Salient Features for Moving Object Detection in Adverse Weather Conditions during Night Time

IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY

Abstract

- Foreground segmentation of moving objects in adverse atmospheric conditions such as fog, rain, low light and dust is a challenging task in computer vision.
- The advantages of thermal infrared imaging at night time under adverse atmospheric conditions have been demonstrated, which are due to the long wavelength.
- However, existing state-of-the-art object detection techniques have not been useful in such scenarios.
- In this paper, we propose an improved background model that utilizes both thermal pixel intensity features and spatial video salient features.
- The proposed spatial video salient features are represented as an Akin-based per-pixel Boolean string over a local region block, and depend on the effect of neighbouring pixels on a centre pixel. The result of this Boolean procedure is referred to as the - '*Akin-based Local Whitening Boolean Pattern (ALWBP),*' which differentiates foreground and background region accurately, even against a cluttered background.

Continue…

- The background model is controlled via
  - (i) the automatic adaptation of parameters such as the decision threshold R_T and, learning parameter L, and
  - (ii) the updating of background samples B_{sample_int} and,- B_{sample_ALWBP} to minimize
    - (a) the effect of the background dynamics of outdoor scenes, and
    - (b) the temperature polarity changes during the maiden appearance of a moving object in thermal frame sequences.

- The performance of this model is evaluated using nine existing standard segmentation performance metrics on our newly created -'*Tripura University Video Dataset at Night time (TU-VDN)'* and on the publicly available CDnet-2014 dataset.
- Our newly created weather-degraded video dataset, namely, TU-VDN, consists of sixty video sequences that represent four atmospheric conditions, namely, low light, dust, rain, and fog.
- The results of a performance comparison with fourteen state-of-the-art detection techniques also demonstrate the high accuracy of the proposed technique.

Background Challenges

- For detecting moving objects, both a visual digital camera and a typical charge-coupled device (CCD) camera have the advantage of high resolution, which renders them more suitable for day time or night time use with a proper lighting setup.
- However, they are ineffective in environments with poor illumination or visibility due to atmospheric conditions because the appearance of objects in the captured images is not as clear as in images that are captured during under normal atmospheric conditions.
- To address the limitations of visual and CCD cameras at night time, many studies have been conducted on methods that detect objects with near/far-infrared (NIR/FIR) based cameras.
  - NIR cameras are robust against darkness, and however, they have a similar drawback to that faced by CCD cameras when the interferences are produced by vehicle headlights. In addition, the attenuation of visual, CCD, and NIR radiation that is produced through atmospheric aerosols is mostly due to their short wavelengths.
  - In contrast, FIR cameras enable robust object detection regardless of the atmospheric conditions because as the spectrum wavelength increases, the effect of bad atmospheric conditions decreases.

Continue…

- However, there have many key issues that are related to object detection at night using an FIR camera, such as the following:
  - **Flat Cluttered Background:** The infrared radiation signal must travel from the target to the camera sensor among adverse atmospheric particles and is attenuated due to scattering; the loss of radiation along the way produces a blurred flat region. In addition, with the thermal sensors, because of large variations in the surface, which includes hot and cool objects such as buildings, vehicles, animals, humans, and light poles, the foreground objects and the background scene become indistinguishable.
  - **Temperature Polarity Changes:** Thermal temperature adjustment during the maiden appearance of a moving object in a video sequence causes illumination-type effects in the background model from the current video frame and, therefore, yields false classifications.
  - **Background Dynamics:** Outdoor scenes are affected by movement in the background, e.g., due to waves or swaying tree leaves.

Contributions

- The primary contributions of this paper are summarized as follows:
  - The paper describes in brief a comprehensive thermal video dataset of outdoor night scenes that are degraded by various adverse weather conditions, such as fog, dust, rain, and low light/poor illumination. This dataset is referred to as Tripura University Video Dataset at Night time (TU-VDN). Researchers can utilize this dataset for testing and ranking of existing and new algorithms for moving object detection.
  - The paper proposes an improved video salient feature-based background model algorithm for detecting moving objects in night videos that were captured under adverse atmospheric conditions, in which thermal intensity information, in addition to spatial information, is fully taken into account.
  - This algorithm can handle key challenging issues in thermal and outdoor adverse atmospheric environments, such as a flat cluttered background, a dynamic background, and thermal temperature polarity changes.
  - The proposed salient-feature-based moving object detection method is successfully applied to our adverse-atmospheric-condition-based thermal night dataset, namely, TU-VDN, and the results demonstrate that it outperforms related state-of-the-art methods in terms of detection performance.
  - The performance of the proposed method is also evaluated on change detection dataset - ‘CDNet 2014’.

Problem Definition

- The thermal infrared radiation signal must travel from the target to the camera detector sensor under adverse weather conditions or through atmospheric particles; therefore, more of the signal can be lost along the way, which produces blurry flat regions.
- The thermal infrared camera produces an image according to the differences in the omitted thermal radiation between an object and the background.
- If the background emits the same amount of thermal radiation as objects, e.g., a cluttered background, the foreground and background regions will be indistinguishable.
- We investigated the performance of a perceptual discrimination salient-feature-based methodology on a flat cluttered background, as shown in Fig. 2 (P.T.O.). The sample frames are collected from our TU-VDN dataset with a flat cluttered background.

Continue…

- The pixel values of the background region in Fig. 2(a) and of the foreground object region in Fig. 2(b) are similar and vary smoothly; hence, the background and foreground true-positive pixel intensity values cannot be properly categorized, thereby resulting in incorrect interpretations.
- The main difficulty that is faced by well-known feature descriptors (LBP, LSBP) on such flat cluttered regions is homogenous neighbouring pixel intensity values.

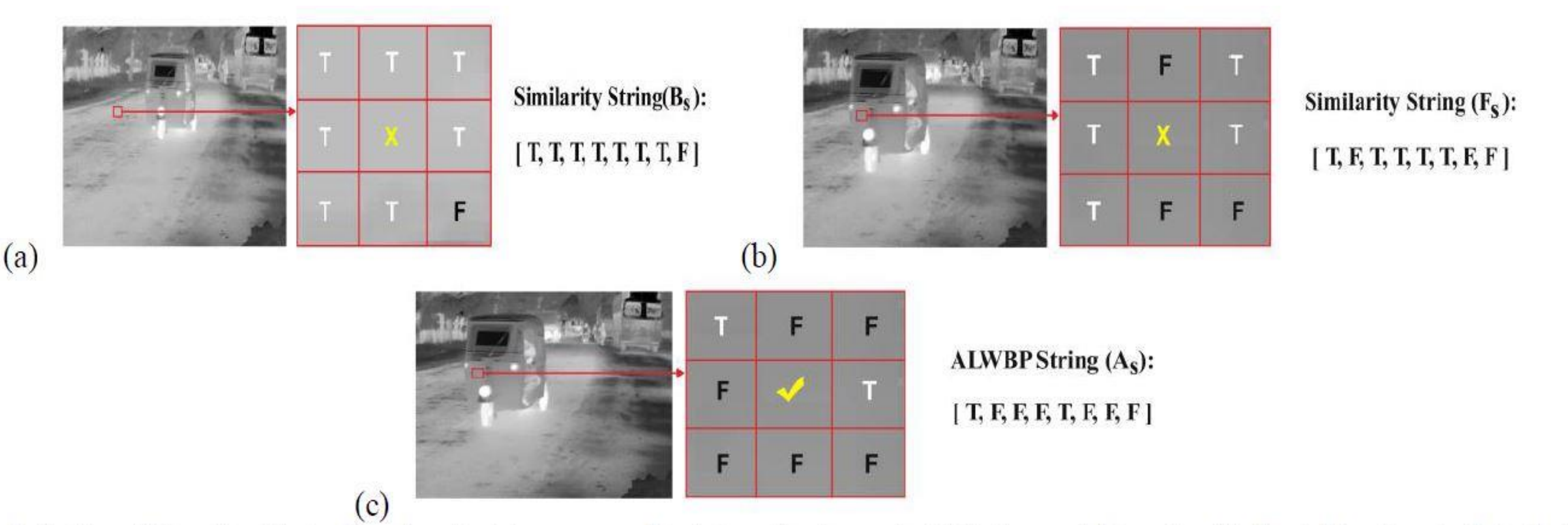


Fig. 2. Outline of the salient-feature-based methodology over a flat cluttered background. (a) Background flat region. Each neighbouring pixel similarity pattern (B_s) is computed using the center pixel (marked as 'x'); (b) Foreground object flat region. The foreground string (F_s) has 6/8 matches with the background similarity string (B_s), which could be categorized as background (incorrectly); (c) Foreground object flat region. The ALWBP descriptor (A_s) is computed using a randomly selected background sample (marked as '√') as a reference center pixel. The foreground string (A_s) has 3/8 matches with the background string (B_s), which is categorized as foreground (correctly).

Continue…

- In Fig. 2(a), we have investigated a background-based local flat region where each neighbouring pixel similarity pattern (B_s) is computed using the centre pixel, where is marked as ‘x’.
- In Fig. 2(b), we have also investigated a foreground-object-based local flat region that is cluttered with the background region. The foreground-region-based similarity pattern (F_s) has 6 matches out of 8 with the background-based similarity pattern (B_s), which could be categorized incorrectly as background.

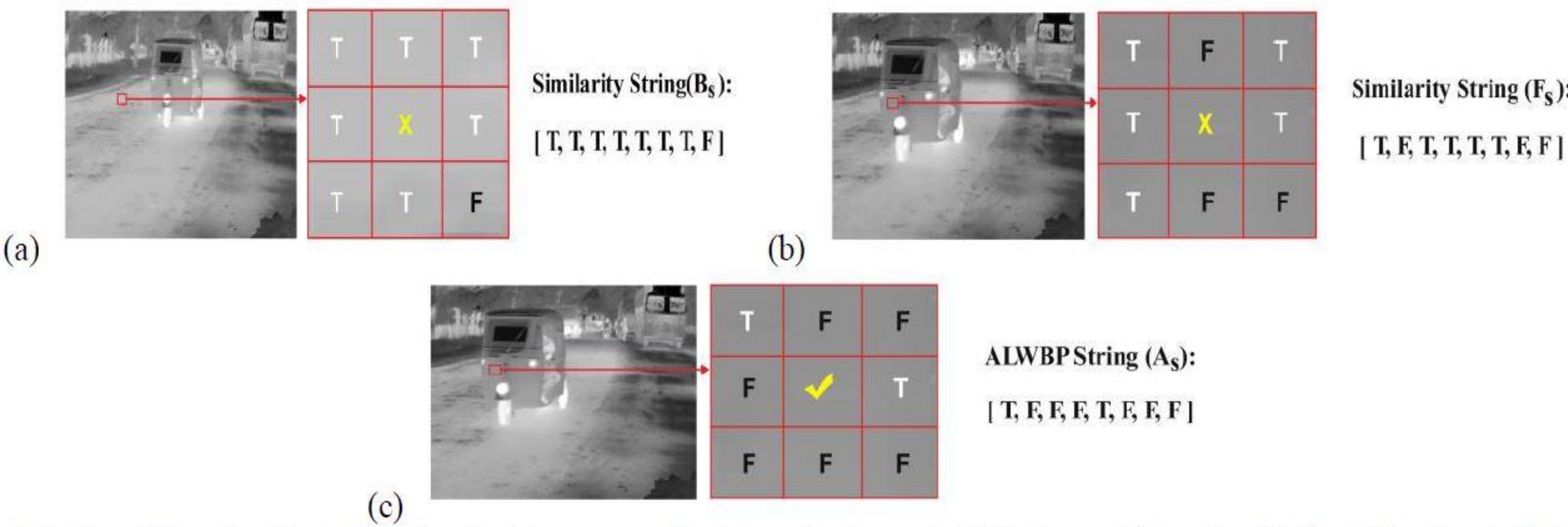


Fig. 2. Outline of the salient-feature-based methodology over a flat cluttered background. (a) Background flat region. Each neighbouring pixel similarity pattern (B_s) is computed using the center pixel (marked as ‘x’); (b) Foreground object flat region. The foreground string (F_s) has 6/8 matches with the background similarity string (B_s), which could be categorized as background (incorrectly); (c) Foreground object flat region. The ALWBP descriptor (A_s) is computed using a randomly selected background sample (marked as ‘√’) as a reference center pixel. The foreground string (A_s) has 3/8 matches with the background string (B_s), which is categorized as foreground (correctly).

Continue…

- We have overcome over this challenge by increasing the robustness of existing local binary feature descriptors, to obtain the ALWBP descriptor.
- In Fig. 2(c), the ALWBP similarity pattern (A_s) is computed using a reference centre pixel, which is marked as '√'. As a result, the foreground similarity pattern (A_s) has 3 matches out of 8 with the background pattern (B_s), which is sufficiently discriminative to be correctly categorized as foreground.

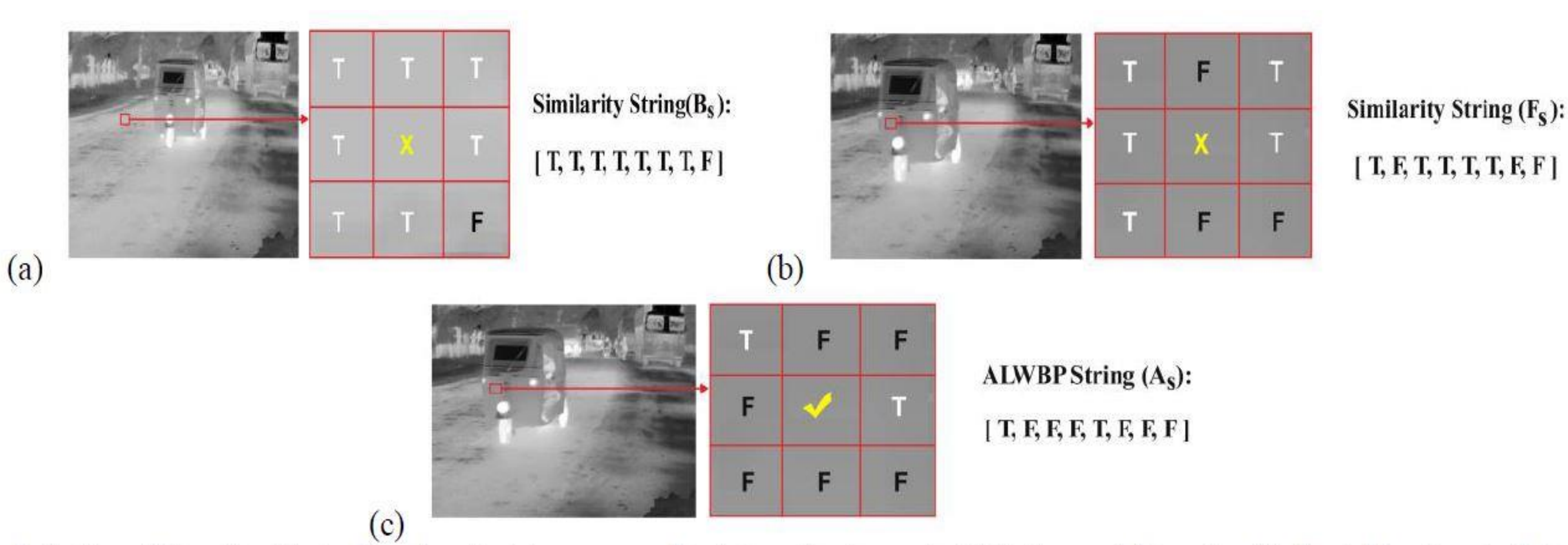


Fig. 2. Outline of the salient-feature-based methodology over a flat cluttered background. (a) Background flat region. Each neighbouring pixel similarity pattern (B_s) is computed using the center pixel (marked as 'x'); (b) Foreground object flat region. The foreground string (F_s) has 6/8 matches with the background similarity string (B_s), which could be categorized as background (incorrectly); (c) Foreground object flat region. The ALWBP descriptor (A_s) is computed using a randomly selected background sample (marked as '√') as a reference center pixel. The foreground string (A_s) has 3/8 matches with the background string (B_s), which is categorized as foreground (correctly).

Proposed Methodology

- The only advantages of thermal cameras are that the captured images are not influenced by illumination and shadows and a pedestrian can be clearly distinguished as a foreground object due to its temperature absorbance.
- Other foreground objects, such as moving vehicles, that are comprised of several body components, such as wheels and headlights are visible, while the remaining components have similar texture to the as background.
- However, finding a satisfactory reference or background model for background subtraction is difficult when there are several real-time objects in thermal frames.
- In this paper, we present a satisfactory background segmentation model that uses the novel Akin-based Local Whitening Boolean Pattern (ALWBP) salient features. It handles flat cluttered regions in thermal frame sequences and increased false-negative ratios.
- The model is inspired by pixel-level and spatiotemporal-level methods because LBP or LBSP features are not robust to flat cluttered regions when neighbouring pixels are similar.
- The overall system pipeline of the proposed background segmentation method is the combination of
  - an ALWBP feature descriptor and
  - a background model generation.

Novel Akin-based Local Whitening Boolean Pattern (ALWBP) salient features

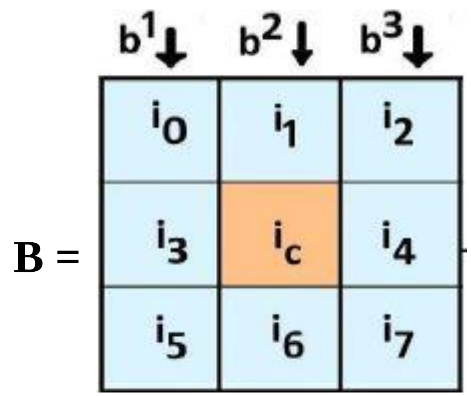
- Existing well-performing and fast local feature descriptors- LBP and LSBP have the following disadvantages:
  - LBP only considers differences between the centre and each neighbouring pixels and
  - LSBP considers the similarity between the centre and each neighbouring pixel, but not the effect of neighbouring pixels on the current similarity between the considered centre and neighbouring pixel.

- These methods are illumination invariant but not robust against low-frequency flat regions and smooth backgrounds or cluttered backgrounds, which has been discussed in the **Problem Definition**.

- These feature descriptors have difficulties on flat cluttered regions due to the homogeneous neighbouring pixel intensity values.

Continue…

- Suppose, we are extracting features on a thermal flat cluttered region block $B \epsilon R^{nxn}$ (B consists of vectors $b^i \epsilon R^n$ for 1<=i<=n) where the values of adjacent pixels are highly correlated. B is a 3x3 block and each column is a set of three pixel values. Each 3x1 column vector is considered as feature vector b^i. Therefore, block contains of three feature samples.



- It is necessary to pre-process each b^i such that the correlation values are lower between adjacent pixels. A very well-known approach is to **whiten** each b^i in the direction of pixel variations that are perpendicular to each other, such that they will have lesser correlation with unit variance.

Continue…

Whitening Over a Local Block B

- To more formally identify the directions of b^1, $b^2,\ldots,b^n$; we compute the matrix covariance, namely, $\sum$, as follows:

$$\Sigma = \frac{1}{n}\sum_{i=1}^{n}(b^i - \bar{b})(b^i - \bar{b})^T \qquad (1)$$

- The eigenvalue decomposition (EVD) can be used to analyse the covariance matrix $\sum$of $B \in R^{nxn}$ as follows:

$$\Sigma = [u_1, u_2, \ldots, u_n][diag(\lambda_1, \lambda_2, \ldots, \lambda_n)][u_1, u_2, \ldots, u_n]^T \qquad (2)$$

where u_1 is the principal vector, namely, the first eigenvector, of $\sum$; u_2 is the second eigenvector; and so on. These vectors are stacked to form an orthogonal matrix, which is denoted as U. Additionally, let λ_1, $\lambda_2,\ldots$, λ_n be the corresponding eigenvalues; they form a diagonal matrix, which is denoted as D. To make our input vectors b^i less correlated with each other, we reflect the original data as follows:

$$b^i_{refl} = U^T b_i \qquad (3)$$

Continue...

- Thus, b^1_{refl}, $b^2_{refl},\dots,b^n_{refl}$ will be less correlated and will satisfy one of our whitening properties. Since U is an orthogonal matrix, it satisfies the property $U^TU=UU^T=I$. Therefore, the reflected vector b^i_{refl} back to original data b^i can be computed via $Ub^i_{refl}=UU^Tb^i=Ib^i=b^i$.
- The unit variance properties of input vectors b^i are imposed by rescaling each reflected vector b^i_{refl} as follows:

$$b^i_{resl} = b^i_{refl} \Big/ \sqrt{\lambda_i+\varepsilon} \qquad (4)$$

- In the scaling step of Eq. (4), a small constant, namely, ε, is added to the eigenvalues to make the feature vectors numerically stable. Altogether, the whitening is defined as follows:

$$B_w = UB_{resl} \qquad (5)$$

$$= U\times\frac{b^i_{refl}}{diag(\sqrt{\lambda_i+\varepsilon})} \qquad \text{using Eq.(4)}$$

$$= U\times diag((\lambda_i+\varepsilon)^{-\frac{1}{2}})\times b^i_{refl}$$

$$= UD^{-\frac{1}{2}}B_{refl} \qquad \{\because\ B_{refl}=[b^1_{refl},b^2_{refl},\dots,b^n_{refl}]\}$$

$$= UD^{-\frac{1}{2}}U^Tb^i \qquad \text{using Eq. (3)}$$

$$= UD^{-\frac{1}{2}}U^T B \qquad (6)$$

Continue...

- The matrix B_w of flat cluttered region block B is white, namely, its vectors $b^1{}_w$, $b^2{}_w$,…,$b^n{}_w$ are less correlated and of unit variance. The covariance of matrix B_w satisfies the following identity property:

$$E\{B_w B_w^T\} = I \tag{7}$$

Continue...

ALWBP Descriptor

- We have obtained a local flat region of pixels via Eq. (6) that are less correlated, and used this region to generate Akin-based local Boolean pattern (ALWBP).
- The term *Akin* indicates a most appropriate similar neighbouring pixel to a centre reference pixel that has more analogous characteristics than the other neighbouring pixels.
- Unlike the traditional LBP and LBSP approaches, which calculates the difference and similarity, respectively, between two pixel values (a centre pixel and a neighbouring pixel), the ALWBP approach considers the effect of other neighbouring pixel values. This Akin-based concept is termed Akinity and is described in Fig. 4.

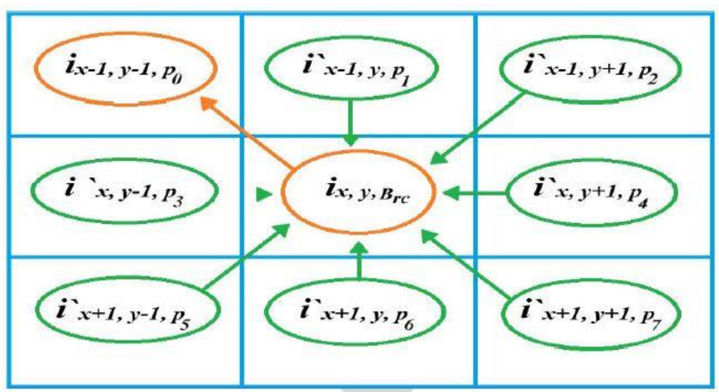


Fig. 4. Akinity $a(i_{x,y,Brc}, i_{x-1,y-1,p0})$ from center pixel $i_{x,y,Brc}$ to candidate Akin pixel $i_{x-1,y-1,p0}$.

Continue...

- B_{rc} is a background intensity sample value at (x,y), which is the reference centre (rc) pixel.
- This differs from the approaches in [SubSense, LOBSTER], where the reference centre pixel is imported from a previous frame intensity value.
- We have altered it because in flat regions, selecting the previous frame reference pixel as centre does not yield substantial discriminative power.
- The value of the reference centre from the background sample is selected randomly from N samples (regarding background samples, wait for Background Model Section).

Continue...

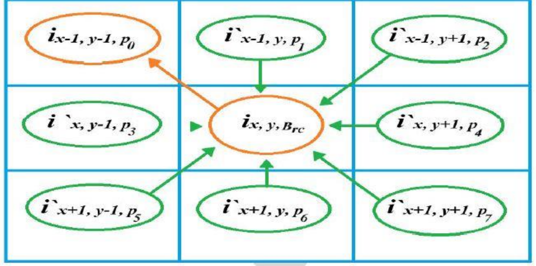


Fig. 4. Akinity $a(i_{x,y,Brc}, i_{x-1,y-1,p0})$ from center pixel $i_{x,y,Brc}$ to candidate Akin pixel $i_{x-1,y-1,p0}$.

Akinity:

- While evaluating a '*candidate Akin neighbouring pixel*' for the '*centre pixel*', we consider other candidate Akin neighbouring pixels as competitors. Fig. 4 shows the Akinity, namely, $a(i_{x,y,Brc}, i_{x-1,y-1,p0})$ from centre pixel $i_{x,y,Brc}$ to a candidate Akin neighbouring pixel $i_{x-1,y-1,p0}$.
- Akin neighbouring pixel $i_{x-1,y-1,p0}$ serves as the most similar candidate for centre pixel $i_{x,y,Brc}$, while other candidate Akin neighbouring pixels i' will compete for centre pixel $i_{x,y,Brc}$.
- Via this approach, we can analyse the similarity between two pixels more intensively than between other neighbouring pixels.
- The Akinity 'a' at location $(i_{x,y,Brc}, i_{x-1,y-1,p0})$ can be calculated via the following formula:

$$
\begin{aligned}
a(i_{x,y,B_{rc}}, i_{x-1,y-1,p_0}) &= sim(i_{x,y,Brc}, i_{x-1,y-1,p_0}) - \min_{i' \neq i}\{sim(i_{x,y,Brc}, i')\} && \text{if} \quad sim(i_{x,y,Brc}, i_{x-1,y-1,p_0}) < T_s \\
&= sim(i_{x,y,Brc}, i_{x-1,y-1,p_0}) + \min_{i' \neq i}\{sim(i_{x,y,Brc}, i')\} && \text{if} \quad sim(i_{x,y,Brc}, i_{x-1,y-1,p_0}) \geq T_s
\end{aligned} \tag{8}
$$

Continue...

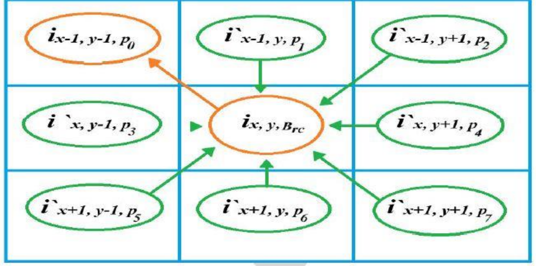


Fig. 4. Akinity $a(i_{x,y,Brc}, i_{x-1,y-1,p0})$ from center pixel $i_{x,y,Brc}$ to candidate Akin pixel $i_{x-1,y-1,p0}$.

Akinity:

- How much higher is the similarity score of a candidate Akin neighbouring pixel $i_{x\text{-}1,y\text{-}1,p0}$ than those of the other competing candidate Akin neighbouring pixels *i'*?
- To answer this, we have subtracted the largest of the similarities among the competing candidate Akin neighbouring pixels *i'* with centre pixel $i_{x,y,Brc}$.
- At this point, we impose a condition: if the similarity between $i_{x,y,Brc}$ and $i_{x\text{-}1,y\text{-}1,p0}$ is less than a similarity threshold, namely, T_s, (which is set to 0.2 in this paper), the value will be subtracted; otherwise it will be added.
- Hence, if there is a more correlated value even after the whitening process, the similarity will be increased, and if there is a slightly uncorrelated value between centre and an Akin neighbouring pixel, the similarity will be decreased.
- In a same manner, the Akinity will be estimated for remaining neighbouring pixels, namely, $p_1, p_2, \ldots, p_7$.

Continue...

Akinity:

- Since the Akinity is estimated among a group of neighboring pixels with a centre point, in some circumstances, replicate values will be obtained, which is called oscillation of numerical values.
- It is important for them to be damped to avoid numerical oscillation.
- Each updated damped Akinity value is set to λ times its previous value plus (1-λ) times its current Akinity value, as follows:

$$a_p(i_{Brc}, i_p) = \lambda \times a(i_{Brc}, i_{p-1}) + (1-\lambda) \times a(i_{Brc}, i_p)\ ;\ 0<\mathrm{p} \le 7 \qquad (9)$$

Continue...

- Now, the ALWBP descriptor Boolean string rule is presented as

$$ALWBP(x, y) = T \quad \text{if } a_p < relative\_tau \;; \quad 0 \le \text{p} \le 7$$
$$= \text{F} \quad \text{Otherwise} \qquad (10)$$

where a_p corresponds to the estimated Akinity value of the p^{th} neighbour of the pixel at *(x,y)* in the current frame and *relative_tau* = $E\text{xi}_c$ is the new energy based threshold value estimate for the current centre pixel at *(x,y)*.

- To capture the micro-texture in a smooth region, the spatial two-dimensional dependence matrix, which is known as the *grey-level co-occurrence matrix* (*G*), of thermal grey palette values is used with displacement vector *d=(dx, dy)*, where *dx=1* and *dy=1*. The feature that measures the randomness of grey-level distribution is the energy, namely, *E*, which is defined using the grey-level co-occurrence matrix as follows:

$$Energy(E) = \sum_x \sum_y G^2[x,y] \qquad (11)$$

Algorithm Summary: Novel ALWBP

Algorithm 1 (ALWBP): Akin-based Local Whitening Boolean Pattern Generation

Input: A frame F, the energy (E) over frame F.

Output: A Boolean pattern string for each pixel in frame F.

1.	**for** x: length(F,1)
1.1	**for** y: length(F,2)
1.1.1	B_{rc} = randomly select a background sample from N recent samples
1.1.2	$relative\_tau = E*F(x,y)$
1.1.3	B = extract a 3x3 block at coordinate (x,y)
1.1.4	Initialize a matrix of size B, $a \leftarrow 0$
1.1.5	Whiten the block B *// to reduce the correlation between adjacent pixels*
1.1.6	$B_w(i_c,i_c) = B_{rc}$ *// center is replaced by a background sample that is used as the reference center pixel*
1.1.7	Estimate *Akinity* 'a' over this whitened block B_w as
	if $sim(i_{Brc},i_p)<T_s$ **then**
1.1.7.1	$a(i_{Brc},i_p) = sim(i_{Brc},i_p)-min\{ i_{Brc},i'_p \}$
1.1.8	**else**
1.1.8.1	$a(i_{Brc},i_p)=sim(i_{Brc},i_p)+min\{ i_{Brc},i'_p \}$
1.1.9	**endif**
1.1.10	Dampen the *Akinity* values to avoid numerical oscillations via
	$a_p(i_{Brc},i_p)= \lambda *a(i_{Brc},i_{p-1}) + (1- \lambda)* a(i_{Brc},i_p)$
1.1.11	**if** $a_p<relative\_tau$ **then**
1.1.11.1	$ALWBP(x,y) = T$
1.1.12	**else**
1.1.12.1	$ALWBP(x,y) = F$
1.1.13	**endif**
1.1.14	**endfor**
1.1.15	**endfor**

Generating the Background Model via ALWBP (BM U ALWBP)

- To generate our non-parametric background model, we represent each background pixel using both spatial-level and pixel-level features, namely, *ALWBP Boolean patterns* and *thermal intensities*.

- To try to match each pixel from the current frame with background integer samples,
  - we first compare the thermal pixel intensity values using the taxicab geometry to a pixel wise dynamic threshold R_T.
  - Second, we compare the ALWBP Boolean patterns over 3x3 blocks on the current frame with background ALWBP samples via a Hamming distance threshold, which is denoted as H_T .

- The methods [SubSense, LOBSTER] are typically not robust against *flat cluttered backgrounds*, whereas our method focuses on this issue, along with other background subtraction problems such as *thermal intensity changes* upon the first appearance of objects and *dynamic backgrounds*.

Continue...

A. Pixel decision via samples of thermal intensity and ALWBP:

- Inspired by [PBAS, VIBE], we develop a random sample consensus framework for modelling both long-term and short-term periodic events.
- Each pixel intensity, namely, *I(x,y)*, is modeled by an array of *N* recently observed background intensity samples, namely, B_{sample_Int} and ALWBP string samples, namely, B_{sample_ALWBP}.

$$B_{sample\_\mathrm{int}}(x,y) = \{BInt_1(x,y), BInt_2(x,y), \ldots, BInt_k(x,y), \ldots, BInt_N(x,y)\} \in \Re^{N\times 1} \quad (12)$$

$$B_{sample\_ALWBP}(x,y) = \{BALWBP_1(x,y), BALWBP_2(x,y), \ldots, BALWBP_k(x,y), \ldots, BALWBP_N(x,y)\} \in S^{N\times 1} \quad (13)$$

- For thermal scenes, *N* must be a small as possible to balance memory consumption and computational complexity (we set *#N=10* in our case).
- Each of these samples is matched against its observation *I(x,y)* or *ALWBP(x,y)* at coordinate *(x,y)* on the current frame for classifying a pixel as foreground (*F(x,y)=1*) or background (*F(x,y)=0*) as follows: (P.T.O.)

Continue...

A. Pixel decision via samples of thermal intensity and ALWBP:

$$F(x,y)=1 \quad \text{if } \{\text{texicab}(\text{I}(x,y),\text{B}_{sample\_\text{int}}(x,y))<\text{R}_T(x,y)$$

$$\&$$

$$\text{XOR}(\text{ALWBP}(x,y),\text{B}_{sample\_ALWBP}(x,y))\leq H_T\} \quad < \quad \text{Threshold}_{\text{min}}$$

$$= 0 \quad \text{otherwise} \qquad (14)$$

F(x,y)=1 corresponds to a per-pixel output segmentation map

$R_T(x,y)$ is the per-pixel distance threshold at pixel *(x,y)*, which should be high for highly dynamic areas and low for static areas

H_T is a fixed Hamming distance threshold (we set $\#H_T=3$)

At last classification, $Threshold_{min}$ is the minimum number of matches with background samples in both the thermal intensity and ALWBP pattern, which is a fixed global parameter (we set $\#Threshold_{min}=2$) that balances the noise resistance.

Continue...

B. Per-pixel adaptation of the distance threshold (R_T):

- A dynamic distance threshold, namely, R_T is defined per-pixel at coordinates *(x,y)*.
- For highly dynamic areas, $R_T(x,y)$ should be high to prevent incorrect classifications as foreground and it should be low for static areas.
- In a video sequence, there can be regions with waving of a water layer or trees in the wind, which will provide higher background dynamics and result in incorrect classifications of foreground objects.
- In addition, there can be regions with small to no changes, which provide low dynamic value.
- Therefore, the background dynamics, namely, $đ_{min}(x,y)$, must be estimated, as inspired by PBAS.

Continue...

B. Per-pixel adaptation of the distance threshold (R_T):

Estimation of Background Dynamics-

- In addition to saving arrays of the *N* recently observed background thermal intensity samples and ALWBP samples in the background maintenance, as in Eq. (12) and (13), we create another array, namely, *D(x,y)* of minimum-distance samples between the current thermal pixel intensity and the background intensity samples as follows:

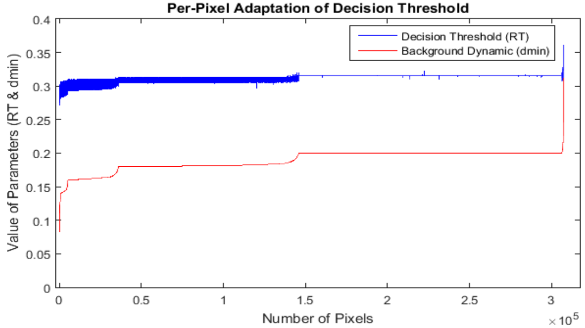
$$D(x, y) = \{D_1(x, y), D_2(x, y), \ldots, D_k(x, y), \ldots, D_N(x, y)\} \quad (15)$$

$$D_k(x, y) = \min\{texicab(I(x, y), B_{sample\_\mathrm{int}}(x, y))\} \quad (16)$$

- To measure the background dynamics at pixel coordinate *(x,y)*, the average of these minimum distance samples is calculated as follows:

$$\bar{d}_{\min}(x, y) = \frac{1}{N} \sum_{k=1}^{N} D_k(x,y) \quad (17)$$

Continue...



B. Per-pixel adaptation of the distance threshold (R_T):

- The dynamic adaptation of distance threshold $R_T(x,y)$ via this measurement of the background dynamics is expressed as follows:

$$R_T(x,y) = (1 - R_{lr}) \times R_T(x,y) + R_{lr} \times \bar{d}_{\min}(x,y) \qquad (18)$$

where R_{lr} is a fixed regulated controller rate for the distance threshold ($R_{lr}=0.02$ in our case).

In completely static regions or less dynamic background regions, namely, $\bar{d}_{min}(x,y)\approx 0$, the value of $R_T(x,y)$ will slowly decrease.

In contrast, under increasing background dynamics, the distance threshold, namely, $R_T(x,y)$, approaches the product value of $R_{lr} x \bar{d}_{min}(x,y)$, which provides a robust, increasing threshold value.

However, in dynamic regions, $R_T(x,y)$ initially slightly decreases by a factor of $(1-R_{lr})$ and subsequently rapidly increases by a factor of R_{lr} as the value of $\bar{d}_{min}(x,y)$ increases.

In above Fig., the decision threshold is plotted.

Continue...

(a) (b)

Fig. 7. Thermal intensity changes upon the first appearance of an object in (a) a thermal frame and (b) the next frame in which the object enters for the first time.

C. Updating the Background Model:

- To account for changes in the background, such as thermal intensity changes upon the first appearance of an object in the frame (as shown in Fig. 7), a waving water layer, and shaking trees, updating the background pixels in the background model, namely, B_{sample_Int}, B_{sample_ALWBP}, is essential.

- We have updated our background model via a similar approach to that in [PBAS].

- A pixel at coordinate (x,y) is updated to one of the background samples if and only if the pixel is categorized as background, namely, $F(x,y)=0$.

- Hence, foreground pixels will be excluded from this update process.

- For a randomly selected index $k \epsilon 1,2,\ldots,N$, the corresponding background sample values, namely, $BInt_k(x,y)$ and $BALWBP_k(x,y)$, are replaced by the current intensity value, namely, $I(x,y)$, and ALWBP pattern, namely, $ALWBP(x,y)$, respectively.

Continue...

C. Updating the Background Model:

- At the same time, we also update a random sample that is selected from 8-neighbouring pixels: $I(x',y') \in N(I(x,y))$.

- The background model at this neighbouring pixel is replaced by its current intensity value, namely, $I(x',y')$, and pattern, namely, $ALWBP(x',y')$.

- Via this neighbouring pixel update process, wrongly classified foreground pixels are gradually incorporated into the background model, as shown in Fig. 8.

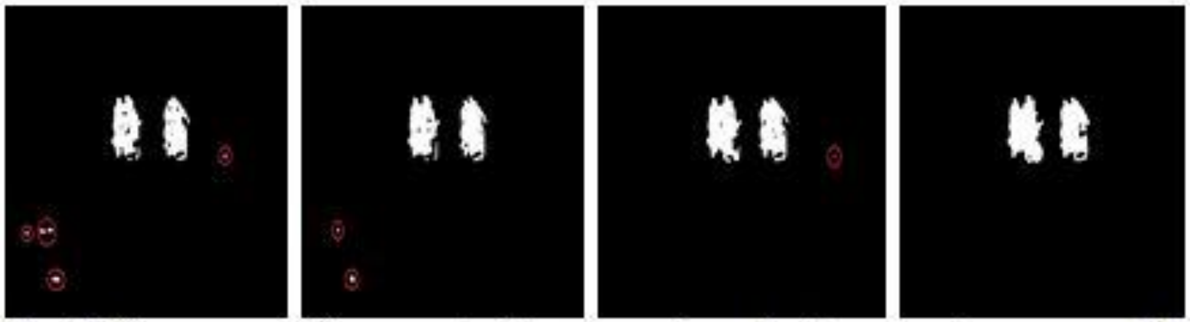


Fig. 8. Sequence of segmented frames, where the incorrectly segmented (marked by red circles) foreground pixels are gradually vanish in subsequent frames.

Continue...

D. Per-pixel adaptation of the learning parameter (L):

- Every pixel, whether foreground or background, that is incorporated into a background sample also depends upon the learning parameter, namely, $L(x,y)$. A higher $L(x,y)$ value indicates that the pixel at (x,y) is more likely to be incorporated into the background model.
- According to the adaptation of the learning parameter $L(x,y)$, pixels those pixels that are wrongly classified as foreground will be merged into background pixels. This strategy is formulated in Eq. (19) as follows:

$$L(x,y) = L(x,y) \times \{(1 - L_{lr} / \bar{d}_{\min}(x,y)) \times F(x,y) + (1 + L_{lr} / \bar{d}_{\min}(x,y)) \times (1 - F(x,y))\} \qquad (19)$$

where L_{lr} is a learning rate (L_{lr}=0.02 in our case).

- The learning parameter of a pixel is decreases fast if the pixel belongs to the foreground, namely, if $F(x,y)=1$, or a plus low dynamic background and slowly decreased in the case of a highly dynamic background. As a result, an incorrectly classified pixel will slowly be identified as background pixel.
- If a pixel belongs to the background, namely, if $F(x,y)=0$, the second term in Eq. (19) (after '+') will increase the learning parameter value by $L_{lr}/đ_{min}(x,y)$. The learning rate will increase based on the value of $đ_{min}(x,y)$. A larger value of $đ_{min}(x,y)$ will slowly increase the learning parameter value, namely, $L(x,y)$, and small value of $đ_{min}(x,y)$ will rapidly increase it.

Algorithm Summary: BM U ALWBP

Algorithm 2 (BM ⋃ ALWBP): Background Model using Akin based Local Whitening Boolean Pattern

Input: Total number N of frame sequences F^i for the generation of the corresponding background model

Output: Corresponding segmented frame sequences F^i

1.	**for** i : N number of frames
2.	Initialization:
	match ← 0
	k ← 0
	R_T ← initialize randomly
	L ← initialize randomly
	$Threshold_{min}$ ← 2
3.	**for** x : length(F^i,1)
3.1	**for** y : length(F^i,2)
3.1.1	**while**(k<=N) **do**
3.1.1.1	**if** [taxicab{I(x,y), $BInt_k$(x,y)}<R_T(x,y) &&… {ALWBP(x,y) ⊕ $BALWBP_k$(x,y)}<H_T(x,y)]
3.1.1.1.1	match = match + 1
3.1.1.2	**endif**
3.1.1.3	k = k + 1
3.1.2	**endwhile**
3.1.3	**if** (match<$Threshold_{min}$) **then**
3.1.3.1	F^i(x,y) = 1
3.1.4	**else**
3.1.4.1	F^i(x,y) = 0
3.1.5	**endif**
3.1.6	Adaptation of distance threshold R_T(x,y) based on dynamic parameter $\bar{d}_{min}(x,y)$
3.1.7	Adaptation of Learning parameter L(x,y) based on dynamic parameter $\bar{d}_{min}(x,y)$ and F^i(x,y)
3.2	**endfor**
4.	**endfor**
5.	**endfor**

Conclusion

- We have described briefly our newly created night video dataset, namely, TU-VDN, for moving object detection in thermal infrared images.

- The dataset consists of degraded atmospheric night outdoor scenes under low-light, dusty, rainy, and foggy conditions.

- We also presented a video salient-feature-based background segmentation technique that uses both spatial features and thermal intensity for the robust investigation of thermal frames.

- We summarize the findings regarding this proposed technique as follows:
  - It handles various key challenges in thermal outdoor scenes, such as dynamic background, flat cluttered background, and thermal intensity adjustment during the maiden appearance of a moving object in the video sequence.
  - In terms of accuracy, F_1-score, and MCC, the results of the comparative experiments on the TU-VDN dataset has demonstrated the superior performance of our proposed method.
  - The results of our analysis on the CDnet-2014 dataset over the night, thermal, and badWeather category sequences have also demonstrated the superior performance of the approach in terms of MCC value and error rate.