Dynamic Background Subtraction Using Texton Co-occurrence Matrix

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: smeher@nitrkl.ac.in

***Abstract*—Moving object detection in the presence of changing illumination and non-stationary background such as swaying of trees, fountains, ripples in water, flag fluttering in the wind, camera jitters, noise, etc., is known to be very difficult and challenging task. Background subtraction (BS) is the most sought after technique for moving object detection. Still, most of the BS techniques do not take into account the spatial relationship between the pixels. In this paper, we have presented a novel BS algorithm using the properties of texton co-occurrence matrix (TCM) for accurately detecting the moving objects. TCM is a popular technique in the field of image retrieval. However, its adoption in BS is not reported in the literature. TCM integrates the colour, texture, and shape features in background modelling. It is computed in a neighbourhood region of the pixel. This implicitly utilizes image features and the spatial relationship between the pixels in the BS. Quantitative and qualitative results of the proposed algorithm is shown with state-of-the-art techniques, to prove its efficacy for moving object detection in presence of dynamic backgrounds.**

***Index Terms*—Visual surveillance, motion detection, background subtraction, non-stationary scene, illumination invariant, GLCM, texton co-occurrence matrix.**

I. Introduction

Automated visual surveillance system is being used in homes, offices, bus stands, railway stations, airports, religious places, banks, government and important buildings etc., to prevent from any kind of unlawful, anti-social and terrorism activities. Moving object detection is the first step in general for every visual surveillance system, succeeded by object classification, person identification, object tracking, action recognition, etc. Thus, moving object detection is a critical pre-processing step in the visual surveillance system. Moving object detection techniques can be divided into temporal differencing, BS and optical flow. BS [1]–[3] is the most popular and widely used technique among the three, for moving object detection in the presence of stationary cameras.

In the last decade, a large number of research papers have been proposed in the field of BS, to solve the problem caused by non-stationary background, illumination variation, addition or removal of background objects, shadow, noise, camera jitters, etc. Despite this, the performance of BS in the presence of non-stationary background such as swaying vegetation, fountain, changing illumination, ripples in water, flag fluttering in the wind, camera jitters etc., remains to be unsolved. BS techniques can be classified into pixel-based and region-based depending on the features and procedures used to construct the background model. The early approaches on BS were of pixel-based [1] and considered every pixel to be independent. The small changes in background will effect the neighbouring pixels in the background model and hence the performance of the pixel-based BS were poor in the presence of dynamic background. But pixel-based provides accurate silhouette of the moving objects. Region-based or block-based BS assume pixels to have some amount of correlation among its neighbours in background modelling. The background model is calculated from the block level features such as covariance [4], histogram, co-occurrence matrix [5], invariant moments [6], local binary pattern (LBP) [7], etc., by dividing the frames into overlapping or non-overlapping blocks. The correlation of pixels inside the block helps to overcome the general problem of non-stationary background and hence they are considered to be more robust approach than their pixel counterpart algorithms at the higher cost of computational complexity. However the detection results does not provide the accurate shape of the moving object.

We observe that, useful correlation exists in the neighbourhood of a pixel and the property of spatial correlation can be used in the greater deal to overcome the general problem of non-stationary scenes. TCM [8] originally proposed for content based image retrieval (CBIR) integrates colour, texture and shape features. The originality of the work lies in the use of TCM for background modelling, which have higher discrimination power than gray level co-occurrence matrix (GLCM). The use of TCM in background modelling incorporates the image features and spatial relationship between pixels without a considerable increase in its complexity. In this paper, we have inherited the advantages of both the pixel-based and region-based in our proposed BS algorithm. We have used the intensity and energy calculated from TCM in the local neighbourhood as a feature for background modelling. It helps in avoiding some of the shortcomings visited in earlier techniques in presence of the non-stationary scenes and illumination variation.

978-1-4799-5364-6/14/$31.00 ©2014 IEEE

The rest of the paper is organized as follows. In Section II, we have shown how TCM is used to compute the feature required for background modelling. The BS algorithm is implemented in Section III. In Section IV, we present a comparison of our proposed scheme with that of the state-of-the-art techniques in terms of visual as well as quantitative measures. Section V concludes the work done.

II. Background Modeling

Background modelling is the very important step in any BS algorithms. It depends on the choice of feature used in background construction. Here, we have chosen intensity and TCM properties as features for modelling the background. The foreground detection and maintenance of the algorithm is in accordance with the GMM algorithm.

A. Edge orientation detection in RGB color space

The edge of an image is an important feature in describing the contour and the texture structures present in the image. If the gradient magnitude and orientation are computed from the gray-scale image, then the chromatic information of the image is lost. In order to incorporate the colour information, we have computed the gradient [9] in RGB color space. Let r, g and b represent the unit vector along the R, G and B axes in RGB colour space. The vectors for the full colour image $f(x, y)$ is given by

$$u = \frac{\partial R}{\partial x}r + \frac{\partial G}{\partial x}g + \frac{\partial B}{\partial x}b \tag{1}$$

$$v = \frac{\partial R}{\partial y}r + \frac{\partial G}{\partial y}g + \frac{\partial B}{\partial y}b \tag{2}$$

The partial derivative required to compute the vectors are computed from the Sobel operator. Let g_{xx}, g_{yy} and g_{xy} denotes the dot product of these vectors:

$$g_{xx} = u^T u = \left|\frac{\partial R}{\partial x}\right|^2 + \left|\frac{\partial G}{\partial x}\right|^2 + \left|\frac{\partial B}{\partial x}\right|^2 \tag{3}$$

$$g_{yy} = v^T v = \left|\frac{\partial R}{\partial y}\right|^2 + \left|\frac{\partial G}{\partial y}\right|^2 + \left|\frac{\partial B}{\partial y}\right|^2 \tag{4}$$

$$g_{xy} = u^T v = \frac{\partial R}{\partial x}\frac{\partial R}{\partial y} + \frac{\partial G}{\partial x}\frac{\partial G}{\partial y} + \frac{\partial B}{\partial x}\frac{\partial B}{\partial y} \tag{5}$$

The direction of the maximum rate of change of I(x,y) is given by

$$\varphi(x, y) = \frac{1}{2}\tan^{-1}\left[\frac{2g_{xy}}{(g_{xx} - g_{yy})}\right] \tag{6}$$

The rate of change of image function I(x,y) is given by the angle as follows:

$$\boldsymbol{G}(x, y) = \left\{\frac{1}{2}\left[\begin{array}{l}(g_{xx} + g_{yy}) + (g_{xx} - g_{yy})cos2\varphi_0 \\ +2g_{xy}\,\sin 2\varphi_0\end{array}\right]\right\}^{1/2} \tag{7}$$

The angle $tan(\theta) = \tan(\theta \pm \pi/2)$, gives two equations for the gradient of image function $I(x, y)$.

$$G_1(x, y) = \left\{\frac{1}{2}\left[\begin{array}{l}(g_{xx} + g_{yy}) + (g_{xx} - g_{yy})cos2\varphi_0 \\ +2g_{xy}\,\sin 2\varphi_0\end{array}\right]\right\}^{1/2} \tag{8}$$

$$G_2(x, y) = \left\{\frac{1}{2}\left[\begin{array}{l}(g_{xx} + g_{yy}) + (g_{xx} - g_{yy})cos2(\varphi_0 + \pi/2) \\ +2g_{xy}\,\sin 2(\varphi_0 + \pi/2)\end{array}\right]\right\}^{1/2} \tag{9}$$

The maximum gradient is represented by M and the minimum gradient by N .

$$M = \max(G_1, G_2) \tag{10}$$

$$N = \min(G_1, G_2) \tag{11}$$

The color gradient M and N are normalized in the range of $[0, 1]$ and then mapped to 256 levels. After this the gradient values are quantized to W levels.

B. Colour quantization in RGB colour space

If the grayscale image is used in TCM computations then the colour information present in the image is lost. In order to avoid this, the original RGB image is first quantized to 256 colours using the intensity values of R, G and B channels as given in the equation below.

$$I(x, y) = 32 * F(R) + 4 * F(G) + F(B) \tag{12}$$

$$\left\{\begin{array}{ll} F(R) = 0, & 0 \leq R \leq 32 \\ F(R) = i, & 32 * i + 1 \leq R \leq 32 * (i + 1) \\ & i \in [1, 2, \ldots, 7] \end{array}\right. \tag{13}$$

$$\left\{\begin{array}{ll} F(G) = 0, & 0 \leq G \leq 32 \\ F(G) = i, & 32 * i + 1 \leq G \leq 32 * (i + 1) \\ & i \in [1, 2, \ldots, 7] \end{array}\right. \tag{14}$$

$$\left\{\begin{array}{ll} F(B) = 0, & 0 \leq B \leq 64 \\ F(B) = i, & 64 * i + 1 \leq B \leq 64 * (i + 1) \\ & i \in [1, 2, 3] \end{array}\right. \tag{15}$$

The 256 colored image is then quantized to W levels for TCM computations.

C. Texton co-occurrence matrix

The statistical TCM is computed for the maximum gradient, minimum gradient, and 256 coloured image. The computation of TCM is shown in Fig. 1. The detailed explanation on how to compute the TCM is given in [8] . The energy feature required for background modelling is calculated from TCM.

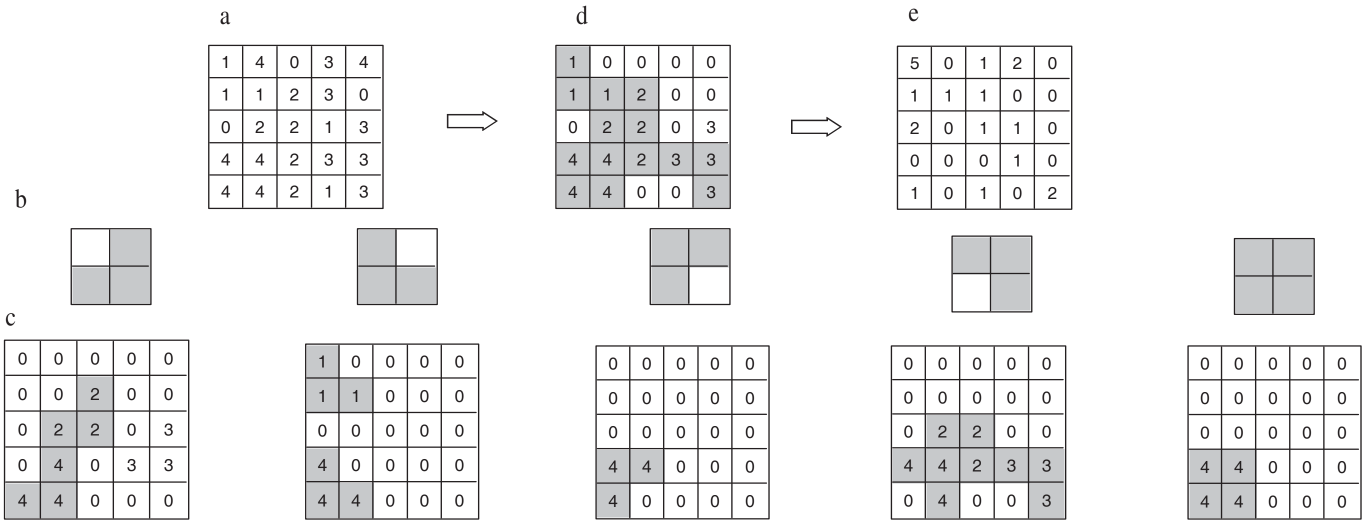

Fig. 1. Flow diagram for texton image computation: (a) original image, (b) different texton used, (c) five texton images, (d) the final texton image is computed by combining five texton images and (e) co-occurrence matrix ($d = 1$, $\theta = 0$) computed from the texton image.

1) Co-occurrence matrix: Co-occurrence matrix is a popular technique for describing spatial image features and finds its application in CBIR, texture analysis, object recognition, object tracking, moving object detection [5], etc. It considers the spatial layout of pixel in the neighbourhood and characterizes how often different combination of values occur at a given distance d and angle θ. The intensity level of I is quantized into W levels as $w \in \{0, 1, \ldots, W-1\}$ with $I(P) = w$. The co-ordinate of the pixel is $P = (x, y)$, let $P_1 = (x_1, y_1), P_2 = (x_2, y_2), I(P_1) = w, I(P_2) = \hat{w}$. The probability Pr of two intensity values $(w, \hat{w})$ occurring at a distance d and angle θ is given by

$$C_{d,\theta}(w, \hat{w}) = \Pr\left(\begin{array}{l} I(p_1) = w \wedge I(p_2) = \hat{w}, \\ \|p_1 - p_2\| = d, \theta \in \{0^\circ, 45^\circ, 90^\circ, 135^\circ\} \end{array}\right) \tag{16}$$

The co-occurrence matrix C is calculated from the texton matrix. Symmetric co-occurrence matrix is given by

$$C_s = C + C^T \tag{17}$$

C_{NS} represents the normalized symmetric co-occurrence matrix.

$$C_{NS} = C_S / \sum_{i=0}^{W-1} \sum_{j=0}^{W-1} c_S(i, j) \tag{18}$$

where c_S represent the element of symmetric co-occurrence matrix. The energy is calculated from the normalized symmetric co-occurrence matrix.

$$E = \sum_{i=0}^{W-1} \sum_{j=0}^{W-1} {c_{NS}}^2(i, j) \tag{19}$$

where c_{NS} represent the element of normalized symmetric co-occurrence matrix. A feature vector **f** is computed for each pixel in the frame is given by:

$$\mathbf{f} = [I \;\; E_I \;\; E_M \;\; E_N] \tag{20}$$

where I represent the intensity of 256 color image. E_I, E_M, E_N represents the energy calculated from the TCM of 256 color image, maximum gradient and minimum gradient respectively. The covariance matrix is computed in a neighbourhood region R of size $n \times n$ for a pixel centred at location (x, y)

$$Cov_{\mathbf{x},t} = \frac{1}{n^2} \sum_{k}^{n^2} (\mathbf{f}_k - \mu_{\mathbf{x},t})(\mathbf{f}_k - \mu_{\mathbf{x},t})^T \tag{21}$$

where $\mu_{\mathbf{x},t}$ is the mean feature vector calculated for a region R centred at (x, y).

III. Background Subtraction

The first frame of the video is used to initialized the mean feature vector and the covariance matrix of the background model. Let $w_{x,i,t}$ represent the weight associated with the ith model. The values of weight are initialize between 0 and 1, such that sum of the weights in the K models equals to 1. For maintaining the background model, the parameters $\mu^b_{\mathbf{x},i,t}$, $Cov^b_{\mathbf{x},i,t}$ and $w_{\mathbf{x},i,t}$ need to be updated based on the current mean feature $\mu_{\mathbf{x},t}$. The current feature vector is said to be matched to one of the background model, If the Mahalanobis distance is less than threshold T_p. Mahalanobis distance is given by

$$\Delta = (\mu_{x,t} - \mu^b_{x,i,t})^T \left(Cov^b_{x,i,t}\right)^{-1} (\mu_{x,t} - \mu^b_{x,i,t})^{1/2} \tag{22}$$

where $\mu_{x,t}$ represents the current mean feature vector. Let $\mu^b_{x,i,t}$ and $Cov^b_{x,i,t}$ denotes the mean feature vector and the covariance matrix of the background model. The minimum Mahalanobis distance whose value is less than T_p is chosen as the best match model. Based on the matching results the parameters of the best matched model, $\mu^b_{x,i,t}$, $Cov^b_{x,i,t}$ and $w_{x,i,t}$ are to be updated.

$$w_{x,i,t} = (1-\alpha)w_{x,i,t-1} + \alpha\,(M_{t,i}) \tag{23}$$

where $M_{t,i}$ is 1 for the best match and 0 for the rest.

$$\mu^b_{x,i,t} = (1-\rho)\mu^b_{x,i,t-1} + \rho\,(\mu_{x,t}) \tag{24}$$

$$Cov^b_{x,i,t} = (1-\rho)Cov^b_{x,i,t-1} + \rho Cov_{x,t} \tag{25}$$

where $0 \leq \alpha \leq 1$ and $0 \leq \rho \leq 1$ are the learning rate respectively.

If the Mahalanobis distance is greater than T_p for all the models, than the no match is found. The minimum weight component in model is replaced with current mean feature vector and the current covariance matrix. The weight of the model is replaced with low value of initial weight.

After this the weights are normalized, such that $\sum_{i=1}^{K} w_{x,i,t} = 1$. The K distributions are arranged in descending order by w/σ. This ordering moves the most likely background with high weight and low variance at the top. The first B Gaussian distribution which satisfies (26) are retained for the background distributions. The threshold T_b accounts for the minimum fraction of the background model. If a small value of T_b is chosen, the background model is uni-modal and is multi-modal, if higher value of T_b is chosen.

$$B \;=\; \arg\min_{b}\left(\sum_{i=1}^{b} w_i \;>\; T_b\right) \tag{26}$$

If the distance between the pixel of the current frame and the model is less than Tp for any one of the background component, then the pixel is marked as background. Otherwise, the pixel will be classified as foreground.

IV. Experimental Results and Comparison

Evaluation and comparison of the BS techniques is an important issue. In this section, we compare the performance of our proposed algorithm quantitatively and qualitatively with some of the well known state-of-the-art techniques in this field such as: GMM [1], Covariance BS [4], Moment BS [6], LBP BS [7] and STBS [5]. To maintain the un-biased comparison, no pre or post processing operations such as median filtering, morphological operation, etc. have been applied to any of these algorithms.

A. Test Sequences

The test sequence used for the performance evaluation of the proposed algorithm contains various challenging situations faced in real life situation such as non-stationary backgrounds, illumination variation, camouflage, etc. The video sequences are shot at both indoor and outdoor environments. All these test videos are taken from the publicly available dataset. The test videos “Curtain”, “Campus”, “Lobby”, “Watersurface” and “Fountain” are taken from I2R dataset [10]. The video sequence “Camouflage” and “Waving trees” are taken from walflower dataset [11].

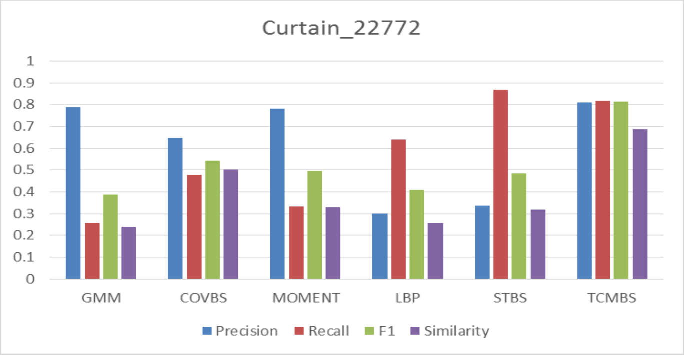

Fig. 2. Performance analysis of Curtain 22772 frame.

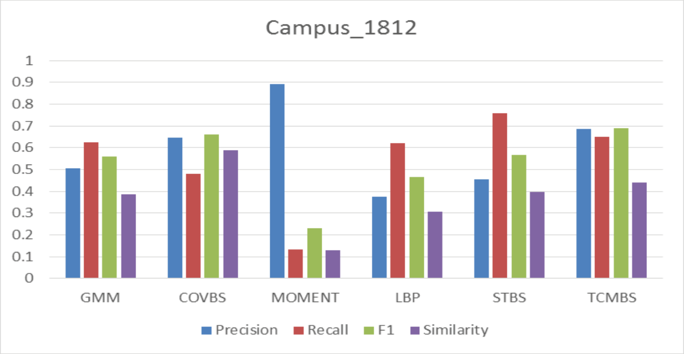

Fig. 3. Performance analysis of Campus 1812 frame.

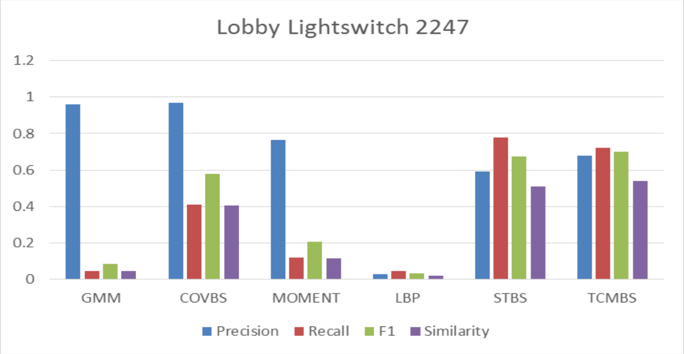

Fig. 4. Performance analysis of Lobby Lightswitch 2247 frame.

The parameter used for the proposed TCMBS are as follows: $R = 5$, $W = 16$, $d = 1$, $\theta = 0$, $K = 3$, $Tb = 0.75$, $\alpha = 0.01$, $\rho = 0.005$ and $Tp = (tuned)$. The parameter selection for all the state-of-the-art BS algorithm is taken same as given in their seminal papers.

B. Quantitative Evaluation Metrics

Precision denotes the ratio of detected true positives as compared to the total number of pixels detected by the method.

$$Precision = \frac{t_p}{t_p + f_p} \tag{27}$$

Recall is the ratio of detected true positives to the total number of pixels present in the ground truth.

$$Recall = \frac{t_p}{t_p + f_n} \tag{28}$$

F_1 metric, also known as Figure of Merit or F-measure. It is the weighted harmonic mean of precision and Recall.

$$F_1 = \frac{2 * Recall * Precision}{Recall + Precision} \tag{29}$$

and finally pixel-based Similarity measure is defined as:

$$Similarity = \frac{t_p}{t_p + f_n + f_p} \tag{30}$$

Here, true positive t_p represents the number of pixels classified correctly as belonging to the foreground and true negative t_n means 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.

The performance analysis of the proposed (TCMBS) algorithm is done using various complex dynamic test videos such as “Curtain”, “Campus” , and “Lobby Lightswitch” . It is shown in Fig. 2, 3 and 4 respectively. The TCMBS algorithm provides the highest value of $F1$ and $similarity$ for the “Curtain” sequence. Similarly TCMBS algorithm provides the highest value of precision, $F1$ and $similarity$ for the “Campus” sequence. The moment based algorithm shows higher value of $precision$ for “Campus” sequence as it gives higher false positive. In ”Lobby” sequence, the GMM, Covariance BS, and “Moment BS” shows higher $precision$ as compared to the TCMBS due to increase in false positive in presence of illumination variation. The proposed algorithm shows higher value of $F1$ and $similarity$ for “Lobby” sequence. The performance of the proposed algorithm is very much satisfactory with all the test sequence and in some instance it outperforms some of the well known state-of-the-art algorithm in BS.

C. Qualitative Analysis

The proposed TCMBS algorithm performs very well in presence of non-stationary scenes, illumination variation, and camouflage. The subjective evaluation of the algorithm is shown in Fig. 5. In “curtain” video sequence there is a non-stationary background movement of the venetian blinds. The proposed algorithm is able to reduce the effect of non-stationary scene and is able to detect the accurate silhouette of the moving object. The “Campus” sequence also contain the dynamic scene due to the non-periodic movement of the leaves of the tree. It is a very challenging sequence to detect the person and the vehicle with accurate segmentation of the foreground object. The “Lobby” sequence is for illumination variation. Persons are moving in the lobby in the amidst of the instantaneous light changes. Our proposed algorithm is able to detect the person and diminishes the effect of light variation in the video. The GMM, COV BS and Moment BS shows an increase in false positive for swaying vegetation, illumination variation and camouflage sequences. LBP BS gives higher false negative as compared to other algorithms. The STBS algorithm shows false negative in “Curtain” and “Camouflage” sequence. The algorithm is missing some part of the foreground object in “Curtain” and “Camouflage” sequence. The results of the proposed TCMBS algorithms gives accurate silhouette of the moving object.

V. Conclusion

In this paper, we have presented a novel BS using TCM for accurately detecting the moving objects. The TCM is computed in a neighbourhood region of the pixel. This implicitly utilizes image features and the spatial relationship between the pixels in the BS. The use of intensity and energy feature calculated from the TCM The algorithm can solve the problem associated with non-stationary background, illumination variation and camouflage. The experimental results of the proposed algorithm, outperforms some of the well established traditional algorithms in this field.

References

[1] C. Stauffer and W. E. L. Grimson, “Learning patterns of activity using real-time tracking,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, pp. 747–757, 2000.

[2] D. Panda and S. Meher, “A gaussian mixture model with gaussian weight learning rate and foreground detection using neighbourhood correlation,” in *IEEE Asia Pacific Conference on Postgraduate Research in Microelectronics and Electronics (PrimeAsia)*. IEEE, 19-21 Dec. 2013, pp. 158 – 163.

[3] ——, “Video object segmentation based on adaptive background and wronskian change detection model,” in *International Conference on Signal Processing and Communication (ICSC),*, 2013, pp. 219 – 224.

[4] S. Zhang, H. Yao, S. Liu, X. Chen, and W. Gao, “A covariance-based method for dynamic background subtraction.” in *ICPR*. IEEE, 2008, pp. 1–4.

[5] P. Chiranjeevi and S. Sengupta, “Moving object detection in the presence of dynamic backgrounds using intensity and textural features,” *Journal of Electronic Imaging*, vol. 20, no. 4, pp. 043 009–1 – 043 009–11, 2011.

[6] R. Marie, A. Potelle, and E. M. Mouaddib, “Dynamic background subtraction using moments.” in *ICIP*. IEEE, 2011, pp. 2369–2372.

[7] M. Heikkilä and M. Pietikäinen, “A texture-based method for modeling the background and detecting moving objects,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 4, pp. 657–662, 2006.

[8] G.-H. Liu and J.-Y. Yang, “Image retrieval based on the texton co-occurrence matrix,” *Pattern Recognition*, vol. 41, no. 12, pp. 3521–3527, Dec. 2008.

[9] S. D. Zenzo, “A note on the gradient of a multi-image,” *Computer Vision, Graphics, and Image Processing*, vol. 33, no. 1, pp. 116–126, 1986.

[10] L. Li, W. Huang, I. Y. H. Gu, and Q. Tian, “Statistical modeling of complex backgrounds for foreground object detection,” *IEEE Transactions on Image Processing*, vol. 13, no. 11, pp. 1459–1472, 2004, http://perception.i2r.a-star.edu.sg/bk_model/bk_index.html.

[11] K. Toyama, J. Krumm, B. Brumitt, and B. Meyers, “Wallflower: principles and practice of background maintenance,” *Proceedings of the Seventh IEEE International Conference on Computer Vision*, vol. 1, no. c, pp. 255–261, 1999, http://research.microsoft.com/en-us/um/people/jckrumm/wallflower/testimages.html.

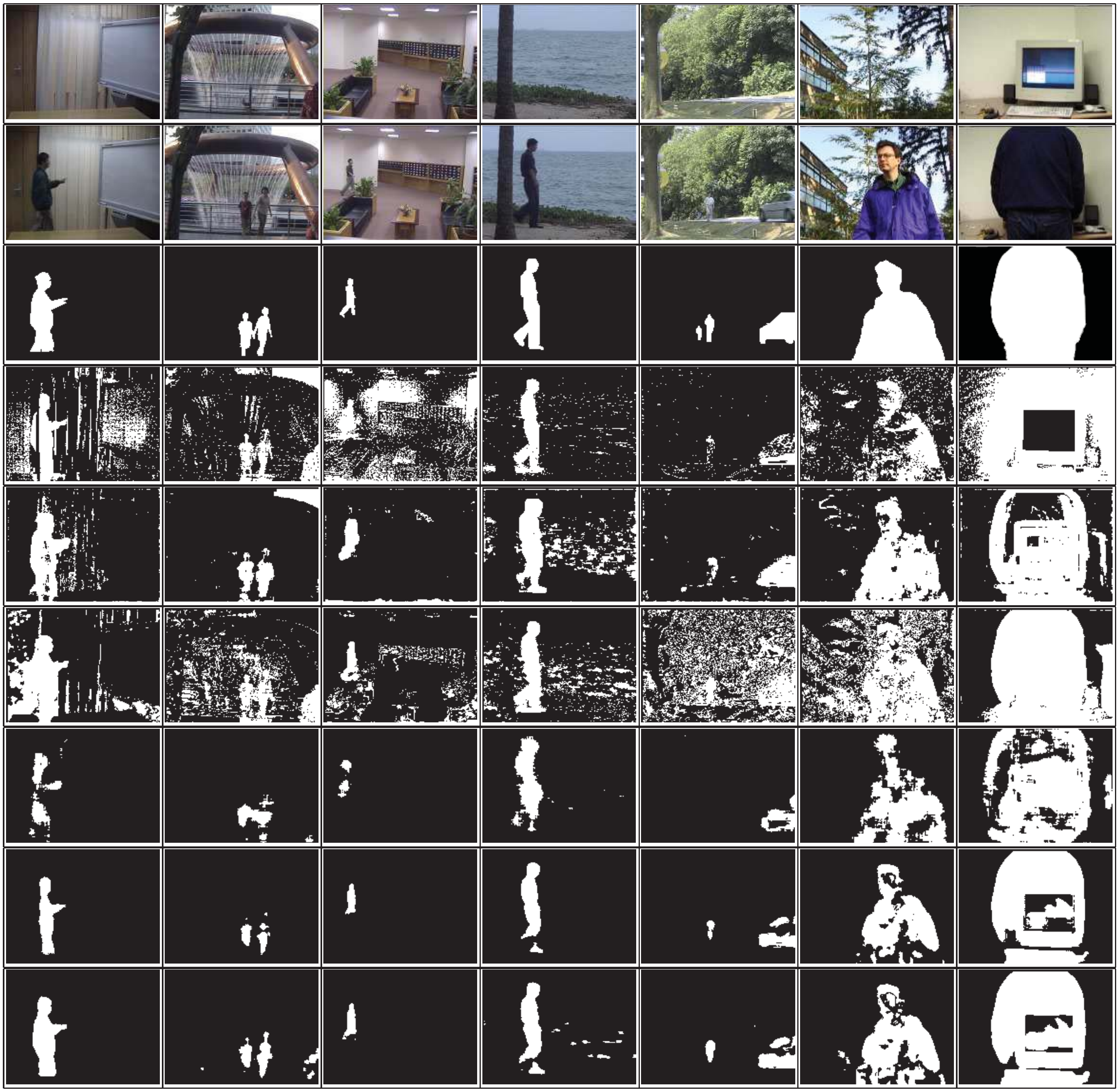

Fig. 5. Left to right: Curtain, Fountain, Lobby, Water Surface, Campus, Waving trees, Camouflage. Top to bottom: Original Image, Test Image, Ground truth, Moving object detection for GMM, Covariance BS, Moment BS, LBP BS, STBS, TCMBS