Cognitive Computation (2019) 11:18–30
https://doi.org/10.1007/s12559-018-9594-5

3D Local Spatio-temporal Ternary Patterns for Moving Object Detection in Complex Scenes

Srikanth Vasamsetti[1,2] · Neerja Mittal[1,2] · Bala Chakravarthy Neelapu[1,2] · Harish Kumar Sardana[1,2]

Received: 7 June 2017 / Accepted: 4 September 2018 / Published online: 14 September 2018

Abstract

Humans possess natural cognitive vision to perceive objects in a 3D space and are able to differentiate foreground and background moving objects using their shape, colour and texture. Moving object detection is a leading and challenging task in complex scenes which involve illumination variation, blurriness, camouflage, moving background objects, etc. Inspired by human cognitive vision, a novel descriptor named 3D local spatio-temporal ternary patterns (3D-LStTP) is proposed for moving object detection. The 3D-LStTP collects multidirectional spatio-temporal information from three consecutive frames in a video by forming a 3D grid structure. The background models are constructed by using texture and colour features. The results obtained after modelling are integrated for foreground moving object detection in complex scenes. The performance of proposed algorithm is validated by conducting five experiments on Fish4Knowledge dataset, four experiments on I2R dataset and four experiments on Change Detection dataset. Qualitative and quantitative analyses are carried out on benchmark datasets. The results after investigation prove that the proposed method outperforms the state-of-the-art techniques for moving object detection in terms of ROC, TPR, FPR, Precision and F- measure.

Keywords Background subtraction · Colour feature · Dynamic background · Moving object detection · Texture feature

Introduction

Moving object detection in complex video sequences is one of the leading task for video processing applications such as tracking and visual surveillance. The object detection in underwater environment is a challenging task as compared to natural video scenes. Underwater environment videos are usually degraded due to the physical properties (colour absorption and light scattering) of water medium. The extraction of feature descriptors in these conditions would raise a challenging problem which comprises noise-corrupted video frames, often few observed to be of poor quality, abrupt changes in illumination, shadow of the moving objects, crowded environment, camouflaged foreground objects, dynamic background such as swaying trees, rippling water and camera movements. These extreme conditions are the motivation for our study in automatic detection of underwater moving objects.

In recent years, foreground and background modelling techniques have been developed by adopting spatial, colour, texture and temporal information in dealing with complex scenarios. However, very few methods were tested under extreme conditions (underwater domain). Background subtraction by utilising different features is a commonly used technique for detection of moving objects in complex scenarios. In the early stage of research, the single Gaussian model [34] was used to model the background at each pixel location of the frame due to which it is robust to static background but could not handle the dynamic background variations. Stauffer and Grimson [28] proposed Gaussian Mixture Model (GMM) to address this problem. Various modified versions of GMM [2, 10] were developed to improve the effectiveness of background modelling and to handle dynamic background situations.

In recent times, techniques for identifying the number of Gaussians "on-the-fly" have been proposed [15, 40], and these methods are not successful. To overcome the complexity of identifying the suitable shape of the probability density function (pdf), non-parametric methods are adopted.

✉ Srikanth Vasamsetti
srikanth.vasamsetti@csio.res.in

[1] Academy of Scientific & Innovative Research (AcSIR), Chandigarh, India

[2] Computational Instrumentation, CSIR-Central Scientific Instruments Organisation, Chandigarh 160030, India

Bayesian modelling for moving object detection in dynamic scenes is proposed in [24]. The work presented in [26] uses a joint domain-range kernel density estimation algorithm for modelling the background and foreground by using texton features which is employed on underwater videos as well for testing under extreme conditions. The development of novel features for background modelling reduces the limitations of normal features, and the integration of several heterogeneous features (texture, colour and/or motion) has improved the performance of the background subtraction technique [6, 32]. Kim et al. proposed a codebook model technique based on clustering methodology [13]. Using colour distortion and brightness distortion, pixel values are clustered into a set of code words to build the background model. Several notable changes [5] were made to improve the original codebook model.

To construct the background model, mostly three types of features, i.e., pixel, region and temporal based were used in background subtraction techniques. Pixel-based features are constructed using the information extracted from a single pixel. Background model is affected due to the variation of pixel characteristics in complex backgrounds. Using pixel-based features alone may not accurately segment the moving objects in dynamic backgrounds. However, these features [7, 34] are effective in extracting the shape of the moving objects. Moving object detection and tracking by using annealed background subtraction technique are presented by [11].

Tu et al. [30] and Tong et al. [36] proposed spatio-temporal saliency-based moving object techniques using human cognitive vision. However, these techniques framed on pixel-based approach for modelling the background and the useful information obtained from consecutive frames were not considered which leads to noise-corrupted signal. A biologically inspired vision-based method for moving object detection in complex scenes is proposed by Zhengzheng et al. [31]. It integrates principal component analysis (PCA) with independent component analysis (ICA) to detect moving objects. The author also mentioned about further improvement in detecting minor objects. Jing et al. [23] proposed an incremental technique for moving object detection named as ISC (where I, S and C stand for incremental, sparsity and connectivity, respectively). The constraints of connectivity and foreground sparsity are combined for the subspace-based objective function. Collaborative Low-Rank And Sparse Separation (CLASS) technique is proposed by Aihua et al. [39]. In this method, sparse outliers and low-rank model are used for foreground and background modelling respectively. The author highlights the failure of technique when the objects in foreground and background possess similar colour. Akilan et al. [1] proposed fusion-based foreground background separation by using multivariate multimodel Gaussian distribution. In this, a novel foreground enhancement approach is introduced by assimilating illumination and colour measures.

Motion cues are the most reliable information for moving object detection. Hence, moving object detection methods are designed based on temporal information such as frame difference [33, 38] and background subtraction [14, 18]. Frame difference method is simple and fast but it is incapable of extracting entire contour of the moving object. Whereas, background subtraction is a typical approach to identify moving objects, which separates the background model from the input frames. To build a background model in complex scenarios, region-based methods [8, 37] are used by taking advantage of local pixel relationships. In [22], local binary pattern (LBP), a basic texture feature descriptor was introduced to encode a spatial relationship among pixels of a specific region in an image. To improve its robustness, Tan and Triggs [29] presented a local ternary pattern (LTP) operator by modifying the thresholding scheme to three levels. LTP is computed from three orthogonal planes in human action classification [19].

Scale invariant local ternary pattern (SILTP) operator, a variant of LTP was proposed by Liao et al. [17] by introducing a scale transform factor to handle the illumination variations in the scene. In general, the texture features are extracted based on the relationship between the centre (reference) pixel and its surrounding neighbours. The LTP [29] and SILTP [17] collect the texture information within a single frame, whereas ST-SILTP [9] collects the spatio-temporal texture information between the two consecutive frames. Humans possess natural cognitive vision to perceive objects in a 3D space and able to differentiate foreground and background moving objects using its shape, colour and texture. The feature descriptors mentioned above have either spatial information or spatio-temporal information. Inspired by human cognitive vision, a new feature descriptor named three-dimensional local spatio-temporal ternary patterns (3D-LStTP) for complex background subtraction is proposed. The main contributions and novelty of this work are stated below:

- A novel 3D-LStTP feature descriptor is proposed in which collection of local spatio-temporal information from three consecutive video frames by forming 3D grid in a particular direction for a given centre pixel.
- Block-based technique along with a single histogram model is implemented for each block.
- Two-level block-based approach is used. Background models are constructed in big blocks for stability, while the final foreground detection decisions are made for small blocks to achieve finer boundary.
- Integration of segmented results obtained after background modelling using 3D-LStTP texture features and L^*, a^*, b^* colour features. The colour and texture

information can compensate for their respective shortcomings and their amalgamation can bring more stable performance on varied video sequences.
– The performance of the proposed technique is improved for both natural and underwater moving object detection applications.

The organisation of the manuscript is as follows: In "Introduction", a brief review of various moving object detection techniques is presented. Section "Local Patterns" presents a concise review of local patterns (the LBP, the LTP and the SILTP). Section "Proposed Moving Object Detection Framework" presents background modelling using texture and colour features. The proposed moving object detection framework is given in "Experimental Results and Discussions". Experimental results of various techniques are analysed in "Conclusion".

Local Patterns

Local Binary Patterns (LBPs)

LBP, a basic texture feature descriptor, was introduced by Ojala et al. [21]. For a given a central pixel in the 3 × 3 region, LBP value is calculated by comparing its intensity value with its neighbours' as follows:

$$LBP_{D,R} = \sum_{d=0}^{D} 2^{(d-1)} \times f\,(I_d - I_c) \tag{1}$$

$$f\,(x) = \begin{cases} 1 & x \geq 0 \\ 0 & x < 0 \end{cases} \tag{2}$$

where R is the radius of neighbours, D is the number of neighbours, I_c is the intensity value of centre pixel and I_d is the intensity value of its neighbours.

Local Ternary Patterns (LTPs)

Tan and Triggs presented a local ternary pattern by extending LBP to three-level threshold values. The intensity values in the range of width $\pm w$ around a central pixel j_c are quantised to zero, those above $j_c + w$ are quantised to +1 and those below $j_c - w$ are quantised to −1, i.e. the function $k(v)$ is replaced with three-valued function and binary LBP code is replaced by a ternary LTP code as follows:

$$K\,(v, j_c, w) = \begin{cases} 1, & v \geq j_c + w \\ 0, & |v - j_c| < w \\ -1, & v \leq j_c - w \end{cases} \tag{3}$$

Scale Invariant Local Ternary Patterns (SILTP)

The LBP descriptor suffers from two problems. It is sensitive to the local noises and second is the intensity scale variation results in false classification of two or more different patterns into the same class. However, in the case of LTP, the first problem is addressed by introducing three-level threshold values and the second problem has occurred only in few cases. To address these problems, scale invariant local ternary pattern (SILTP) operator was presented by Liao et al. [17]. A scale transform factor is introduced to handle the illumination variations in the scene. For a given central pixel in the 3 × 3 region, SILTP pattern is computed as

$$SILTP^{\tau}_{D,R} = \bigoplus_{d=0}^{D-1} f\,(I_d, I_c) \tag{4}$$

$$f(x) = \begin{cases} 01 & I_d > (1+\tau) \times I_c \\ 10 & I_d < (1-\tau) \times I_c \\ 00 & otherwise \end{cases} \tag{5}$$

where I_c is the intensity value of centre pixel, I_d is the intensity value of its neighbours, $\oplus$ denotes concatenation operator and τ is a scale factor.

Proposed Moving Object Detection Framework

In this section, we propose a 3D-LStTP-based method for underwater moving object detection. Three consecutive frames are selected for 3D local spatio-temporal ternary patterns calculation by forming 3D grid. Two consecutive frames are chosen for generating temporal background in colour feature extraction. Background is modelled using 3D-LStTP features and Lab colour model. The flowchart of proposed underwater moving object detection method is shown in Fig. 1.

3D Local Spatio-Temporal Ternary Patterns (3D-LStTP)

The proposed 3D-LStTP collects the texture information from the 3D grid which is formed by collecting t, $t-1$ and $t-2$ frames in a given video. Figure 2 illustrates the 3D grid formation for a given reference frame t. After forming the 3D grid for a reference frame t, the features are coded based on the directional spatio-temporal information as given in Eq. 6.

$$V^{\alpha}\big|_{P=8} = \begin{bmatrix} \{F^t(P-\alpha-1), F^t(C), F^t(\Gamma),\ F^{t-1}(\Gamma), F^{t-2}(\Gamma), \\ F^{t-2}(C), F^{t-2}(P-\alpha-1), F^{t-1}(P-\alpha-1)\} \\ \forall\,\Gamma = \mathrm{mod}((x+P), P);\ x = floor\left(\frac{(P-2\alpha-1)}{2}\right) \end{bmatrix} \tag{6}$$

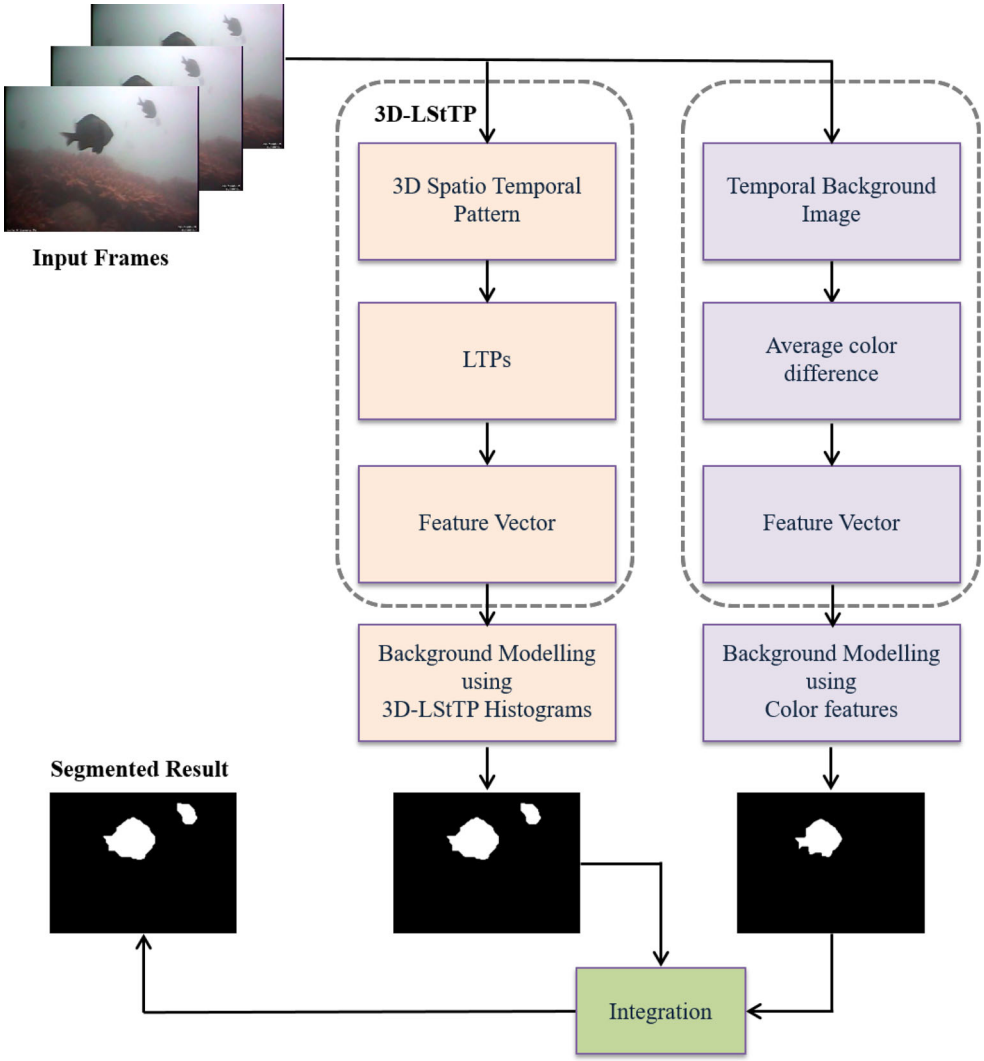

Fig. 1 Schematic diagram of the proposed method

where, $mod(r, s)$ provides the remainder for r/s operation, P represents the number of neighbours and represents the possible direction ($\alpha = 1, 2, 3, 4$) for spatio-temporal neighbours' collection.

After collecting spatio-temporal information from the neighbours, a 3D-LStTP is constructed by calculating the relationship between centre pixel $F^{t-1}(C)$ and its surrounding neighbours as follows:

$$Y^{\alpha}|_{P=8} = \left\{ \begin{bmatrix} f_1\left(F^t(P-\alpha-1), F^{t-1}(C)\right), f_1\left(F^t(C), F^{t-1}(C)\right), \\ f_1\left(F^t(\Gamma), F^{t-1}(C)\right), f_1\left(F^{t-1}(\Gamma), F^{t-1}(C)\right), \\ f_1\left(F^{t-2}(\Gamma), F^{t-1}(C)\right), f_1\left(F^{t-2}(C)\right), F^{t-1}(C)\right) \\ f_1\left(F^{t-2}(P-\alpha-1), F^{t-1}(C)\right), \\ f_1\left(F^{t-1}(P-\alpha-1), F^{t-1}(C)\right) \end{bmatrix} \right. \\ \left. \forall\, \Gamma = \mathrm{mod}((x+P), P);\ x = floor\left(\frac{(P-2\alpha-1)}{2}\right) \right. \tag{7}$$

$$f_1(a, b)|_{\tau=0.05} = \begin{cases} 2, & a > b(1+\tau) \\ 1, & a < b(1-\tau) \\ 0, & else \end{cases} \tag{8}$$

The 3D-LStTP (D) value for a particular selected direction is computed by Eq. 9.

$$D_P^{\alpha} = \sum_{i=0}^{P-1} 3^i \times Y^{\alpha}(i) \tag{9}$$

Three consecutive frames were selected for 3D grid structure formation from the given video sequences. An example of 3D-LStTP calculation for a given centre pixel (intensity value 200) highlighted with red colour in the 3D grid is explained in Fig. 3. Let the direction selected for 3D-LStTP calculation is "$\alpha = 1$". The 3D spatio-temporal neighbours collected based on Eq. 6 are {213, 199, 221, 93, 112, 111, 205, 203}. The LTP values were calculated using Eqs. 6 and 7 by taking the difference between neighbour intensity values and a scale transform factor (τ =0.05) multiplied with centre pixel value. The threshold computed after multiplying centre pixel intensity value with upper scale transform factor ($a > b(1+\tau)$) is 210 and lower scale transform factor ($a < b(1-\tau)$) is

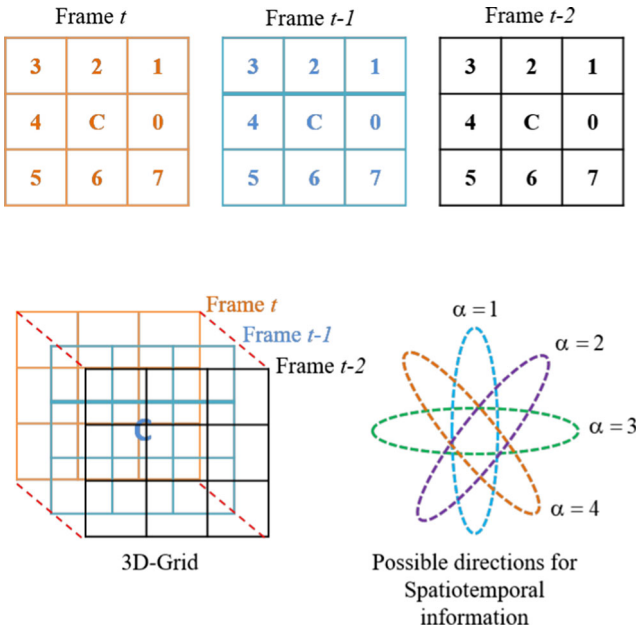

Fig. 2 3D grid formation and possible directions for spatio-temporal information collection

190. In the first case, LTP values are coded by replacing positive values with 2 and replacing 0 for negative values. In the second case, LTP values are coded by replacing negative values with 1 and replacing 0 for positive values. Finally, 3D-LStTP value (371) is coded based on Eq. 9 by multiplying LTPs with weights for representation of spatio-temporal structure of local 3D volume.

After identifying the local pattern, 3D-LStTP features are used for background modelling. The feature vector length calculated for all the four possible directions (α=1,2,3,4) is 26244 (4×3^8), which increases the computational complexity of the algorithm. To reduce the same, only two directions (α=1,3) are considered for feature extraction.

Background Modelling Based on 3D-LStTP Features

Each incoming frame is divided into overlapping big blocks and non-overlapping small blocks. A histogram is calculated for each big block using 3D-LStTP descriptor. Suppose $h_k^1, h_k^2, ..., h_k^t$ be the 3D-LStTP histograms and $b_k^1, b_k^2, ..., b_k^t$ be the background model of each big block of a frame at time $1, 2, ..., t$ respectively. To initialise the background model, b_k value for each big block $(b_k^0(1), b_k^0(2), ..., b_k^0(t))$ is taken as $1/N_{bb}$. Background model needs to be updated for acquiring pattern changes in each block area, which is given as:

$$b_k^t(i) = (1-\beta) \times b_k^{t-1}(i) + \beta \times h_k^t(i) \tag{10}$$

where, β is the learning rate and $b_k^t(i)$ is the i^{th} bin of background model at time t. The learning rate (α) = 0.005 is used for model updating. After background modelling, probability is calculated for each big block, defined as:

$$P_b^{bb} = \sum_{i=1}^{N_{bb}} h_k(i) \times X\left(b_k(i), \frac{1}{N_{bb}}\right) \tag{11}$$

where,

$$X\left(b_k(i), \frac{1}{N_{bb}}\right) = \begin{bmatrix} 1 & b_k(i) \geq \frac{1}{N_{bb}} \\ 0 & b_k(i) < \frac{1}{N_{bb}} \end{bmatrix} \tag{12}$$

where, P_b^{bb} is the probability of big block belonging to the background and N_{bb} is the maximum possible histogram bins. In Eq. 12, only those background patterns are taken into account whose value is greater than the threshold. The outcome of Eq. 12 is further used for computing the probability of big blocks which are summed up in Eq. 11 to get the final background patterns. Based on the probability of each big block, small block probabilities are computed. Further, the probability of a small block belonging to the background is calculated as follows:

$$P_b^{sb} = \frac{\sum_{i=1}^{m} P_b^{bb(i)}}{m} \tag{13}$$

where, P_b^{sb} is the background likelihood of the small, m is the number of shared big blocks by each small block and $P_b^{bb(i)}$ is the probability of big block. If $P_b^{sb} > Th$, then it indicates the small block belongs to the background, else it is foreground. The threshold Th is calculated for every incoming frame using Otsu's method [12]. Otsu introduced an automatic threshold selection technique for picture segmentation. An optimal threshold is selected by the discriminant criterion, namely by maximising the discriminant measure (or the measure of separability of the resultant classes in gray levels). The proposed method is characterised by its nonparametric and unsupervised nature of threshold selection. The procedure is very simple, utilising only the zeroth- and the first-order cumulative moments of the gray-level histogram.

Colour Features

The colour and texture information can compensate for their respective shortcomings and their amalgamation can bring more stable performance on varied video sequences. For each video frame, a temporary background image is constructed according to following equation

$$Z_b^t = \begin{cases} 0 & if\ \mathrm{t} = 0 \\ (1-\gamma) Z_b^{t-1} + \gamma Z_{\inf} & if\ \mathrm{t} > 0 \end{cases} \tag{14}$$

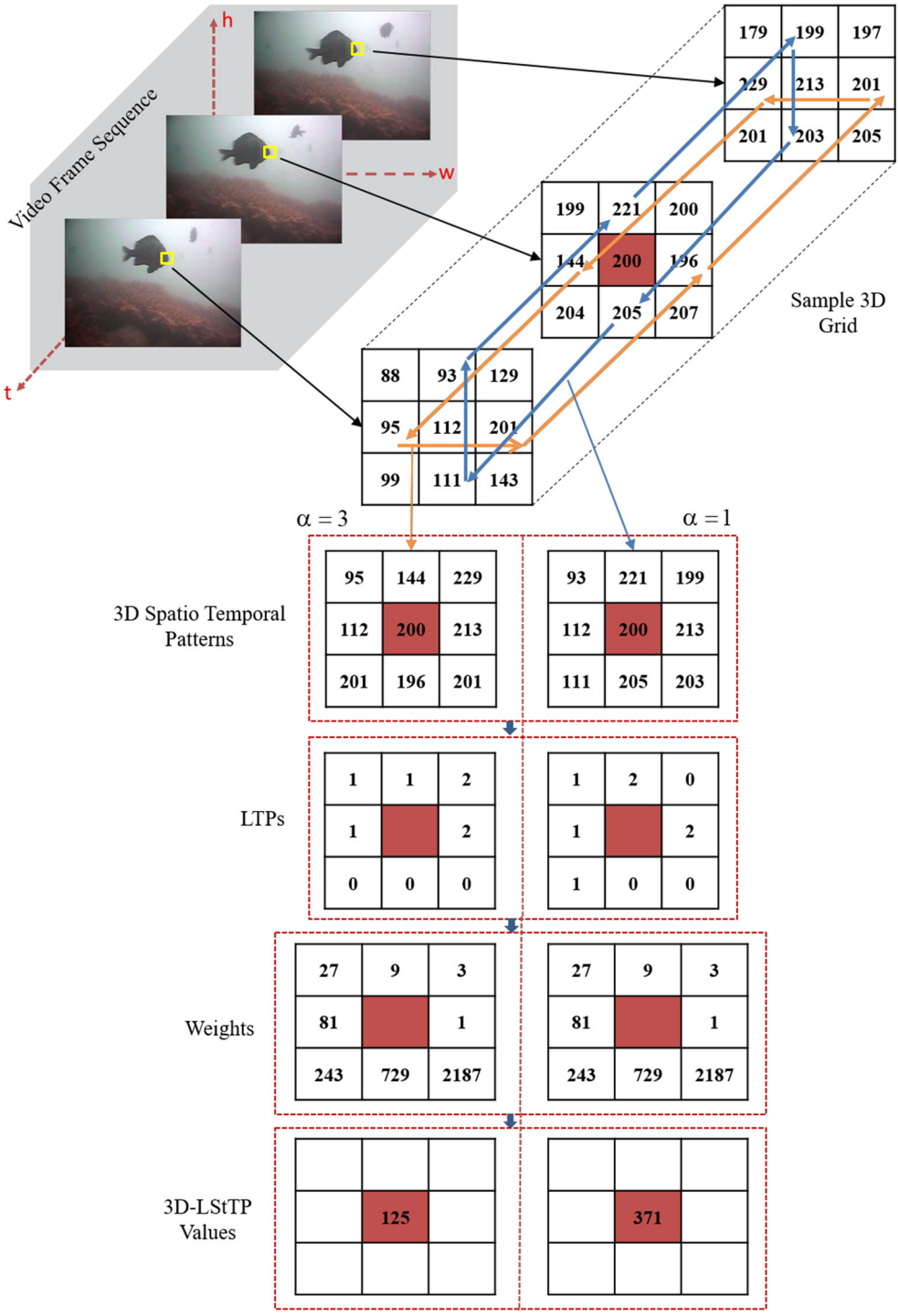

Fig. 3 Example to obtain the 3D-LStTP in two selected directions

where,

$$\gamma = \begin{cases} \frac{1}{\frac{U}{e^{\ln \ln(U) \frac{t-U}{U-1}}}} & if\ 1 \leq t < U \\ \frac{1}{U} & if\ \mathrm{t} \geq U \end{cases} \quad (15)$$

where, Z_b^t is the temporary background image, t is the index for current frame, $Z_{\inf}$ is the incoming frame, γ is the updating rate and U is the updating time window size.

For every incoming video frame, a colour difference value is calculated between each small blocks of the

frame and corresponding small blocks of the temporary background image. Average of the difference values is calculated for three colour channels individually followed by combing the average values of each colour component. Let N_s be the number of pixels present in the small block, then average colour difference value is computed as:

$$I^{\Delta} = \sum_{i=1}^{N_s} \left(R_b^{\Delta}(i) - R_{\text{inf}}^{\Delta}(i) \right); \quad \Delta \in \{L*, a*, b*\} \tag{16}$$

$$I = \frac{1}{255 \times 255 \times 3} \sum_{\Delta=\{L*,a*,b*\}} \left(\frac{I^{\Delta}}{N_s} \right)^2 \tag{17}$$

where I^{Δ} is the summation of the colour channel difference between the small blocks of the temporary background image and incoming frame. Where R_b^{Δ} is the colour channel intensity value of the i^{th} pixel in the small block of the temporary background image, R_{inf}^{Δ} is the colour channel intensity value of the i^{th} pixel in the small block of the incoming frame and Δ represents Lab colour model. I is the final colour information difference of the small blocks of the temporary background image and the incoming video frame. It has been converted into binary image containing foreground moving objects by selecting a threshold. The final foreground image is generated by integrating binary images obtained after background modelling using texture and colour features.

The steps described above can be summarised in Algorithm as follows:

Algorithm 1

1: Load the image sequences.
2: Three consecutive frames are selected for 3D grid structure formation.
3: Produce the 3D volume for 3D-LStTP calculation.
4: Collect 3D spatio-temporal pattern from three dimensional grid in two selected directions.
5: Construct LTPs from 3D spatio-temporal patterns using upper and lower threshold values.
6: Calculate 3D-LStTP by applying weights on LTPs.
7: Construct histogram for 3D-LStTP features in 0° and 90°.
8: Build a feature vector by concatenating the histograms.
9: Colour features are extracted for each small block.
10: Collect separate background models using 3D-LStTP texture and colour features.
11: Integration of 3D-LStTP and colour background models leads to final segmented result.

Experimental Results and Discussions

In this work, the proposed method is tested on three different benchmark datasets (Fish4Knowledge, I2R and Change Detection datasets). The proposed algorithm is implemented in MATLAB R2013a, 64-bit Windows8 platform with Intel Xenon CPU@2.80 GHz and 16 GB of RAM. The performance of the proposed method is analysed by comparing the results with existing state-of-the-art approaches: *TBGS*, Texture BGS; *MBGS*, Multi-Layer BGS; *MCBGS*, Multi-Cue BGS; *SSENSE*, SubSENSE; *FI*, Fuzzy Integral; *PM*, Proposed Method. In order to avoid any implementation bias, the results of the state-of-the-art methods are from the BGSLibrary [25]. The quantitative performance of the proposed method is carried out with the following equations:

$$Precision = \frac{TP}{TP + FP} \tag{18}$$

$$TPR/Recall = \frac{TP}{TP + FN} \tag{19}$$

$$F - measure = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{20}$$

$$FPR = \frac{FP}{FP + TN} \tag{21}$$

where *TP* (true positive) is the number of pixels that are correctly labelled as foreground, *FP* (false positive) is the number of pixels that are incorrectly labelled as foreground, *TN* (true negative) is the number of pixels that are correctly labelled as background and *FN* (false negative) is the number of pixels that are incorrectly labelled as background.

Fish4Knowledge Dataset

The dataset [12] is classified into five categories (*Blurred, Camouflage, Complex background, Dynamic background and Luminosity*) with spatial resolution ranging from 320 × 240 to 640 × 480 pixels. The parameters, τ = 0.05, size of big block = 9 × 9 pixels, size of small block = 3 × 3 pixels, α = 0.005, N_{bb} = 6560, Th = 0.7, U = 50 and T =0.1, are used for experimentation. Table 1 illustrates the comparison of the proposed method with the state-of-the-art methods (TBGS [8], MBGS [35], MCBGS [20] and SSENSE [27]) in terms of F-measure values on *Fish4Knowledge Dataset*.

It can be observed from the results (as shown in Fig. 4.) that complete object contours are detected for luminosity and blurred sequences using proposed approach. High true positive rate (TP) indicates good performance of the proposed method in all the categories except dynamic background case (as shown in Table 1). TBGS is a background

Table 1 Performance of various techniques in terms of F-measure on Fish4Knowledge dataset

Category	TBGS	MBGS	MCBGS	SSENSE	PM
Luminosity	0.56	0.9	0.85	0.82	*0.91*
Blurred	0.66	0.68	0.86	0.59	*0.93*
Complex background	0.69	0.58	0.48	0.21	*0.8*
Camouflage	0.42	0.66	0.77	0.42	*0.83*
Dynamic	0.43	0.32	0.33	*0.81*	0.76

The best performance for each category is given in italics

modelling technique by utilising LBP as a texture feature. The algorithm fails in uniform region and small changes in intensity value of the given pixel rise to different LBP code. Moreover, LBP histogram is not capable to perfectly model the non-stationary scenes.

The detection results in case of dynamic background are not better, due to inclusion of background in foreground objects. This resulted in poor performance of algorithm with high false positive (FP) rates. Whereas, a non-parametric approach, i.e. SSENSE, efficiently handles the dynamic background environment. However, SSENSE approach in the complex background scenes has not observed objects resulting in high false negative (FN) values which could significantly lower the algorithm performance.

MBGS is a combination of texture and colour features which enhance the background modelling performance as compared to TBGS. However, MBGS approach is not efficient for precise moving object segmentation due to extraction of texture features based on the LBP operator which is exploited by image noise in uniform regions. MCBGS technique had low F-measure values in complex and dynamic background scenes, as it fails to detect the objects in foreground and background possessing similar colour and texture.

Further, the performance of the proposed method is also analysed with ROC curve as shown in Fig. 6. From the results presented in Table 1 and Figs. 4, 5 and 6, it is clearly observed that the proposed method outperforms the existing methods on Fish4Knowledge dataset.

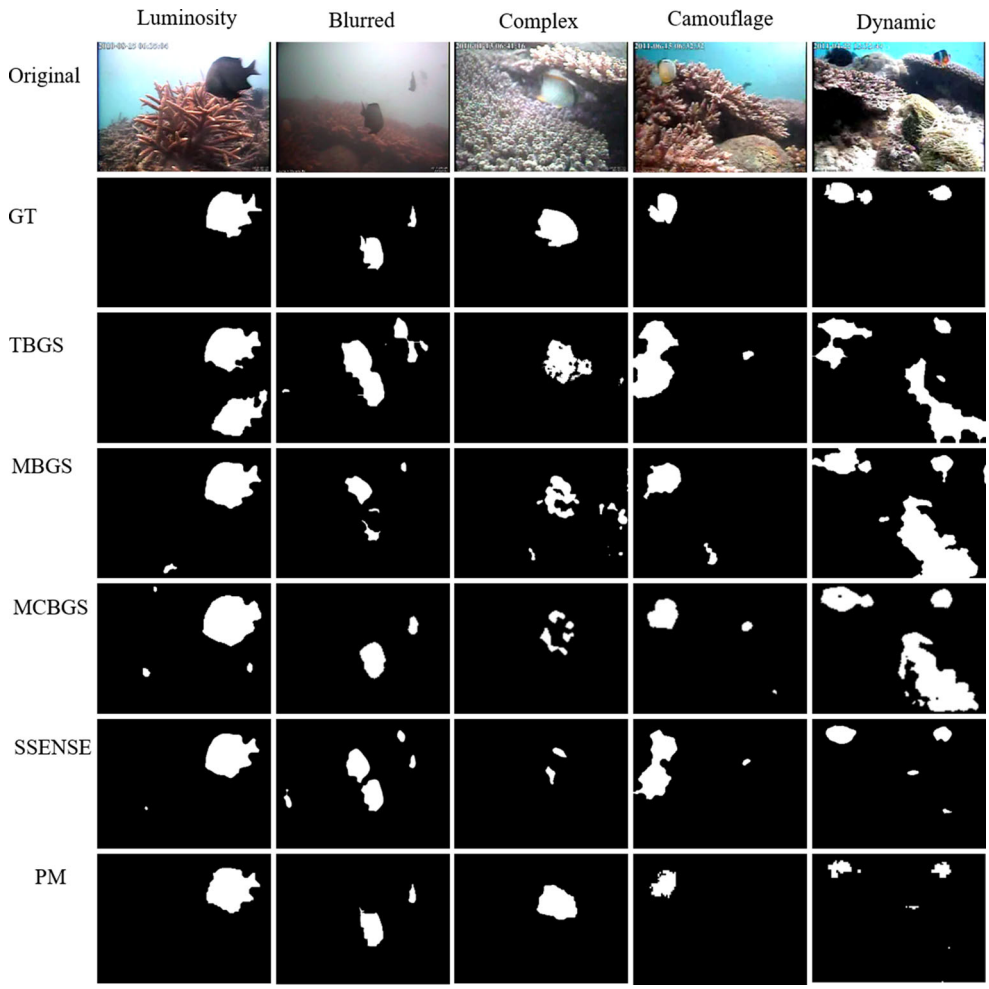

Fig. 4 Qualitative comparison of the proposed technique with state-of-the-art methods on Fish4Knowledge

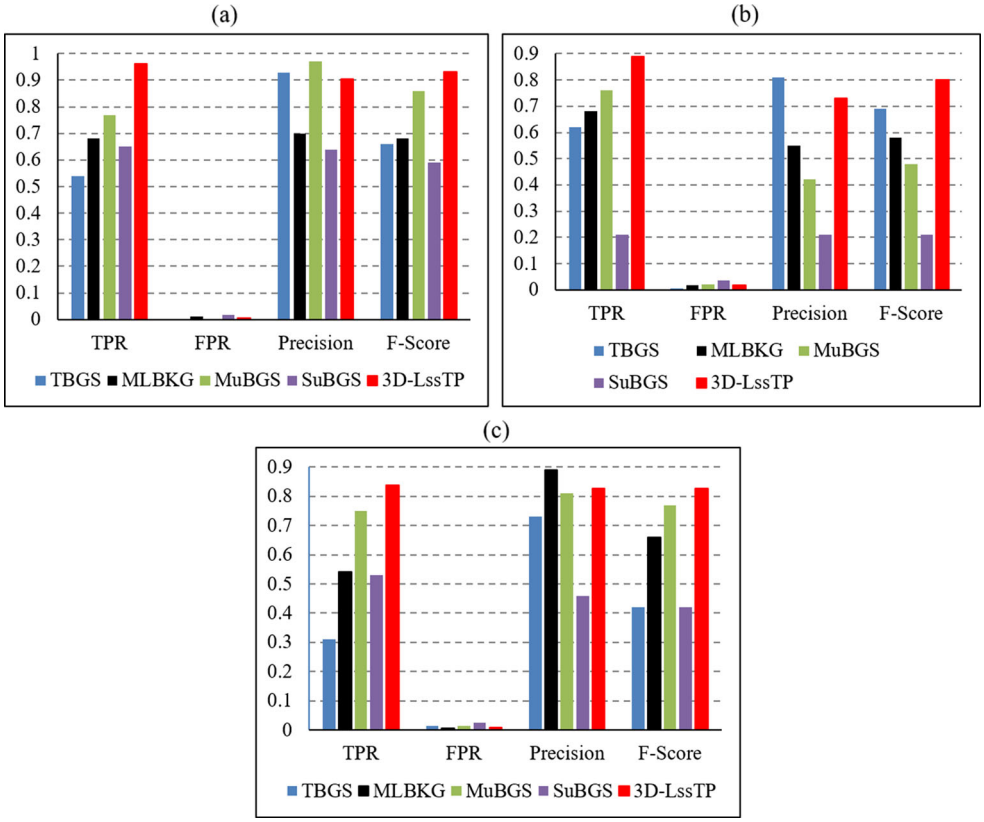

Fig. 5 The bar charts of TFR, FPR, Precision and F-Score metrics for **a** Blurred, **b** Complex Background and **c** Camouflage Background sequences of Fish4Knowledge dataset

Change Detection Dataset

The dataset [4] consists of four video sequences (*Office, Fall, canoe and fountain02*) are considered for experimentation. Figure 7 illustrates the qualitative comparison of the proposed method with existing techniques on *change detection dataset*. Table 2 illustrates the comparison of the proposed method with the state-of-the-art methods (Pfinder [34], TBGS [8], FI [3] and MBGS [35]) in terms of F-measure values on Change Detection Dataset.

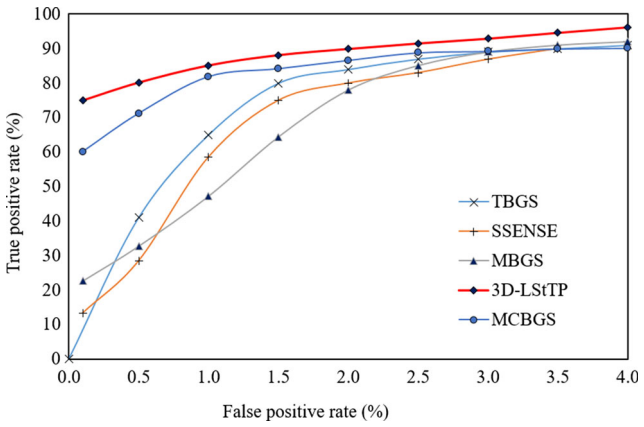

Fig. 6 Comparison of various algorithms in ROC for blurred

First video sequence (Office) depicts indoor scene which shows objects in dynamic and stationary positions with conjugative time intervals. Other three sequences contain outdoor scenes exhibiting dynamic background motion. The objective of the experiments is to illustrate the ability of the proposed method to deal with complex dynamic scenes.

Pfinder method models the background using single Gaussian probability density function (pdf). Single Gaussian pdf is not suitable for dynamic scenes. It performed poorly for Canoe and Fountain02 categories and achieved comparable F-measure values in Office and Fall categories.

In FI technique, foreground detection is made using the Choquet integral by aggregating colour and texture similarity measures. It performed well in Office, Fall and Canoe categories, but cannot able to detect most of foreground regions leading to very poor F-measure value for Fountain02 video sequence. TBGS method displayed comparable performance in all categories with fewer false positives (FP) rates. MBGS technique resulted with poor F-measure value in all the categories due to more number of false positives (FP) and less true positives (TP) values except in Fountain02 category.

From Table 2 and Fig. 7, it is evident that the proposed method outperforms the state-of-the-art methods on Change Detection Dataset.

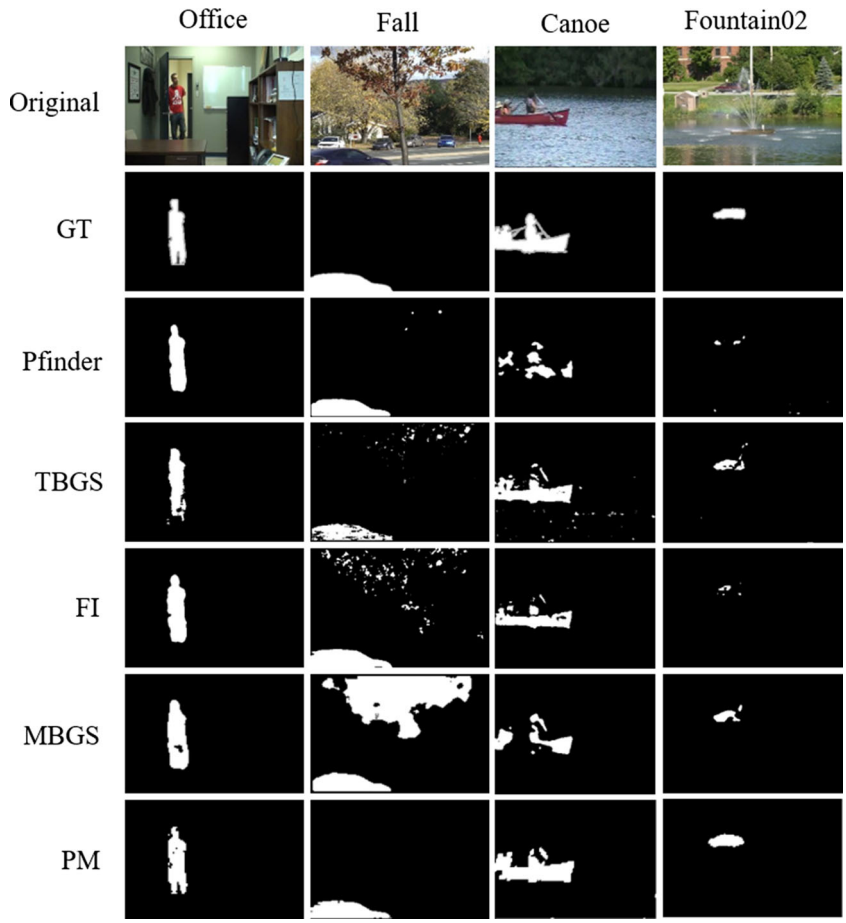

Fig. 7 Qualitative comparison of the proposed technique with state-of-the-art methods on Change Detection dataset

I2R Dataset

In this dataset [16], four dynamic texture video sequences (*Campus, Curtain, Lobby and Fountain*) are considered for experimentation. Figure 8 illustrates the qualitative comparison of the proposed method with existing methodologies on *Campus, Curtain and Fountain* video sequences. Table 3 illustrates the comparison of the proposed method with the state-of-the-art methods (TBGS, MBGS, MCBGS, FI, ISC [23], CS-SILTP [18], SILTP [17], ST-SILTP [9]) in terms of F-measure values on I2R Dataset.

The Fountain sequence contains people moving against a background of a fountain with varying illumination. The foreground objects in the dataset are stationary for a particular time period. In pixel information-based methods such as TBGS, MBGS, FI and ISC except MCBGS, most of the foreground information is lost as a result of pixel-level processing. In contrast, block-based techniques (CS-SILTP, SILTP, ST-SILTP and the proposed method) are able to extract most of the foreground objects, due to the consideration of neighbouring pixel information.

In Campus sequence, proposed and ST-SILTP techniques are sensitive enough to detect small foreground movements against dynamic backgrounds in an improved manner in comparison to others. TBGS and ISC methods fails to detect the foreground objects, whereas MBGS and SILTP are able to extract the smaller objects against dynamic backgrounds with some FP values. MCBGS and FI techniques showed comparable performance with other techniques.

In Curtain video sequence, waving curtains are the main disturbing factor in the background. The MCBGS, FI and CS-SILTP methods are able to extract foreground object with having few FP values, causing a small F-score. TBGS and MBGS output contains small holes inside the foreground object shown in Fig. 8, whereas ST-SILTP, SILTP and our proposed methods are able to segment the entire foreground object resulted in higher F-measure values.

Table 2 F-Measure values of various techniques on Change Detection dataset

Category	Pfinder	TBGS	FI	MBGS	PM
Office	0.8	0.72	0.76	0.74	*0.83*
Fall	0.78	0.79	0.74	0.71	*0.82*
Canoe	0.47	0.7	0.71	0.53	*0.73*
Fountain02	0.53	0.75	0.44	0.8	*0.81*

The best performance for each category is given in italics

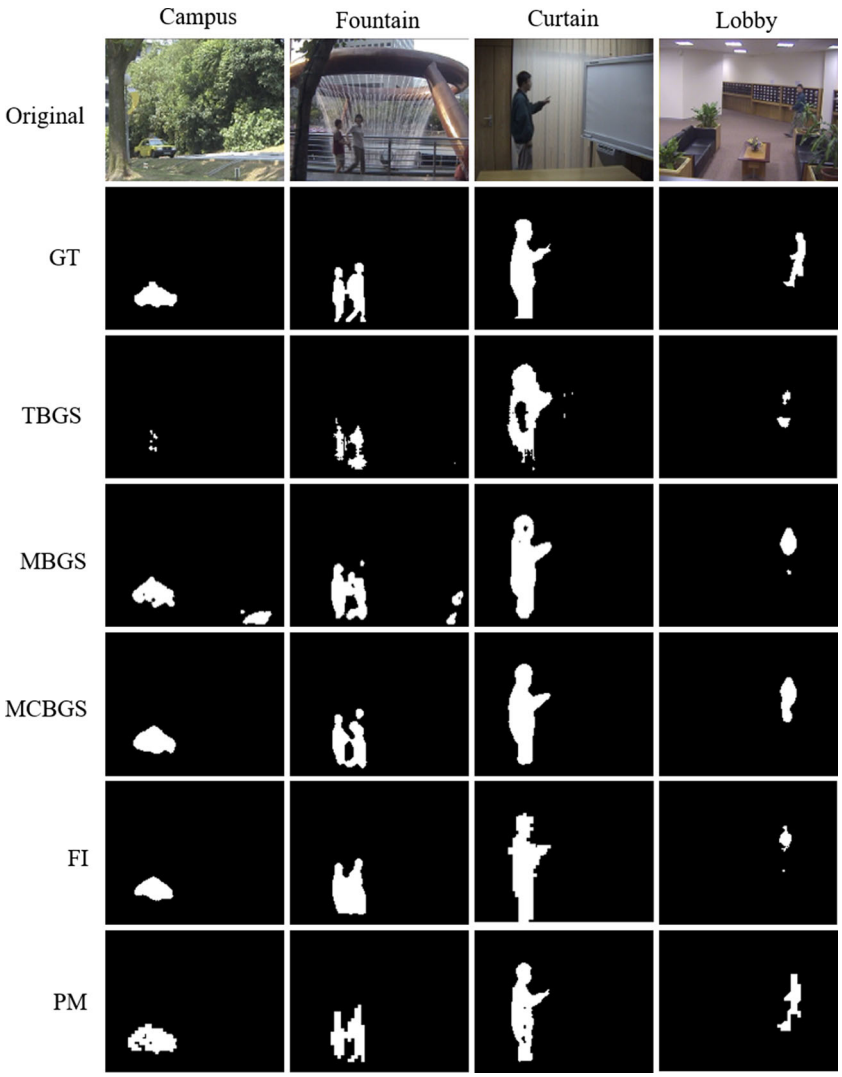

Fig. 8 Qualitative comparison of the proposed technique with state-of-the-art methods on I2R dataset

Lobby video frame is an indoor scene with non-uniform illumination. TBGS, MBGS, MCBGS, FI and ISC are pixel information-based techniques which cannot adapt to global illumination variations. CS-SILTP, SILTP, ST-SLITP and proposed method are block-based techniques which are observed to be more stable in similar cases.

Table 3 F-Measure values of various techniques on I2R dataset

Category	Fountain	Campus	Curtain	Lobby
TBGS	0.77	0.68	0.7	0.62
MBGS	0.74	0.75	0.75	0.65
MCBGS	0.82	0.79	0.76	0.67
FI	0.71	0.8	0.68	0.72
ISC	0.71	0.28	0.29	0.57
CS-SILTP	0.85	0.75	0.74	0.77
SILTP	0.85	0.68	0.92	0.79
ST-SILTP	0.86	0.85	*0.94*	0.86
PM	*0.87*	*0.86*	0.88	*0.87*

The best performance for each category is given in italics

Campus and Curtain video sequences consist of large-area dynamic background. The foreground connectivity gives rise to negative effect in ISC technique. A misconception of moving object is observed by assuming background as foreground objects. Therefore, the ISC method produced smaller F-measure values as shown in Table 3.

From Table 3 and Fig. 8, it is clear that the proposed technique performed well on Fountain, Campus and Lobby video sequences and comparable on Curtain video sequence by comparing with the state-of-the-art methods.

Computational Efficiency

The total algorithm complexity lies in texture and colour computation. Texture feature computation is $O(RCb)$, where R and C are the numbers of big blocks row- and

Table 4 Average processing time per frame on Fish4Knowledge dataset

Methods	TBGS	MBGS	MCBGS	SSENSE	PM
Time (s)	0.658	1.195	1.153	0.158	0.721

columnwise, and b is the size of the big block. Whereas colour feature computation is $O(rcs)$, where r and c are the numbers of small blocks row- and columnwise, and s is the size of the small block. The computation complexity of the colour feature is very less as compared to that of texture feature. So, the colour feature complexity was neglected while calculating entire image computation.

For the entire video, computation is $O(TXYRCb)$, where T is the number of frames and $X \times Y$ is the frame size. The average processing time per frame of the proposed method in comparison to the state-of-the-art techniques is indicated in Table 4. From Table 4, we observed that our approach's average processing time per frame is less as compared to the existing methods of MBGS and MultiCueBGS and more as compared to the existing methods of Texture-BGS and SubSENSEBGS.

Conclusion

In this work, a new 3D-LStTP feature descriptor has been proposed. In 3D-LStTP descriptor calculation, collection of the neighbours is based on local spatio-temporal information from three consecutive video frames by forming a 3D grid in a particular direction for a given centre pixel. The background models are constructed using 3D-LStTP texture and colour features. The segmented results of texture and colour features after background modelling are combined for detection of moving objects in the presence of various extreme conditions. The effectiveness of the proposed method is tested by conducting five experiments on Fish4Knowledge dataset, four experiments on I2R dataset and four experiments on Change Detection dataset. The performance of proposed method outperforms that of state-of-the-art methods qualitatively and quantitatively under various complex scenarios. The processing time of the proposed method can be reduced by optimising the parameters and implementing the code on real-time platforms like OpenCV.

In future work, aim will be on deep learning techniques for moving object segmentation and also improving our proposed method to handle the similar appearance of foreground and background objects. An approach for adaptively adjusting the weights of texture and colour features may boost the segmentation results. Further, making the proposed method computationally efficient for real-time application also needs to be investigated.

Acknowledgements The authors would like to express their sincere thanks to the funding agency Council for Scientific and Industrial Research, India under Network Project (ESC0113) for supporting this work. The authors are grateful to Dr. Maia Hoeberechts and team for providing the Ocean Networks Canada Dataset.

Funding Information This study was funded by Council for Scientific and Industrial Research, India under Network Project (grant/project number: ESC0113).

Compliance with Ethical Standards

Conflict of interests The authors declare that they have no conflict of interest.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

Informed consent Informed consent is not necessary for the present study.

References

1. Akilan T, Wu QJ, Yang Y. Fusion-based foreground enhancement for background subtraction using multivariate multi-model gaussian distribution. Inf Sci. 2018;430:414–31.
2. Bouwmans T, El Baf F, Vachon B. Background modeling using mixture of gaussians for foreground detection-a survey. Recent Patents on Computer Science. 2008;1(3):219–37.
3. El Baf F, Bouwmans T, Vachon B. Fuzzy integral for moving object detection. In: IEEE international conference on Fuzzy systems, 2008. FUZZ-IEEE 2008.(IEEE world congress on computational intelligence). IEEE; 2008. p. 1729–36.
4. Goyette N, Jodoin PM, Porikli F, Konrad J, Ishwar P. Changedetection. net: a new change detection benchmark dataset. In: 2012 IEEE computer society conference on computer vision and pattern recognition workshops (CVPRW). IEEE; 2012. p. 1–8.
5. Guo JM, Liu YF, Hsia CH, Shih MH, Hsu CS. Hierarchical method for foreground detection using codebook model. IEEE Transactions on Circuits and Systems for Video Technology. 2011;21(6):804–15.
6. Han B, Davis LS. Density-based multifeature background subtraction with support vector machine. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2012;34(5):1017–23.
7. Hati KK, Sa PK, Majhi B. Intensity range based background subtraction for effective object detection. IEEE Signal Processing Letters. 2013;20(8):759–62.
8. Heikkila M, Pietikainen M. A texture-based method for modeling the background and detecting moving objects. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2006;28(4):657–62.
9. Ji Z, Wang W. Detect foreground objects via adaptive fusing model in a hybrid feature space. Pattern Recogn. 2014;47(9): 2952–61.
10. KaewTraKulPong P, Bowden R. An improved adaptive background mixture model for real-time tracking with shadow detection. In: Video-based surveillance systems. Boston: Springer; 2002. p. 135–144.
11. Karasulu B, Korukoglu S. Moving object detection and tracking by using annealed background subtraction method in videos: performance optimization. Expert Syst Appl. 2012;39(1):33–43.

12. Kavasidis I, Palazzo S, Di Salvo R, Giordano D, Spampinato C. An innovative web-based collaborative platform for video annotation. Multimedia Tools and Applications. 2014;70(1):413–32.
13. Kim K, Chalidabhongse TH, Harwood D, Davis L. Real-time foreground–background segmentation using codebook model. Real-time imaging. 2005;11(3):172–85.
14. Kim W, Kim Y. Background subtraction using illumination-invariant structural complexity. IEEE Signal Processing Letters. 2016;23(5):634–8.
15. Lee DS. Effective gaussian mixture learning for video background subtraction. IEEE transactions on Pattern Analysis And Machine Intelligence. 2005;27(5):827–32.
16. Li L, Huang W, Gu IY, Tian Q. Foreground object detection from videos containing complex background. In: Proceedings of the eleventh ACM international conference on multimedia. ACM; 2003. p. 2–10.
17. Liao S, Zhao G, Kellokumpu V, Pietikäinen M, Li SZ. Modeling pixel process with scale invariant local patterns for background subtraction in complex scenes. In: 2010 IEEE conference on computer vision and pattern recognition (CVPR). IEEE; 2010. p. 1301–06.
18. Lin L, Xu Y, Liang X, Lai J. Complex background subtraction by pursuing dynamic spatio-temporal models. IEEE Trans Image Process. 2014;23(7):3191–202.
19. Nanni L, Brahnam S, Lumini A. Local ternary patterns from three orthogonal planes for human action classification. Expert Syst Appl. 2011;38(5):5125–28.
20. Noh S, Jeon M. A new framework for background subtraction using multiple cues. In: Asian conference on computer vision. Berlin, Heidelberg: Springer; 2012. p. 493–506.
21. Ojala T, Pietikäinen M, Harwood D. A comparative study of texture measures with classification based on featured distributions. Pattern Recognit. 1996;29(1):51–9.
22. Ojala T, Pietikainen M, Maenpaa T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2002;24(7):971–87.
23. Pan J, Li X, Li X, Pang Y. Incrementally detecting moving objects in video with sparsity and connectivity. Cognitive Computation. 2016;8(3):420–8.
24. Sheikh Y, Shah M. Bayesian modeling of dynamic scenes for object detection. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2005;27(11):1778–92.
25. Sobral A, Vacavant A. A comprehensive review of background subtraction algorithms evaluated with synthetic and real videos. Comput Vis Image Underst. 2014;122:4–21.
26. Spampinato C, Palazzo S, Kavasidis I. A texton-based kernel density estimation approach for background modeling under extreme conditions. Comput Vis Image Underst. 2014;122:74–83.
27. St-Charles PL, Bilodeau GA, Bergevin R. Flexible background subtraction with self-balanced local sensitivity. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops; 2014. p. 408–13.
28. Stauffer C, Grimson WEL. Adaptive background mixture models for real-time tracking. In: IEEE computer society conference on computer vision and pattern recognition, 1999. vol 2, IEEE; 1999. p. 246–52.
29. Tan X, Triggs B. Enhanced local texture feature sets for face recognition under difficult lighting conditions. IEEE Trans Image Process. 2010;19(6):1635–50.
30. Tu Z, Abel A, Zhang L, Luo B, Hussain A. A new spatio-temporal saliency-based video object segmentation. Cognitive Computation. 2016;8(4):629–47.
31. Tu Z, Zheng A, Yang E, Luo B, Hussain A. A biologically inspired vision-based approach for detecting multiple moving objects in complex outdoor scenes. Cognitive Computation. 2015;7(5):539–51.
32. Vasamsetti S, Setia S, Mittal N, Sardana HK, Babbar G. Automatic underwater moving object detection using multi-feature integration framework in complex backgrounds. IET Computer Vision. 2018.
33. Wang Z, Liao K, Xiong J, Zhang Q. Moving object detection based on temporal information. IEEE Signal Processing Letters. 2014;21(11):1403–07.
34. Wren CR, Azarbayejani A, Darrell T, Pentland AP. Pfinder: Real-time tracking of the human body. IEEE Transactions on Pattern Analysis and Machine Intelligence. 1997;19(7):780–85.
35. Yao J, Odobez JM. Multi-layer background subtraction based on color and texture. In: IEEE conference on computer vision and pattern recognition, 2007. CVPR'07. IEEE; 2007. p. 1–8.
36. Yubing T, Cheikh FA, Guraya FFE, Konik H, Trémeau A. A spatiotemporal saliency model for video surveillance. Cognitive Computation. 2011;3(1):241–63.
37. Zhang S, Li N, Cheng X, Wu Z. Adaptive object detection by implicit sub-class sharing features. Signal Proc. 2013;93(6):1458–70.
38. Zhang Y, Wang X, Qu B. Three-frame difference algorithm research based on mathematical morphology. Procedia Engineering. 2012;29:2705–9.
39. Zheng A, Xu M, Luo B, Zhou Z, Li C. Class: collaborative low-rank and sparse separation for moving object detection. Cognitive Computation. 2017;9(2):180–93.
40. Zivkovic Z, Van Der Heijden F. Efficient adaptive density estimation per image pixel for the task of background subtraction. Pattern Recogn Lett. 2006;27(7):773–80.