel value, the least weighted component is replaced by a distribution with the current value as its mean, a high variance, and a low value of weight parameter is chosen.

$$k = \arg\min_{i=1,\ldots,K} w_{t-1,x,i} \quad (8)$$

$$\begin{aligned} \mu_{t,x,k} &= I_{t,x}, \\ \sigma^2_{t,x,k} &= \sigma^2_0, \\ w_{t,x,k} &= w_0, \end{aligned} \quad (9)$$

Thereafter, the weights are normalized, such that $\sum_{i=1}^{K} w_{t,x,i} = 1$.

$$w_{t,x,i} = w_{t,x,i} / \sum_{i=1}^{K} w_{t,x,i}$$

The K distributions are sorted in descending order by w/σ. This ordering moves the most probable background with high weight and low variance at the top. The first B Gaussian distribution which exceed certain threshold T are retained for the background distributions. If a small value of T is chosen, the background model is uni-modal and is multi-modal, if higher value of T is chosen.

$$B \;=\; \arg\min_{b}\left(\sum_{i=1}^{b} w_i \;>\; T\right) \quad (10)$$

If a pixel $I_{t,x}$ does not matches with any one of the background component, then the pixel is marked as foreground. The noise in the foreground binary mask is removed through proper connected component labeling.

However, GMM suffers from slow learning [10], if a small value of learning rate is chosen, especially in busy and fast changing environments. In addition, a moving shadow region may be wrongly classified as foreground. The problem aggravates further in presence of a quick illumination change such as light switched on/off and can increase the number of falsely detected foreground pixels. Consequently this may give erroneous results in tracking and classification. If the position of the background object is changed or any new object is brought into the background. The pixel value of the new background will not match with the estimated K Gaussian and will be classified as foreground. These small changes are of little interest and should be made part of the background. Otherwise a large percentage of image will be classified as foreground.

IV. Proposed Method

GMM based background subtraction is a very popular and powerful technique for moving object detection due to its good compromise between robustness to various practical environments and real-time constraints. In our proposed approach we assume background pixel follows Gaussian distribution temporally and spatially. The learning rate for parameter updation is based on Gaussian weight learning rate over a spatial neighbourhood. The background pixel can be dynamic, so we have exploited neighborhood correlation in our proposed GMM scheme for foreground detection. The proposed technique is well illustrated in Algorithm 1 and 2.

A. Learning rate using Gaussian weights

The GMM is appealing because it is adaptive, real-time, and allows to overcome the problem of multi-modal background processes. Stauffer and Grimson [5], [6] updated the parameter of each Gaussian with an IIR filter equation. The learning rate α is used to update the weight and a learning rate ρ is used for the updation of mean and variance. The learning rate α decides the sensitivity of the algorithm in varying condition and learning rate ρ control the adaptation rate to a new environment. If the learning rate ρ is chosen low then the convergence of Gaussian to the new background changes will be slow but will maintain the long learning history. On the other hand, a high value will improve the rate of convergence, but the new background model will not maintain sufficient historical information of previous images.

We have used an Gaussian weight learning rate scheme for GMM based background subtraction. In GMM, the parameters are updated over a single pixel and the concept of spatial redundancy is not taken into consideration for background maintenance. In our proposed scheme, If a match is found with one of the K Gaussian, then parameters of the GMM are updated in a small neighbourhood $\Omega(u)$ centred at u. The learning rate is given in eq (11).

$$\alpha = c \begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix} \tag{11}$$

If none of the pixel get matched to the existing Gaussian, then the least weighted component parameter is replaced with mean as its current pixel value, variance as an initial high variance and weight with low value of weight. The weights are normalized and then the ratio of w/σ are sorted in decreasing order. This ordering moves the most probable background with high weight and low variance at the top. The first B Gaussian distribution which exceed certain threshold T are retained for the background distributions. This is well illustrated in Algorithm 1.

B. Foreground Detection using Neighbourhood Correlation

A background contains both static and dynamic pixels. In our method, we have exploited neighbourhood correlation of pixels in foreground detection to avoid classification of any dynamic background pixels from becoming foreground. The first B Gaussian distribution calculated from Algorithm 1 is used in foreground detection. In Algorithm 2, for every foreground pixel we check in the neighbourhood using eq (4) for every B background component, whether the pixel is a dynamic background or not. If the condition satisfies then we classify the pixel as background and label the pixel as background pixel, otherwise the pixel remains labeled as foreground.

V. Experimental Results

The effectiveness of the proposed scheme is demonstrated on publicly available video sequences such as "MSA" [12], "Campus", "Intelligent Room", "Curtain", "Water Surface", "Fountain" and "Train". The sequence "Campus" and "Intelligent Room" are from CVRR Laboratory ATON project [13]. It contain shadows of foreground object. "Campus" is an outdoor sequence with larger shadow size as compared to the indoor video sequence of "Intelligent Room". The video sequence namely "Curtain", "Water Surface", and "Fountain" are complex video sequence with un-stationary background in the scene, are taken from I2R dataset [14].

To validate the proposed scheme, results obtained by it are compared with those of manual thresholding-based background subtraction (SBS) [15], $\Sigma - \Delta$ background subtraction (SDE) [2], W^4 background subtraction [3], ICA [16], RBS [17] and GMM [5], [6]. No post-processing operations such as morphological operation are applied in any of these algorithms to maintain the fairness in comparison.

A. Qualitative Evaluations

The first sequence used for the evaluation is "MSA" [12]. It consits of 528 frames with 320×240 spatial resolution. The video is acquired at a frame rate of 30 fps. In this video, a man comes, places the bag and then leaves the place. The video is used for the evaluation of shadow, as the person in

Algorithm 1 Gaussian weight learning rate for GMM based background subtraction

Parameters: $\sigma_0^2 \left(= 10^2\right)$, $w_0 \left(= 0.01\right)$
match = 0
for $i = 1 \rightarrow K$ **do**
  if $|I_{t,x} - \mu_{t-1,x,i}| \leq T_\sigma \, \sigma_{t-1,x,i}$ **then**
    \\ A match is found with one of the K Gaussian.
    match = 1
    for $u = 1 \rightarrow \Omega$ **do**
      \\ Update is done over a neighbourhood $\Omega\left(u\right)$
      $w_{t,x,i} = (1-\alpha) w_{t-1,x,i} + \alpha\left(M_{t,i}\right)$
      $\rho = \alpha\left(u\right) / w_{t,u,i}$
      $\mu_{t,u,i} = (1-\rho)\mu_{t-1,u,i} + \rho\left(I_{t,u}\right)$
      $\sigma_{t,u,i}^2 = \left(1-\rho\right)\sigma_{t,u,i}^2 + \rho\left(I_{t,u} - \mu_{t,u,i}\right)^2$
    end for
  else
    $w_{t,x,i} = (1-\alpha) w_{t-1,x,i}$
  end if
end for
if $(match = 0)$ **then**
  \\ Replacement phase
  $k = \arg\min_{i=1,\ldots,K} \; w_{t-1,x,i}$
  $\mu_{t,x,k} = I_{t,x}$
  $\sigma_{t,x,k}^2 = \sigma_0^2$
  $w_{t,x,k} = w_0$
end if
$w_{t,x,i} = w_{t,x,i} / \sum_{i=1}^{K} w_{t,x,i}$
$B = \arg\min_b \left(\sum_{i=1}^{b} w_i > T \right)$

Algorithm 2 Foreground Detection using Neighbourhood Correlation

$flag = 0$
for $i = 1 \rightarrow B$ **do**
  if $|I_{t,x} - \mu_{t-1,x,i}| \leq T_\sigma \, \sigma_{t-1,x,i}$ **then**
    $flag = 1$
    break
  end if
end for
if $flag == 1$ **then**
  $bl_{t,x} = 0$
else
  $bl_{t,x} = 1$
  $count = 0$
  for $i = 1 \rightarrow b$ **do**
    for $u = 1 \rightarrow \Omega$ **do**
      if $|I_{t,u} - \mu_{t-1,u,i}| \leq T_\sigma \, \sigma_{t-1,u,i}$ **then**
        $count = count + 1$
      end if
    end for
  end for
  if $count == Th$ **then**
    $bl_{t,x} = 0$
  end if
end if

the sequence cast a shadow on the floor and on the pillar in the total length of the sequence. The proposed technique is able to minimise the effect of shadow in the detection of the person. The second sequence is "Campus". This is used for the shadow detection. It is downloadable from ATON project site [13]. The video is of 288×352 spatial resolution and there are 1179 frames. The proposed sequence shows negligible effect of shadow in the detection of the car. The "Intelligent Room" is also available for download from the ATON project site [13]. The sequence consists of 300 frames of 320×240 spatial resolution. This is also used for the detection of shadow. Our scheme is able to detect the person in the presence of shadow. The "Curtain" video sequence can be easily downloaded from [14]. The spatial resolution of the video is 128×160 and consist of 2964 frames. In this sequence, a person is moving in the room in the backdrop of moving curtain (Venetian blinds). The clothe color of the person gets easily camouflaged with the Venetian blinds. Its a difficult scene to detect the person. In our proposed scheme, the person silhouette is accurately identified. The"Water Surface" [14], sequence consist of a 633 frames with a spatial resolution of 128×160. In this sequence, a person is moving on the bank of the river. The water in the river is moving and are falsely detected as foreground in the the much of the algorithm used for the comparison. We have also used "Fountain" [14] for the evaluation of un-stationary background. Water coming out from the fountain makes the detection of the person quite difficult. The proposed technique identifies the person in the backdrop of fountain. The last sequence used for the evaluation is 'Train". In this sequence the person is running with a brief case in his hand toward the train and in the meantime the train has started moving toward the person. This is unique where both the train and the person is moving against each other. Our interest lies in detecting the person and not the train. Its a difficult scenario to detect the person, when the train is also moving. The proposed technique is able to demonstrate its effectiveness in achieving the desired result. Experimental results demonstrate the effectiveness of the proposed method in providing a promising detection in presence of complex video sequence as shown in Fig. 1.

B. Quantitative Evaluations

The accuracy metric is obtained from Percentage of correct classification (PCC), given by

$$PCC = \frac{t_p + t_n}{t_p + t_n + f_p + f_n} \times 100 \tag{12}$$

Here true positive (t_p) represents the number of pixels classified correctly as belonging to the foreground and true negative (t_n), which counts the number of background pixel classified correctly. The false positive (f_p) is the number of pixels that are incorrectly classified as foreground and false negatives (f_n) represents the number of pixels which are wrongly labelled as background but should have been classified as foreground. PCC would attain values in [1-100]. The higher is the value, better is the accuracy.

Table I
COMPARATIVE ANALYSIS OF PCC

Approach	MSA	Campus	IR	Curtain	WS	Fountain	Train
SBS	98.58	95.77	99.7	98.05	97.13	96.1	91.71
SDE	90.82	86.28	92.54	88.82	96.04	95.77	87.7
W4	98.92	93.52	74.72	87.49	90.9	74.38	77.02
ICA	98.97	95.05	99.26	98.74	97.33	95.52	97.13
RBS	99.36	94.48	96.55	95.6	97.1	91.84	87.25
GMM	99.29	95.68	98.71	95.57	96.79	89.72	85.18
Proposed	**99.74**	**96.25**	**99.77**	**98.63**	**98.48**	**98.46**	**98.75**

Results on frames of "MSA", "Campus", "Intelligent Room", "Curtain", "Water Surface", "Fountain", and "Train" video sequences are provided in Table I. The higher PCC is obtained for the proposed BGS scheme as compared with those of other considered BGS techniques for the evaluation.

VI. Conclusion

In this paper we have presented a new Gaussian weight learning rate for Gaussian mixture model based background subtraction. The foreground detection is done by exploiting the neighborhood correlation of a pixel. The strength of the scheme lies in simple changes to the existing update equation and foreground detection of GMM model by considering the neighborhood of a pixel. A comparison has been made between the proposed algorithm and the state-of-the-art methods. The results obtained by the proposed scheme are found to provide accurate silhouette of moving objects in complex video sequence. The algorithm fails to detect shadow. In our future research, this problem will be at the center stage.

References

[1] M. Piccardi, "Background subtraction techniques: a review," in *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics*, vol. 4, 2004, pp. 3099–3104.

[2] A. Manzanera and J. C. Richefeu, "A new motion detection algorithm based on $\Sigma - \Delta$ background estimation," *Pattern Recognition Letters*, vol. 28, no. 3, pp. 320–328, Feb. 2007.

[3] I. Haritaoglu, D. Harwood, and L. S. Davis, "W4: Real-time surveillance of people and their activities," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, pp. 809–830, 2000.

[4] C. R. Wren, A. Azarbayejani, T. Darrell, and A. P. Pentl, "Pfinder: Real-time tracking of the human body," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, pp. 780–785, 1997.

[5] C. Stauffer and W. E. L. Grimson, "Adaptive background mixture models for real-time tracking," in *Computer Vision and Pattern Recognition*. IEEE Computer Society, 1999, pp. 2246–2252.

[6] ——, "Learning patterns of activity using real-time tracking," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, pp. 747–757, 2000.

[7] T. Bouwmans, F. E. Baf, and B. Vachon, "Background modeling using mixture of gaussians for foreground detection - a survey," *Recent Patents on Computer Science*, vol. 1, no. 3, pp. 219–237, 2008.

[8] K. Kim, T. H. Chalidabhongse, D. Harwood, and L. Davis, "Real-time foreground-background segmentation using codebook model," *Real-Time Imaging*, vol. 11, no. 3, pp. 172–185, Jun. 2005.

[9] P. Kaewtrakulpong and R. Bowden, "An improved adaptive background mixture model for real-time tracking with shadow detection," in *Proceedings of 2nd European Workshop on Advanced Video Based Surveillance Systems*, vol. 5308, 2001.

[10] D. S. Lee, "Effective Gaussian Mixture Learning for Video Background Subtraction," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 5, pp. 827–832, 2005.

[11] H.-H. Lin, J.-H. Chuang, and T.-L. Liu, "Regularized background adaptation: A novel learning rate control scheme for gaussian mixture modeling." *IEEE Transactions on Image Processing*, vol. 20, no. 3, pp. 822–836, 2011.

[12] "MSA test sequence," http://cvprlab.uniparthenope.it/index.php/download/61.html, accessed: 02-11-2013.

[13] "ATON project," http://cvrr.ucsd.edu/aton/shadow/, accessed: 02-11-2013.

[14] "Statistical modeling of complex background for foreground object detection," http://perception.i2r.a-star.edu.sg/bk_model/bk_index.html, accessed: 14-08-2013.

[15] A. C. Bovik, *The Essential Guide to Video Processing*, 2nd ed. Academic Press, 2009.

[16] D.-M. Tsai and S.-C. Lai, "Independent component analysis-based background subtraction for indoor surveillance," *IEEE Transactions on Image Processing*, vol. 18, no. 1, Jan. 2009.

[17] T. Horprasert, D. Harwood, and L. S. Davis, "A statistical approach for real-time robust background subtraction and shadow detection," in *Proc. IEEE International Conference on Computer Vision*, vol. 99, 1999, pp. 1–19.

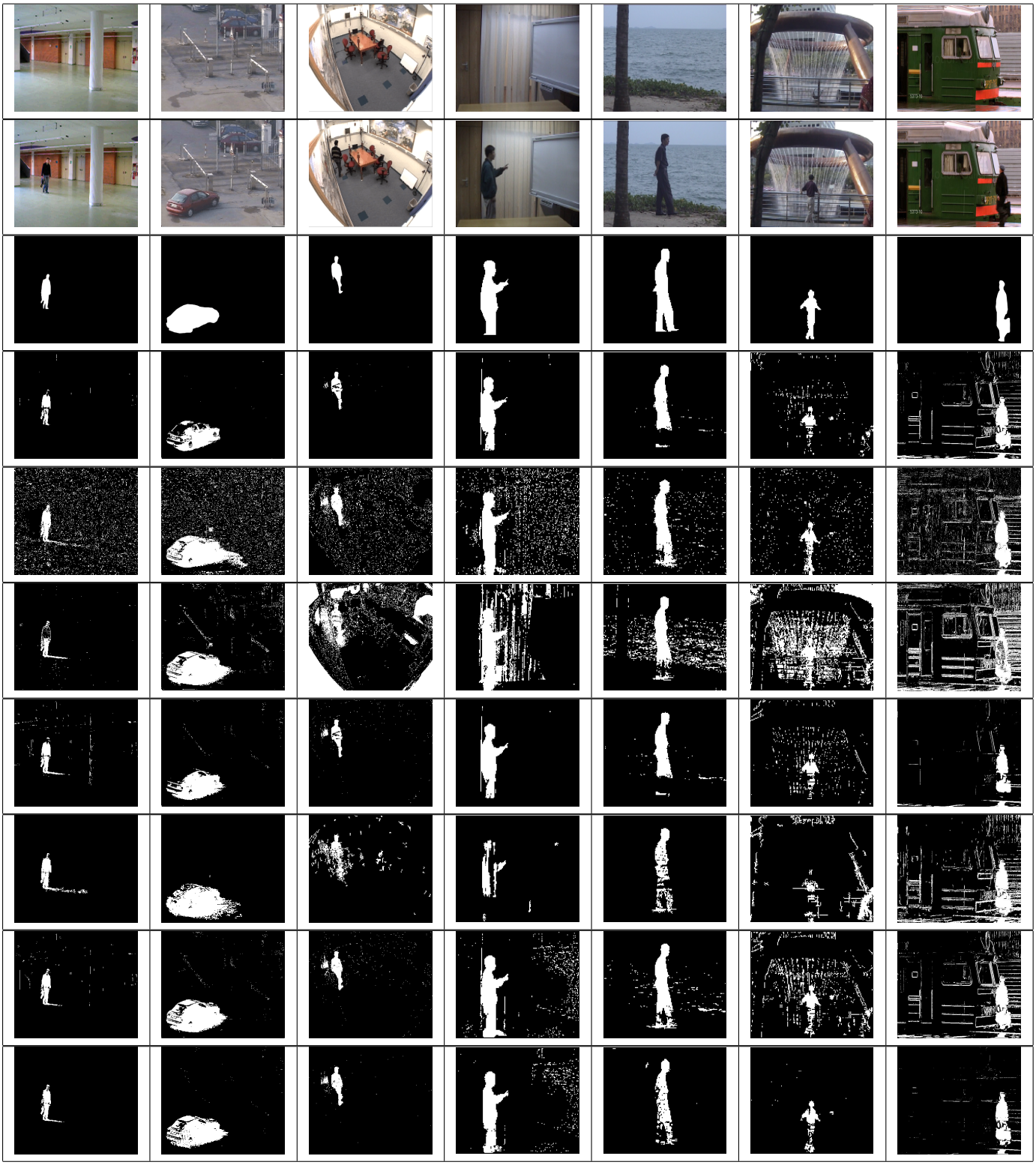

Figure 1. Left to right: MSA, Campus, Intelligent Room, Curtain, Water Surface, Fountain, Train. Top to bottom: Original Image, Test Image, Ground truth, Moving object detection for SBS, SDE, W4, ICA, RBS, GMM, Proposed scheme