

Background Subtraction using Local SVD Binary Pattern

Lili Guo¹, Dan Xu^{*1}, Zhenping Qiang^{1,2}

¹School of Information and Engineering, Yunnan University

²Department of Computer and Information Science, Southwest Forestry University

guoxiaopu@ynu.edu.cn, danxu@ynu.edu.cn, qzp@swfu.edu.cn

Abstract

Background subtraction is a basic problem for change detection in videos and also the first step of high-level computer vision applications. Most background subtraction methods rely on color and texture feature. However, due to illuminations changes in different scenes and affections of noise pixels, those methods often resulted in high false positives in a complex environment. To solve this problem, we propose an adaptive background subtraction model which uses a novel Local SVD Binary Pattern (named LSBP) feature instead of simply depending on color intensity. This feature can describe the potential structure of the local regions in a given image, thus, it can enhance the robustness to illumination variation, noise, and shadows. We use a sample consensus model which is well suited for our LSBP feature. Experimental results on CDnet 2012 dataset demonstrate that our background subtraction method using LSBP feature is more effective than many state-of-the-art methods.

1. Introduction

Segment moving foreground objects from a mostly static background is a fundamental problem in many computer vision tasks such as visual surveillance, traffic control, medical image processing [25], object identification [23] and tracking. Accurate segmentation results can significantly improve the overall performance of the application employing it. Background subtraction is generally regarded as an effective method for extracting the foreground, and it has moved forward from simply comparing a static background frame with current frame to establishing a sophisticated background model of the scene with periodic updates. As we know, illumination variation is one of the major challenges in background subtraction.

Generally, background subtraction is composed of two modules: reference model construction and feature repre-

sentation. The main objective of reference model construction is to obtain an effective and efficient background model for foreground object detection. In the past decade, a very popular background model is to model each pixel with a mixture of Gaussians[20], proposed by Stauffer and Grimson. As further development, more elaborate and recursive update techniques are discussed in [26]. In the ViBe [1] and PBAS [9] presented a sample based classification model that maintained a fixed number of samples for each pixel and classified a new observation as background when it matched with a predefined number of samples. In [5], Elgammal *et al.* proposed kernel density estimate (KDE) technique that has been successfully applied to background subtraction. In [14], Maddalena *et al.* proposed a self-organizing artificial neural network for background subtraction (SOBS). A more detailed discussion of these conventional techniques can be found in recent surveys [3].

The goal of feature representation is to effectively reflect the intrinsic structural properties of scene pixels. Color intensities are commonly used to characterize local pixel representations in pixel-based models. Color features only reflect the visual perception properties of scene pixels, and often ignore the spatial information between adjacent pixels, resulting in the sensitivity to noise and illumination changes.

For introducing the spatial information, a classic method [7] use local binary pattern(LBP) descriptors to handle illumination variation and noise. LBP [7] feature is invariant to local illumination variations such as cast shadow because LBP is obtained by comparing local pixels values. The original LBP operator labels the pixels of an image by thresholding the 3×3 -neighborhood of each pixel with the center value and considering the result as a binary string. It is a powerful mean of texture descriptors. The Center-Symmetric LBP was proposed in [8] to further improve the computational efficiency. In [21], Tan and Triggs extended LBP to LTP (Local Ternary Pattern) by thresholding the graylevel differences with a small value, to enhance the effectiveness on flat image regions. Scale-Invariant Local Ternary Pattern (SILTP) [11] utilizes only one single LBP-

*Dan Xu is the corresponding author.

like pattern as feature, it can be used directly at the pixel level to detect illumination changes. Center Symmetric Spatio-Temporal Local Ternary Pattern (CS-STLTP) [12] is designed to compactly encode the video bricks against illumination variations. Local Binary Similarity Pattern (LBSP) [2] introduced inter and intra LBSP information in background model to enhance the discriminability. Chen *et al.* [4] proposed a powerful and robust local descriptor named the Weber Local Descriptor (WLD) used for texture classification and face detection. Qi *et al.* [16] extend the traditional LBP feature to a pairwise rotation invariant co-occurrence LBP feature used for dynamic texture and scene recognition and dynamic facial expression recognition.

In this paper, we present an efficient background subtraction model using novel Local SVD Binary Pattern feature (named LSBP), it handles illumination variations on the feature level. Our work is motivated by Heikkilä and Pietikäinen [7]. Their method can improve robustness against illumination variations and reduce false classifications caused by camouflaged foreground objects. However, LBP operator is not robust to local image noises when neighboring pixels are similar. In this work, we extend LBP with Local singular value decomposition (SVD) operator. As we know, local binary patterns are not numerical values, they are binary strings. As a result, traditional numerical value based methods, either GMM-like [26] or KDE-like [5], can not be used directly for modeling local patterns into background. Therefore we introduce sample consensus(SACON) model [24] to fit our patterns. This model is well suited for the description of pixels via complex features. In our method, each pixel is modeled by LSBP feature and color intensity separately, and we have verified its effectiveness against illumination variations and noises.

In summary, the main contributions of this paper are: (1) we propose a novel LSBP feature descriptor, it has the ability to gain the potential structures of local regions, it also inhibit the effect of illumination changes especially cast shadows and noise; (2) we introduce an efficient background subtraction model using LSBP and evaluate this model on the CDnet2012 dataset [6]. Experiment results show that our model outperforms several state-of-the-art methods.

2. Local SVD Binary Pattern

Local Binary Pattern (LBP) is proved to be a powerful and fast local image descriptor [15]. It offers an effective way of analyzing textures. The encoding is monotonically invariant to gray scale transforms. However, the LBP operator is not robust to local image noises when neighboring pixels are similar [13, 11]. So we need to introduce other more robustness characteristics to extend LBP.

Singular value decomposition (SVD) is a generalization of the Eigen decomposition which can be used to analyze rectangular matrices (the Eigen-decomposition is defined

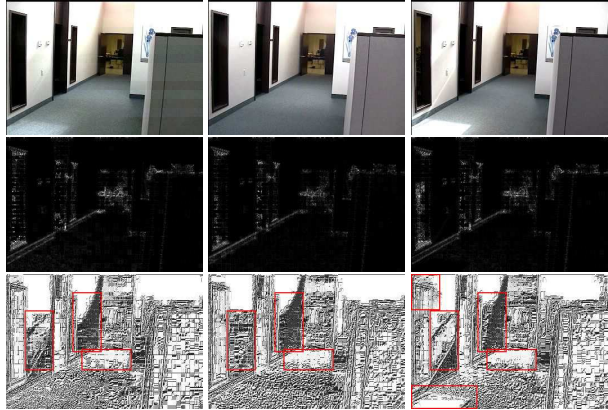


Figure 1: Comparing illumination variations maps. Row (1): input image, row (2): our local structural invariant maps, row (3): LBP map. Note that potential structure almost unchanged under different illumination condition, but LBP suffered from sudden illumination variations showed in red box.

only for squared matrix). Paper [10] uses normalized coefficients of SVD on local intensities as its illumination-invariant face representation. They utilize the Lambertian model which defines the pixel value as a product of reflectance and illumination components. SVD face values are identical in illuminated, penumbra, and umbra areas on the same object. Paper [22] defined two measures to differentiation of image features based on SVD coefficients, that two measures are not sensitive to local perturbations, changes in lighting. We can see that SVD coefficients (i.e., singular values) are likely to reveal the illumination-invariant characteristics. Therefore, we try to apply SVD to local regions. We define block B (centered by (x, y)) of $M \times M$ pixels:

$$B(x, y) = U\Sigma V^T \quad (1)$$

where U and V are orthogonal matrixes; $\Sigma = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$ is a nonnegative diagonal matrix with decreasing entries along the diagonal, it is called the singular values of $B(x, y)$. And Σ is used as the invariant to construct normalized coefficients of SVD on local color intensities. We first use a 3×3 block as an unresolved SVD matrix and obtain the singular values, then we divide the second and third diagonal singular values (λ_2, λ_3) by the largest first one λ_1 , finally sum these two values as an expression of local structural invariant. We define this scalar at each pixel position for a given frame as:

$$g(x, y) = \sum_{j=2}^M \tilde{\lambda}_j, \quad \text{and} \quad \tilde{\lambda}_j = \lambda_j / \lambda_1 \quad (2)$$

where λ_j indicates the j th singular value. Figure.1 shows the local structural invariant of local SVD. Note that po-

tential structure almost unchanged under different illumination condition, but LBP suffered from sudden illumination changes showed in red box.

Inspired by [10], we prove that normalized coefficient of SVD about local region pixels are illumination-invariant. According to Lambertian model, an image acquired by a camera, the intensity ($V(x, y)$) at 2D image position (x, y) can be defined by the product of the illumination component $I(x, y)$ and reflectance component of the object surface $F(x, y)$ [19]:

$$V(x, y) = I(x, y) \times F(x, y) \quad (3)$$

$I(x, y)$ is computed as the amount of light power per receiving object surface area and it is a function of α , α is angle between the direction of the light source and the object surface normal [19]:

$$\begin{aligned} \text{Illuminated area : } I(x, y) &= c_a + c_p \cdot \cos(\alpha) \\ \text{Penumbra area : } I(x, y) &= c_a + t(x, y) \cdot c_p \cdot \cos(\alpha) \\ \text{Umbra area : } I(x, y) &= c_a \end{aligned} \quad (4)$$

where c_a is intensities of ambient light; c_p is intensities of light source; t is transition inside the penumbra which depends on the light source and scene geometry, and ($0 \leq t(x, y) \leq 1$). Now, we denote B_i, B_p, B_u as three small image blocks of $M \times M$ pixels, they come from the same region under illuminated, penumbra (a soft transition from dark to bright), and umbra (without any light from the light source) conditions, respectively. Note that $F(x, y)$ is invariant for three light conditions based on the assumptions in [19]. We thought that $I(x_1, y_1)$ is very close to $I(x_2, y_2)$ where $(x_1, y_1), (x_2, y_2) \in B_k(x, y), k \in \{i, p, u\}$ based on the assumption that light source intensity c_p is high. Note that c_a, t, c_p and α can be seen as approximate constants in a small image block. We can redefine the equation of a small image block as follows:

$$B_p = C_p \cdot B_i, B_u = C_u \cdot B_i \quad (5)$$

then, we apply SVD to each small image block (5): $B_k = U_k \Sigma_k V_k^T$, get the singular values and describe the relationship between the three small blocks as follows:

$$\Sigma_p = C_p \cdot \Sigma_i, \Sigma_u = C_u \cdot \Sigma_i \quad (6)$$

where $\Sigma_k = \text{diag}(\lambda_1^k, \lambda_2^k, \dots, \lambda_N^k), k \in \{i, p, u\}$.

Based on (5), (3) and (4), we know

$$\begin{aligned} C_p = V_p/V_i &= (I_p \times F_p)/(I_i \times F_i) = I_p/I_i \\ &= (c_a + t \cdot c_p \cdot \cos(\alpha))/(c_a + c_p \cdot \cos(\alpha)) \end{aligned} \quad (7)$$

$$\begin{aligned} C_u = V_u/V_i &= (I_u \times F_u)/(I_i \times F_i) = I_u/I_i \\ &= c_a/(c_a + c_p \cdot \cos(\alpha)) \end{aligned} \quad (8)$$

LSBP	8-bit string	16-bit string
CDnet 2012	0.7592	0.7671

Table 1: Overall results in F1 with different bits of LSBP feature.

According (2) and (6), we can clearly obtained the following equations:

$$g_i(x, y) = \sum_{j=2}^M \lambda_j^i / \lambda_1^i \quad (9)$$

$$g_p(x, y) = \sum_{j=2}^M \lambda_j^p / \lambda_1^p = \sum_{j=2}^M (C_p \cdot \lambda_j^i) / (C_p \cdot \lambda_1^i) = g_i(x, y) \quad (10)$$

$$g_u(x, y) = \sum_{j=2}^M \lambda_j^u / \lambda_1^u = \sum_{j=2}^M (C_u \cdot \lambda_j^i) / (C_u \cdot \lambda_1^i) = g_i(x, y) \quad (11)$$

Based on the above description, we can conclude that normalized coefficient of SVD about small region pixels are illumination-invariant.

Through Equation (2), we obtained local structural invariant of each pixel for every frame, then we use it to extend LBP to LSBP. The principle of LSBP is to compare a central point value with neighbor values and check whether they are similar or not. And local structural invariant is applied for central point value and neighbor values. Texture at point (x_c, y_c) is modeled using a local neighborhood of radius R, which is sampled at P points. The LSBP binary string at a given location (x_c, y_c) can be derived from the following formula:

$$LSBP(x_c, y_c) = \sum_{p=0}^{p-1} s(i_p, i_c) 2^p \quad (12)$$

where i_c is the central point value obtained from Equation (2), i_p represents the N-neighborhood point value also obtained from Equation (2). τ is the similarity threshold which is set to 0.05 in this paper. $S(\cdot)$ is a sign function defined as follows:

$$s(i_p, i_c) = \begin{cases} 0 & \text{if } |i_p - i_c| \leq \tau \\ 1 & \text{otherwise} \end{cases} \quad (13)$$

We test our pixel models using $LSBP_{8,1}(P = 8, R = 1)$ and $LSBP_{16,4}(P = 16, R = 4)$ respectively, and the experimental results are shown in Table 1. We can see LSBP with 16-bit string is more discriminative than 8-bit vector in the task of change detection.

We also take a comparison experiment between LBP and LSBP, the result is exhibited in Figure. 2, where the selected background pixel is similar to its neighborhood, and the statistics of the pixel processes (300 frames) with LBP and

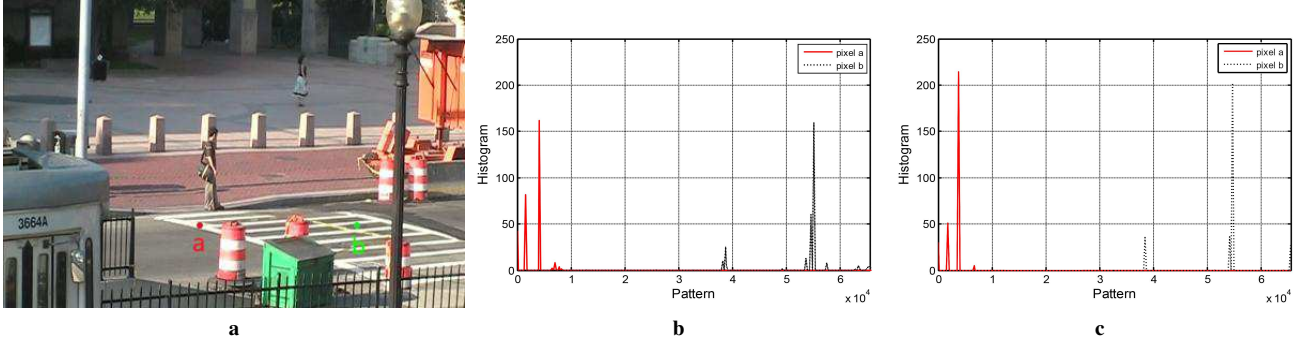


Figure 2: Comparison of LBP and LSBP features for two background pixels on real video. (a) Shows a frame from "tramstop" video, with two marked pixels. (b) and (c) are the histograms of two pixels from frames overtime, with LBP and LSBP descriptors respectively (300frames counted).

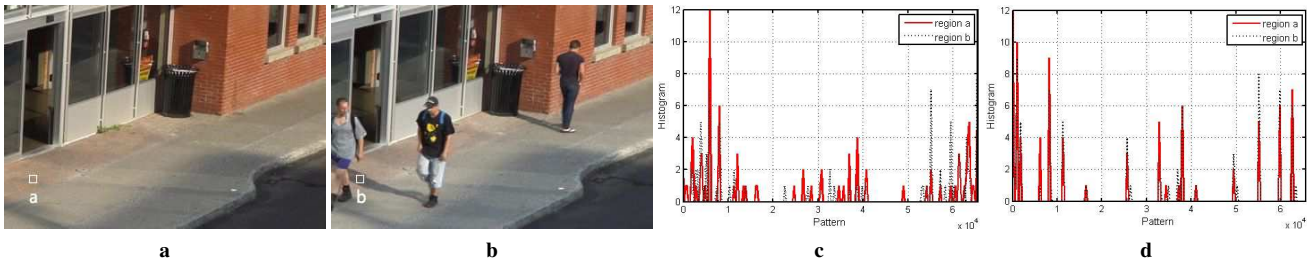


Figure 3: Comparison of LBP and LSBP features with shadow. (a) and (b) are two frames from the "busStation" video, with two 10×10 regions drawn. Regions contain the same background with and without shadows. (c) LBP histogram of two regions. (d) LSBP histograms of two regions.

LSBP descriptors are displayed. The results demonstrate that the LBP is more variable than LSBP, and the latter is almost invariant among all the 300 frames counted.

In Figure. 3, the same 10×10 region in two frames with and without shadows were compared. As can be seen from the histograms, for background with and without shadows, the LSBP operator performs perfectly, almost not influenced by shadows as only a few patterns being different between the two marked image regions, while LBP histogram shows larger difference.

Experiment results in [18] show that only using LBP comparisons usually does not fit in noisy or blurred regions. Therefore, label assignment should also rely on color intensity comparisons in order to reduce the false negative rate of our final method.

In summary, for our pixel-level modeling, we define a single background pixel description using both LSBP binary strings and color intensities. When trying to match current frame with background model, we need to compare color value with the background samples using L1 distance, meanwhile compare LSBP binary string with the background samples using Hamming distance (XOR). Therefore, to consider a current pixel similar to a background sample, both color value and LSBP binary string should be matched correctly.

3. Modeling Background using Local SVD Binary Pattern

To segment the foreground (FG) / background (BG) correctly, we think about construct reference model in a pixel-based manner. But local texture patterns are not numerical values, they have local ordering relationships. So traditional numerical value based methods (GMM [26], KDE [5] etc.) can not be used directly for modeling LSBP into background. Fortunately, sample consensus (e.g ViBe [1], PBAS [9]) model is very well suited for the description of pixels via complex features. Inspired by [9], we develop a sample consensus model that is suitable for LSBP descriptors.

The overview of our method is presented in Figure.4. The central component of our method is the FG/BG classifications block which decides a new observation for or against foreground based on the current frame and background model $B(x, y)$. This classification is based on the per-pixel threshold $R(x, y)$ and H_{LSBP} . In our method, each pixel $P(x, y)$ is modeled by an array of N recently observed background samples, the samples contain both $BInt_{index}(x, y)$ and $BLSBP_{index}(x, y)$.

$$B(x, y) = \{B_1(x, y), \dots, B_{index}(x, y), \dots, B_N(x, y)\} \quad (14)$$

Index represent the sequence number of background samples. N is the number of samples in background model.

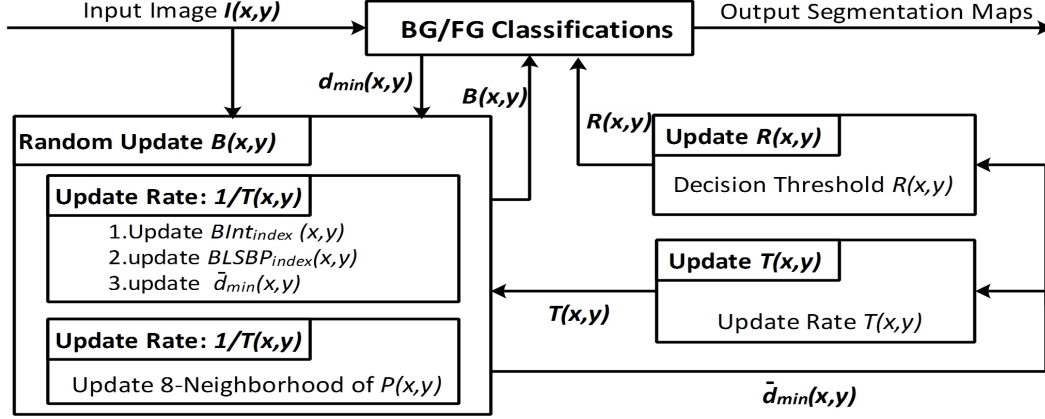


Figure 4: Illustration of framework of the proposed method.

And N is used to balance the sensitivity and precision of sample-based methods. To classify a pixel at coordinate (x, y) , the current frame should be matched against their samples. The pixel value $Int(x, y)$ and $LSBP(x, y)$ value are both need to be matched correctly. We call this combined verification.

$$\begin{aligned} (H(LSBP(x, y), BLSBP_{index}(x, y)) \leq H_{LSBP}) \\ \&\&(L1dist(Int(x, y), BInt_{index}(x, y)) < R(x, y)) \end{aligned} \quad (15)$$

The logic value of Equation (15) equals 1 demonstrates we get a match. $\#min$ is the minimum count of matches needed for classification. Like[1][9], we fixed $\#min = 2$ for our method to be a reasonable balance between noise resistance and computational complexity. For LSBP comparison, we use Hamming distance (XOR) operator similar to [18], and we fixed Hamming distance threshold H_{LSBP} as 4. $Int(x, y)$ is the color intensity at (x, y) . $R(x, y)$ is the per-pixel color intensity distance threshold. For highly dynamic areas, $R(x, y)$ should be higher, and for static regions, $R(x, y)$ should be lower. Because using L2 distance to calculate the similarity between two 3-channel samples is time-consuming, we select simpler L1 distance for the color intensity comparison.

Furthermore, the background model need to be updated over time, it allows for gradual background changes add to the background model depending on a per-pixel update parameter $T(x, y)$. We update our pixel models using a conservative, stochastic approach similar to [9]. Conservative means only update the pixel which is classified as background. Then, for a pixel $P(x, y)$ classified as background, stochastic update means that for certain random select index, the corresponding background model values ($BInt_{index}(x, y)$ and $BLSBP_{index}(x, y)$) are replaced by the current pixel value $Int(x, y)$ and LSBP value $LSBP(x, y)$ respectively. And this update is only realized with probability $p = 1/T(x, y)$. The higher $T(x, y)$, the less likely a pixel will be updated. At the same

Algorithm 1 Background Subtraction for FG/BG segmentation using LSBP feature.

Initialization:

- 1: **for** each pixel of the first N frames **do**
- 2: Extract the LSBP descriptor for each pixels using Equation (12)
- 3: Push color intensities into $BInt_{index}(x, y)$ and LSBP features into $BLSBP_{index}(x, y)$ as the background model
- 4: Compute $\bar{d}_{min}(x, y)$ for each pixel.
- 5: **end for**

Mainloop:

- 6: **for** each pixel of newly appearing frame **do**
- 7: Extract $Int(x, y)$ and $LSBP(x, y)$
- 8: **end for**
- 9: $matches \leftarrow 0$
- 10: $index \leftarrow 0$
- 11: **for** each pixel in current frame **do**
- 12: **while** $((index \leq N) \&\&(matches < \#min))$ **do**
- 13: computer $L1dist(Int(x, y), BInt_{index}(x, y))$ and $H(LSBP(x, y), BLSBP_{index}(x, y))$
- 14: **if** $((L1dist(x, y) < R(x, y)) \&\&(H(x, y) \leq H_{LSBP}))$ **then**
- 15: $matches += matches$
- 16: **end if**
- 17: $index += index$
- 18: **end while**
- 19: **if** $(matches < \#min)$ **then**
- 20: $Foreground$
- 21: **else**
- 22: $Background$
- 23: **end if**
- 24: **end for**

time, we also random update one samples of randomly-

selected 8-neighboring pixel of $P(x, y)$ with the probability of $1/T(x, y)$, the background model at this neighboring pixel are replaced by its current color intensity and LSBP value.

Both the two per-pixel thresholds ($R(x, y)$ and $T(x, y)$) are dynamically changed based on an estimate of the background dynamics $\bar{d}_{min}(x, y)$ inspired by [9]. At first, besides saving an array of recently observed pixel values and LSBP strings in the background model $B(x, y)$, we also create an array $D(x, y) = \{D_1(x, y), \dots, D_{index}(x, y), \dots, D_N(x, y)\}$ of minimal decision distances. Whenever an update of $B_{index}(x, y)$ is carried out, the currently observed minimal distance $d_{min}(x, y) = \min_{index} dist(Int(x, y), BInt(x, y))$ is written to this array: $D_{index}(x, y) \leftarrow d_{min}(x, y)$. Thus, we create a history of minimal decision distances. The average of these values $\bar{d}_{min}(x, y) = 1/N \sum_{index} D_{index}(x, y)$ is a measure of the background dynamics. Other parameters are fixed in the experiments, including the size of block 3×3 , similar thresholding $\tau = 0.05$, Hamming distance thresholding $H_{LSBP} = 4$, others similar settings in [9]. Per-channel FG / BG segmentation using both LSBP feature and color intensity is present in Algorithm 1. When the number of samples (index) is less than N and the matches less than $\#min$, continue the loop. Otherwise, we enter the classification step: if the matches less than $\#min$, the observation is classified as foreground, else as background. Note that $P(x, y)$: pixel at coordinate (x, y) ; $Int(x, y)$: current pixel value at $P(x, y)$; $LSBP(x, y)$: current LSBP string at $P(x, y)$; $BInt_{index}(x, y)$: pixel value of number index background sample at $P(x, y)$; $BLSBP_{index}(x, y)$: LSBP string of number index background sample at $P(x, y)$.

4. Experimental Results

We evaluate our method on the CDnet 2012 database which provided for the Change Detection [6]. This database features 31 real-world sequences including six different categories: baseline, camera jitter, dynamic background, intermittent object motion, shadow and thermal. Manually labeled ground truth is available for all scenarios and is used for performance evaluation.

We compare the proposed method with the six classical state-of-the-art pixel-based background subtraction algorithms: Gaussian Mixture Model by Zivkovic (GMM) [26], the improved adaptive KDE by Elgammal [5], SOBS [14], ViBe [1], SuBSENSE [18] and PBAS [9]. To provide a better understanding about the classification results, typical segmentation results for various sequences of the CDnet2012 dataset are shown in Figure.5. We select the following sences: "highway" and "PETS2006" from the "baseline" category, "copyMachine" from "shadow", "overpass" from "dynamicBackground", "sofa" from "intermittentObject Motion". Segmentation results of PBAS, SuBSENSE,

GMM and KDE methods are obtained from BGSLibrary [17], result of SOBS is come from the CDNET public website.

The "highway" sequences contain dynamic branches and leaves and their shadows on the ground surface, the color of the car often similar to the ground (dark gray) or shadows (black). They should be tolerated in foreground detection. The proposed method separated the background and foreground satisfactorily. The results have shown that the other methods also can provide good performance in handling such non-stationary background, SuBSENSE is relatively better.

"CopyMachine" is a indoor sence but contain intensive light from outdoor. There are shadows about curtains and persons, and the people stand for a while then walk forward. From the results shown in Figure.5 in second column, we can know that this is a challenge problem. The results have shown that the proposed method has detected the person quite well in such an environment. SOBS and GMM can not get the complete foreground results of the person.

"PETS2006" is a environment of railway station. Every frame contains people, they are walk up and down. Soft shadows of moving persons cast on the ground from different directions. The proposed method has obtained the satisfactory results in this environments, detected and removed shadows successfully. In this video SuBSENSE obtained the best results.

"Overpass" which contains dynamic Background, shows pedestrians passing int front of a tree shaken by the wind. This is a challenge sence. In this cases, the proposed method can yield superior performance than the several former works in terms of the test results. But no one get the complete foreground. Result of SOBS in this video is much better than others.

"Sofa" sequences contain challenge about intermittent Object Motion. There are two person move around, then stop for a short while. One person is seen wearing a dark color trousers which is similar to the color of the sofa. The results shown that the proposed method has detected the persons quite well in such an environment. PBAS and SuBSENSE also got good results.

Visually, the results of proposed method look better and are the closest to ground-truth references. This is confirmed by the results of quantitative evaluation.

With standardized evaluation tools, we can easily compare our results to other state-of-the-art methods based on the following official metrics: recall (Re), precision (Pr) and F-measure ($F1$). Recall, also known as detection rate, gives the percentage of detected true positives as compared to the total number of true positives in the ground truth

$$recall = \frac{TP}{TP + FN} \quad (16)$$

where TP is the total number of true positives, and FN is

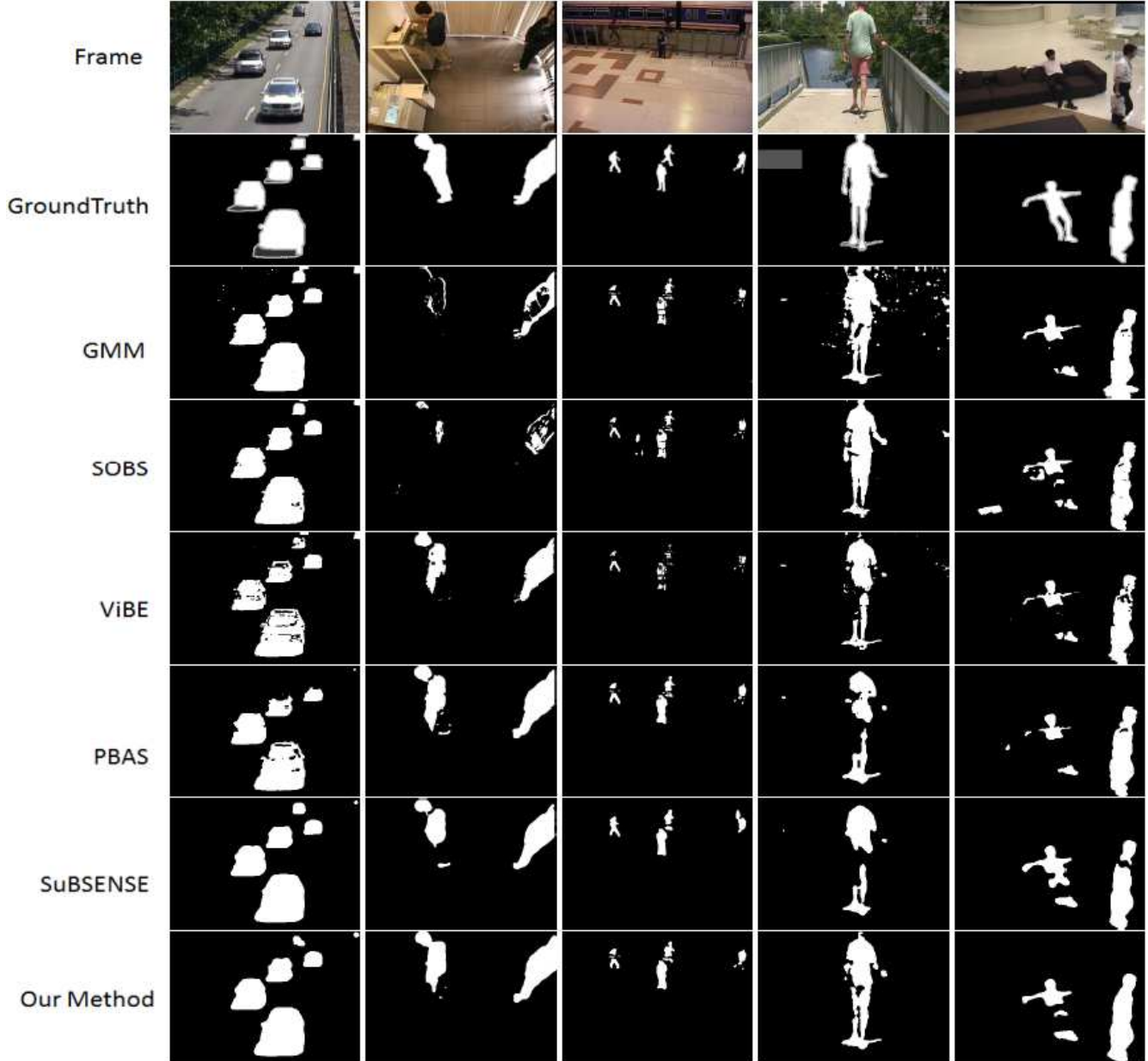


Figure 5: Typical segmentation results for various sequences of the CDnet2012 dataset; row (1) shows input frame, row (2) shows groundtruth, (3) shows GMM results, (4) SOBS results, (5) ViBE results, (6) PBAS results, (7) SuBSENSE results, (8) Our results. From left to right, the sequences are highway (baseline), copyMachine (shadow), PETS2006 (baseline), overpass(dynamic Background), sofa (intermittent Object Motion).

the total number of false negatives, which accounts for the number of foreground pixels incorrectly classified as background.

$$precision = \frac{TP}{TP + FP} \quad (17)$$

Precision, also known as positive prediction, that gives the percentage of detected true positives as compared to the total number of pixels detected by the method, is generally

used in conjunction with the recall. Where FP is the total number of false positives. Generally, a method is considered good if it reaches high recall values, without sacrificing precision. So, F-measure ($F1$) metric also adopted, that is mainly used to compare the performance of different methods which ensures the segmentation accuracy by balance

Scenarios	Recall	Precision	F1
Baseline	0.9535	0.9465	0.9289
Camera Jitter	0.7375	0.7495	0.7332
Dynamic Background	0.7197	0.7875	0.6924
Intermittent Object Motion	0.6827	0.6920	0.5873
Shadow	0.9193	0.8568	0.8865
Thermal	0.7821	0.8805	0.7741
Overall	0.7991	0.8188	0.7671

Table 2: Results for our method of the 2012 CDnet dataset.

Methods	Recall	Precision	F1
GMM[26]	0.7108	0.7012	0.6623
KDE[5]	0.7442	0.6843	0.6719
SOBS[14]	0.7882	0.7179	0.7159
ViBe[1]	0.6821	0.7357	0.6683
SuBSENSE[18]	0.8280	0.8580	0.8260
PBAS[9]	0.7840	0.8160	0.7532
Our method	0.7991	0.8188	0.7671

Table 3: Comparison using Recall, Precision and F1 performance measures with six different methods on the 2012 CDnet dataset.

recall and precision. The $F1$ is defined as:

$$F1 = 2 \frac{\text{recall} \cdot \text{precision}}{\text{recall} + \text{precision}} = \frac{2TP}{2TP + FN + FP} \quad (18)$$

In Table 2, we exhibit our average result of each category in CDnet 2012 dataset. As a whole, we can see that the "shadow" and "thermal" categories exhibit the best improvements while the "dynamicbackground" and "baseline" categories seem to perform at a level comparable to PBAS (most likely as a side-effect of the increased recall). As a side note, PBAS is one of the best methods according to the evaluation results on the CDnet 2012, SuBSENSE is the first place in CDnet 2014 in terms of $F1$ at present.

In Table 3 we present the overall averaged results of our method, lined up with the other state-of-the-art algorithms. The Results show that our LSBP-based background subtraction method outperforms most of them except SuBSENSE.

5. Conclusion

We have proposed an adaptive background subtraction method, it used a novel Local SVD Binary Pattern. Our method outperformed several state-of-the-art algorithms. Experiments have demonstrated that incorporated LSBP feature in our adaptive pixel-based sample consensus method could enhance robustness to illumination changes, shadows and noise. For future work, we will apply LSBP

Experimental results under comparison are come from paper [9, 18].

features for pixel-level feedback scheme which automatically adjusts internal sensitivity to change and update rates.

Acknowledgements

This work was supported by National Natural Science Foundation of China (NSFC) under 61540062, 61271361, 61262067, 61462093.

References

- [1] O. Barnich and M. Van Droogenbroeck. ViBe: A universal background subtraction algorithm for video sequences. *IEEE Transactions on Image Processing*, 20(6):1709 – 1724, 2011.
- [2] G.-A. Bilodeau, J.-P. Jodoin, and N. Saunier. Change detection in feature space using local binary similarity patterns. In *Proceedings - 2013 International Conference on Computer and Robot Vision, CRV 2013*, pages 106–112, 2013.
- [3] T. Bouwmans. Traditional and recent approaches in background modeling for foreground detection: An overview. *Computer Science Review*, 11-12:31 – 66, 2014.
- [4] J. chen, S. G. Shan, H. Chu, G. Y. Zhao, P. Matti, X. Chen, and W. Gao. WLD: a robust local image descriptor. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(9):1705 – 1720, 2010.
- [5] A. Elgammal, R. Duraiswami, D. Harwood, and L. S. Davis. Background and foreground modeling using nonparametric kernel density estimation for visual surveillance. *Proceedings of the IEEE*, 90(7):1151 – 1162, 2002.
- [6] N. Goyette, P.-M. Jodoin, F. Porikli, J. Konrad, and P. Ishwar. changedetection.net: A new change detection benchmark dataset. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, pages 1 – 8, 2012.
- [7] M. Heikkila and M. Pietikainen. A texture-based method for modeling the background and detecting moving objects. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(4):657 – 662, 2006.
- [8] M. Heikkila, M. Pietikainen, and C. Schmid. Description of interest regions with local binary patterns. *Pattern Recognition*, 42(3):425 – 436, 2009.
- [9] M. Hofmann, P. Tiefenbacher, and G. Rigoll. Background segmentation with feedback: The pixel-based adaptive segmenter. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, pages 38 – 43, Providence, RI, United states, 2012.
- [10] W. Kim, S. Suh, W. Hwang, and J.-J. Han. SVD face: Illumination-invariant face representation. *IEEE Signal Processing Letters*, 21(11):1336 – 1340, 2014.
- [11] S. Liao, G. Zhao, V. Kellokumpu, M. Pietikainen, and S. Z. Li. Modeling pixel process with scale invariant local patterns for background subtraction in complex scenes. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 1301–1306, 2010.
- [12] L. Lin, Y. Xu, X. Liang, and J. Lai. Complex background subtraction by pursuing dynamic spatio-temporal models. *IEEE Transactions on Image Processing*, 23(7):3191 – 3202, 2014.

- [13] X. Liu and C. Qi. Future-data driven modeling of complex backgrounds using mixture of gaussians. *Neurocomputing*, 119:439 – 453, 2013.
- [14] L. Maddalena and A. Petrosino. The SOBS algorithm: What are the limits? In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, pages 21 – 26, 2012.
- [15] T. Ojala, M. Pietikainen, and T. Maenpaa. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(7):971 – 987, 2002.
- [16] X. Qi, R. Xiao, C. G. Li, Y. Qiao, J. Guo, and X. Tang. Pair-wise Rotation invariant Co-Occurrence Local Binary Pattern. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(11):2199 – 2213, 2014.
- [17] A. Sobral. BGSLibrary: An openCV C++ background subtraction library. In *IX Workshop de Visao Computacional (WVC'2013)*, Rio de Janeiro, Brazil, Jun 2013.
- [18] P.-L. St-Charles, G.-A. Bilodeau, and R. Bergevin. SuB-SENSE: A universal change detection method with local adaptive sensitivity. *IEEE Transactions on Image Processing*, 24(1):359 – 373, 2015.
- [19] J. Stander, R. Mech, and J. Ostermann. Detection of moving cast shadows for object segmentation. *IEEE Transactions on Multimedia*, 1(1):65 – 76, 1999.
- [20] C. Stauffer and W. Grimson. Adaptive background mixture models for real-time tracking. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2:246 – 252, 1999.
- [21] X. Tan and B. Triggs. Enhanced local texture feature sets for face recognition under difficult lighting conditions. *IEEE Transactions on Image Processing*, 19(6):1635 – 1650, 2010.
- [22] A. T. Targhi and A. Shademan. Clustering of singular value decomposition of image data with applications to texture classification. In *Proceedings of SPIE - The International Society for Optical Engineering*, volume 5150 II, pages 972 – 979, 2003.
- [23] D. Varga, L. Havasi, and T. Sziranyi. Pedestrian detection in surveillance videos based on CS-LBP feature. In *2015 International Conference on Models and Technologies for Intelligent Transportation Systems, MT-ITS 2015*, pages 413 – 417, 2015.
- [24] H. Wang and D. Suter. A consensus-based method for tracking: Modelling background scenario and foreground appearance. *Pattern Recognition*, 40(3):1091 – 1105, 2007.
- [25] J. Yao, Z. Xu, X. Huang, and J. Huang. Accelerated dynamic MRI reconstruction with total variation and nuclear norm regularization. volume 9350, pages 635 – 642, Munich, Germany, 2015.
- [26] Z. Zivkovic. Improved adaptive Gaussian mixture model for background subtraction. In *Proceedings - International Conference on Pattern Recognition*, volume 2, pages 28–31, 2004.