Spatio-temporal Center-symmetric Local Derivative Patterns for Objects Detection in Video Surveillance

Marwa Jmal[1,2], Wided Souidene[1] and Rabah Attia[1]
[1]*SERCOM, Ecole Polytechnique de Tunisie, Université de Carthage, B.P.743, 2078., La Marsa, Tunisia*
[2]*Telnet Innovation Labs, Telnet Holding, Ariana, Tunisia*

Keywords: Local Derivative Patterns, Spatio-temporal Features, Background Modeling, Background Subtraction.

Abstract: Nowadays, more attention is being focused on background subtraction methods regarding their importance in many computer vision applications. Most of the proposed approaches are classified as pixel-based due to their low complexity and processing speed. Other methods are considered as spatiotemporal-based as they consider the surroundings of each analyzed pixel. In this context, we propose a new texture descriptor that is suitable for this task. We benefit from the advantages of local binary patterns variants to introduce a novel spatio-temporal center-symmetric local derivative patterns (STCS-LDP). Several improvements and restrictions are set in the neighboring pixels comparison level, to make the descriptor less sensitive to noise while maintaining robustness to illumination changes. We also present a simple background subtraction algorithm which is based on our STCS-LDP descriptor. Experiments on multiple video sequences proved that our method is efficient and produces comparable results to the state of the art.

1 INTRODUCTION

Ensuring humans' security, either in public or private spaces, is becoming a major priority for all nations. This issue arises the need for video surveillance systems which consist basically in objects detection, objects tracking and behavior understanding (Jain and Favorskaya, 2015). The most important task in surveillance systems is moving objects detection. Thus, a robust detection will highly increase the effectiveness of the surveillance.

For many decades, a significant amount of research in the computer vision field has been devoted to the task of objects detection. The most straightforward technique employed in this context, is background subtraction. In its simplest form, it aims to extract the foreground which represents the relevant objects that remain always in motion. Even though it seems to be simple, this technique has to cope with different challenges occurring from dynamic backgrounds (waving trees, water fountains), illumination variations, camera jitter as well as other challenges that are well depicted in (Bouwmans et al., 2010). To deal with these situations, several works (Sobral and Vacavant, 2014; Shahbaz et al., 2015; Benezeth et al., 2010) have been achieved. Although, some of them focused on videos captured by freely moving camera (Megrhi et al., 2015; Sheikh et al., 2009), the majority of models conceived background subtraction approaches for videos captured by static cameras.

In order to detect sudden events, real-time processing is a requirement in video surveillance systems. This is behind the fact that most employed methods in this field are based on independent pixel-level models which are then integrated in a global background model.

Color-based methods consist in dynamically comparing pixel colors at different positions against a threshold. These methods are very sensitive to illumination changes in the scene. Some of them will not be detected as they involve groups of pixels in which some independent pixels may preserve an appearance similar to the background. The remedy to this issue is to formulate the problem in the feature space: instead of employing pixels colors for comparison, features in the current frame are compared with features in the background model. Lately, Local Binary Pattern (LBP) (Ojala et al., 2002) was adapted to the task of background subtraction. It describes a pixel by a series of bits basing on the gray intensity levels of its surrounding neighbors. LBP was first employed by (Heikkilä and Pietikäinen, 2006) in this context. It was proven that this descriptor is simple, invariant to illumination and computationally effective. More-

Jmal, M., Souidene, W. and Attia, R.
Spatio-temporal Center-symmetric Local Derivative Patterns for Objects Detection in Video Surveillance.
DOI: 10.5220/0005787702150220
In *Proceedings of the 11th Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2016) - Volume 3: VISAPP*, pages 217-222
ISBN: 978-989-758-175-5

over, representing LBP basing on histograms, makes it invariant to translations. Since then, many attempts to build robust descriptor were proposed. However, most of them are either computationally expensive or result in long histograms. In (Silva et al., 2015) a comparison of these methods is provided.

In this paper, we propose a new feature descriptor named STCS-LDP (Spatio-Temporal Center Symmetric Local Derivative Patterns) which is an optimized and enhanced version of an LBP variant proposed in (Xue et al., 2011). Our main improvements lies in the neighboring pixels comparison level. To validate STCS-LDP, we integrated it in a simple background subtraction process. Experimental results, carried out on a subset of the CDNet dataset (Wang et al., 2014), showed that the proposed descriptor is robust to illumination changes and produces a short descriptor. The remainder of the paper is organized as follows: Section 2 presents the related works to pixel- and LBP-based methods for background subtraction. Section 3 provides a description of the proposed descriptor. Experimental results are depicted in Section 4 while Section 5 draws some conclusions and perspectives.

2 LITERATURE REVIEW

In general, pixel-based background subtraction methods are simple and robust in several scenarios. Their major drawback is the sensitivity to illumination changes. In order to handle dynamic backgrounds, more than one pixel value should be associated to the background pixel model. In this context, parametric and non-parametric background models like Gaussian Mixture Model (GMM)(Stauffer and Grimson, 1999) and Kernel Density Estimation (KDE)(Elgammal et al., 2000) were proposed. They are well-known methods on which countless variations and improvements are made, such as in (Zivkovic, 2004). Other pixel-based methods, like VIBE (Barnich and Van Droogenbroeck, 2011), PBAS (Hofmann et al., 2012) and SuBSENCE (St-Charles et al., 2015), focused on selecting background samples randomly and diffusion labelling instead of building a probability distribution of the background of a pixel.

Ordinary pixel-based methods are based only on the use of temporal correlation between pixel values while ignoring the spatial relationship between them where an important amount of information may be lost. Subsequently, some methods attempted to formulate the problem in feature space. Heikkila et al (Heikkilä and Pietikäinen, 2006) are the first to adapt these features for dynamic background modelling. However, the produced LBP operator is long since it considers the first-order gradient information between pixel and its neighbors. Center-Symmetric local binary Pattern (CS-LBP) (Heikkilä et al., 2009) is an extension for LBP where only the relation between center symmetric neighbor pairs is considered. Although, it produces a shorter feature descriptor, it does not carry enough information for background modelling as it ignores the value of the center pixel. A Local Binary Similarity Patterns (LBSP) descriptor was proposed in (Bilodeau et al., 2013). Contrarily to histogram based patterns, this descriptor is based on absolute differences and is calculated within one image and between two images. As a consequence, LBSP succeeded to capture both texture and intensity changes. In (Xue et al., 2011), the authors apply high-order local derivative pattern to produce a center-symmetric local derivative pattern descriptor to capture more local information. This descriptor is then concatenated with CS-LBP to produce a shorter descriptor with low complexity and robust foreground detection. The disadvantage of this method, along with texture based methods, is that it detects only changes in texture while neglecting intensity values which could bring useful information. Also, even though, it is a concatenation of two short descriptors, it is really time consuming. To solve the drawbacks of both LBP and the descriptor presented in (Xue et al., 2011), we propose a new feature descriptor that is binary and captures both changes in texture and intensity.

3 METHODOLOGY

3.1 STCS-LDP

Binary feature descriptors are employed in background subtraction methods thanks to their speed, discrimination, low complexity and invariance to illumination. However, since LBP produce long feature vectors and CS-LBP ignores the central pixel information, Xue et al. (Xue et al., 2011) proposed the use of local derivative patterns which are able to capture more information in center-symmetric direction without discarding the information brought by the central pixel.

Figure 1 presents the diagrams of the three descriptors (LBP, CS-LBP and CS-LDP) with eight neighbors around the center i_c. LBP encodes in all eight direction to produce 8 bits binary sequence while CS-LBP and CS-LDP pattern encode in four directions and produce 4 bits sequence. The CS-LDP descriptor at time t is computed as follows:

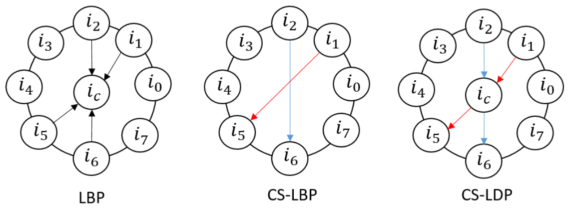

Figure 1: Example of LBP, CS-LBP and CS-LDP features over eight neighbors.

$$CSLDP^t{}_{R,N}(x_{t,c}, y_{t,c}) = \sum_{p=0}^{(N/2)-1} f[(i_{t,p} - i_{t,c}) \times (i_{t,c} - i_{t,p+N/2})]2^p \quad (1)$$

where i_c corresponds to the value of central pixel (x_c, y_c); i_p and $i_{p+N/2}$ are the values of neighborhood pixels in center symmetric direction of N equally spaced pixels on a circle R. The threshold function $f(.,.)$ is used to determine the type of local pattern transition and is defined as:

$$f(x_1 \times x_2) = \begin{cases} 1 & if\ (x_1 \times x_2) \leq 0 \\ 0 & otherwise \end{cases} \quad (2)$$

However, because CS-LDP is computed based on comparisons with a center pixel i_c, change will not be detected if the intensity of i_c remains greater (or smaller) than all neighbors i_p after a change in a scene. The solution here, is to employ a parameter T_d as a threshold when computing the descriptor. This parameter accounts for noise affecting i_c (Eq 3).

$$f(x_1 \times x_2) = \begin{cases} 1 & if\ (x_1 \times x_2) \leq T_d \\ 0 & otherwise \end{cases} \quad (3)$$

Even though local features are proved to be very discriminative, it is not guaranteed that they perform well when applied to the task of background subtraction where features are computed in every position in the image. The solution to this problem is to compute features (i) within an image to account for spatial information (Spatial CS-LDP) by selecting the center pixel to be in the same region as the neighboring pixels, and (ii) between two images to account for temporal information (Temporal CS-LDP) by selecting the center pixel to be in another region in the same image or in another one. The region in which the descriptor is computed should be small in order to capture more discriminative information. Moreover, authors in (Bilodeau et al., 2013) pointed the fact that features should not be based strictly on comparisons in order to handle the situation of large intensity changes. Thus, using absolute difference allows detecting large changes in intensity toward larger or smaller values. Finally, replacing the threshold T_d with a value that is relative to the center pixel will improve the specificity of the descriptor in high illumination variation situations. In summary, the threshold function of the STCS-LDP becomes:

$$f(x_1 \times x_2) = \begin{cases} 1 & if\ |x_1 \times x_2| \leq T_d \times i_c \\ 0 & otherwise \end{cases} \quad (4)$$

3.2 Background Subtraction with STCS-LDP

The goal of this work, is to prove that local features may produce better results in background subtraction than methods based only on intensity. To study the benefits of our STCS-LDP, we propose a simple background subtraction method that focuses mainly on the performance of the descriptor. Our method has no update process. Each new frame is compared to a background model constructed from F first frames. Within the first F frames, we compute the feature descriptor for each pixel using spatial CS-LDP (i_c and i_p are selected from the same image) and an histogram is produced. A pixel is labelled as background, if its histogram is repeated for at least B consecutive times. The repetition is measured by the degree of similarity between two consecutive histograms while the similarity is measured in terms of histogram intersection measure defined as:

$$H(x_1, x_2) = \sum_i min(x_{1,i}, x_{2,i}) \quad (5)$$

where x_1, x_2 are two normalized histograms and i is the bin index of the histogram. It is also possible to employ other distance measures such Chi-squared. This measure its chosen regarding to its simplicity and robustness as it explicitly discards features occurring only once in one of the histograms. A user settable threshold T_{desc} is used to be compared with the similarity value. The produced background model is an array of Spatial CS-LDP histograms that once created, it remains unchanged during the whole process.

In the foreground detection phase, the new frame is represented with Temporal CS-LDP (the center is selected from the current frame and the neighbors from the background model). In this level, we propose another improvement to the local features. In fact, in some situations, the Spatial CS-LDP may not perform well in noisy regions due to the fact that the center pixel of the foreground have the same value as the center pixel in the background model. To correct this problem and reduce the false negatives, a comparison of intensity values is also performed. In order to keep computations simple, we use the L1 distance measure for intensities comparison. when dealing with color images, per-channel comparisons are performed. The whole method is depicted in Algorithm 1. Note that $int_{(x,y,ch)}$ and $desc_{(x,y,ch)}$ are respectively the color intensity of channel ch and the Spatial

CS-LDP descriptor of the background model at position (x,y), *histDist* refers to the intersection measure and $TCSLDP_{(x,y,ch)}$ is the Temporal CS-LDP descriptor at position (x,y).

Algorithm 1: Background Subtraction with STCS-LDP in videos.

```
Require: Image frame Set
Ensure: Labelled frames
  Create background model;
  TotIntDist ← 0 ;
  TotDesDist ← 0 ;
  for x ← 0 : numCols do
    for y ← 0 : numRows do
      for ch ← 1 : numChannels do
        intDist ← L1(int_(x,y,ch), i_(x,y,ch));
        desDist ← histDist(desc_(x,y,ch), TCSLDP_(x,y,ch));
        TotIntDist ← TotIntDist + intDist;
        TotDesDist ← TotDesDist + desDist;
      end for
      if (TotIntDist ≥ T_int & TotDesDist ≤ T_desc) then
        p(x,y) is foreground;
      else
        p(x,y) is background;
      end if
    end for
  end for
```

4 PERFORMANCE EVALUATION

We evaluate the use of STCS-LDP in background subtraction by means of the CDnet dataset (Wang et al., 2014). Since our method does not include any update process for the background model, we have tested our background subtraction only on the baseline and thermal video subsets (9 videos, 27149 frames). We have used exactly the same metrics provided in (Wang et al., 2014). Let TP the number of true positives, TN the number of true negatives, FN the number of false negatives, and FP the number of false positives. The 7 metrics used are:

1. Recall (Re): $TP/(TP+FN)$
2. Specificity (Sp): $TN/(TN+FP)$
3. False Positive Rate (FPR): $FP/(FP+TN)$
4. False Negative Rate (FNR): $FN/(TN+FP)$
5. Percentage of Wrong Classifications (PWC): $100\times(FN+FP)/(TP+FN+FP+TN)$
6. Precision (Pr): $TP/(TP+FP)$
7. F measure: $2\times Pr\times Re/(Pr+Re)$

These metrics are provided with evaluation tools which are available online (http://www.changedetection.net). The parameters used for our method are:

- $T_{desc}=30$: threshold used to determine if an input pixel matches the background model based on the intersection measure,
- $T_{int}=90$: threshold used to determine if an input pixel matches the background model based on the L1 distance,
- $T_d=8$: the STCS-LDP descriptor threshold,
- $F=100$: number of frames considered to build the background model,
- $B=30$: required number of similar histograms to label a pixel as background.

We first investigate the effect of T_d and T_desc on the subtraction results. Then, we compare the performance of our STCS-LDP background subtraction technique against some methods from the state of the art.

4.1 Parameters Analysis

We investigated the effect of the parameters on the performance of background subtraction. We made the computation with $T_{desc}\in[5,45]$ and $T_d\in[1,20]$. The obtained results revealed that T_d has more effect on STCS-LDP performance than T_{desc}. In fact, when T_d is low, the descriptor models textures resulted from small changes in intensity, thus it becomes more sensitive to noise. If T_d is moderate, textures of small details disappear. Histograms will be robust to noise, but at the expense of detailed texture models. However, if T_d is high, detailed texture of important changes totally disappear. For T_{desc}, if it is low, any change in texture or intensity will be detected, then the performance of the descriptor will depend on the value of T_d. If T_{desc} is high, only relevant textures of intensity changes will be detected. Therefore, the value of T_d is more critical than T_{desc}. T_d should be set to be higher than noise.

4.2 Comparison with the State of the Art

To evaluate the use of our proposed descriptor in background subtraction, we compared it with some methods tested on the same dataset (Wang et al., 2014). We selected some of both best and classical methods (see Tables 1 and 2). Note that we didn't apply any morphological operations as preprocessing

Table 1: Results on Baseline Dataset (Wang et al., 2014).

Method	Re	Sp	FPR	FNR	PWC	Pr	F-measure
SubSENCE(St-Charles et al., 2015)	0.952	0.9982	0.0018	0.0480	0.3574	0.9495	0.9503
STCS-LDP	0.843	0.9985	0.0014	0.156	0.6118	0.9455	0.8915
LBSP(Bilodeau et al., 2013)	0.806	0.9977	0.0023	0.0074	0.9168	0.9275	0.8623
Euclidean(Benezeth et al., 2010)	0.838	0.9955	0.0045	0.1615	1.026	0.872	0.9114
KDE(Elgammal et al., 2000)	0.747	0.9954	0.0046	0.2528	1.8058	0.7392	0.7998
GMM(Stauffer and Grimson, 1999)	0.586	0.9987	0.0013	0.4137	1.9381	0.7119	0.9532

Table 2: Results on Thermal Dataset (Wang et al., 2014).

Method	Re	Sp	FPR	FNR	PWC	Pr	F-measure
SubSENCE(St-Charles et al., 2015)	0.8161	0.9908	0.0092	0.1839	2.0125	0.8328	0.8171
STCS-LDP	0.8354	0.9877	0.0123	0.1646	2.3626	0.8463	0.8408
LBSP(Bilodeau et al., 2013)	0.6535	0.9916	0.0083	0.0142	2.0774	0.7794	0.6924
Euclidean(Benezeth et al., 2010)	0.5111	0.9907	0.0093	0.4889	3.8516	0.6313	0.8877
KDE(Elgammal et al., 2000)	0.4147	0.9981	0.0019	0.5853	5.4152	0.4989	0.9164
GMM(Stauffer and Grimson, 1999)	0.3395	0.9993	0.0007	0.6605	4.8419	0.4767	0.9709

of the obtained results. Our method is characterized by a low FNR, high precision, average PWC, average recall, and average FRR. It performs better than the euclidean measure which is based only on pixel colors. This proves that texture based features are less sensitive to illumination changes. Compared to the LBSP, STCS-LDP reached slightly better results even though they use a larger descriptors (5×5) than ours. This is due to the fact that local derivative patterns provide a robust descriptor even when extracted over a small region. Although it combines detections of both color and texture, the KDE method was not very successful due to the failure of texture in uniform regions where only individual pixel colors may detect changes. In our method, we boost texture comparison between histograms with intensity comparison between pixel values. Finally, we believe that STCSLDP performed really well on both datasets even when compared to SubSENCE, one of the best ranked methods. Contrarily to our simple method, SubSENCE is based on pixel-level feedback loops to dynamically adjust internal parameters without user intervention.

5 CONCLUSION

In this work, we propose a new texture descriptor for the purposes of background subtraction. The proposed STCS-LDP may be adapted to slightly dynamic situations as it can be computed spatially (in the same frame) or temporally (between two frames). we proposed some improvements in the neighboring pixels comparison level to make the descriptor less sensitive to noise while maintaining robustness to illumination changes. Moreover, we compared our descriptor against some state of the art algorithms and showed that it achieved comparable results. Future work will be to bring more enhancements on the descriptor and integrate it in more sophisticated background subtractor that is based on an update process to deal with more complex situations.

ACKNOWLEDGEMENTS

This research and innovation work is carried out within a MOBIDOC thesis funded by the European Union under the PASRI project and administered by the ANPR.

REFERENCES

Barnich, O. and Van Droogenbroeck, M. (2011). Vibe: A universal background subtraction algorithm for video sequences. *Image Processing, IEEE Transactions on*, 20(6):1709–1724.

Benezeth, Y., Jodoin, P.-M., Emile, B., Laurent, H., and Rosenberger, C. (2010). Comparative study of background subtraction algorithms. *Journal of Electronic Imaging*, 19(3):033003–033003.

Bilodeau, G.-A., Jodoin, J.-P., and Saunier, N. (2013). Change detection in feature space using local binary similarity patterns. In *Computer and Robot Vision (CRV), 2013 International Conference on*, pages 106–112. IEEE.

Bouwmans, T., El Baf, F., Vachon, B., et al. (2010). Statistical background modeling for foreground detection: A survey. *Handbook of Pattern Recognition and Computer Vision*, 4(2):181–189.

Elgammal, A., Harwood, D., and Davis, L. (2000). Non-parametric model for background subtraction. In *Computer VisionECCV 2000*, pages 751–767. Springer.

Heikkilä, M. and Pietikäinen, M. (2006). A texture-based method for modeling the background and detecting moving objects. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 28(4):657–662.

Heikkilä, M., Pietikäinen, M., and Schmid, C. (2009). Description of interest regions with local binary patterns. *Pattern recognition*, 42(3):425–436.

Hofmann, M., Tiefenbacher, P., and Rigoll, G. (2012). Background segmentation with feedback: The pixel-based adaptive segmenter. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2012 IEEE Computer Society Conference on*, pages 38–43. IEEE.

Jain, L. C. and Favorskaya, M. N. (2015). Practical matters in computer vision. In *Computer Vision in Control Systems-2*, pages 1–10. Springer.

Megrhi, S., Jmal, M., Beghdadi, A., and Mseddi, W. (2015). Spatio-temporal action localization for human action recognition in large dataset. In *IS&T/SPIE Electronic Imaging*, pages 94070O–94070O. International Society for Optics and Photonics.

Ojala, T., Pietikäinen, M., and Mäenpää, T. (2002). Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 24(7):971–987.

Shahbaz, A., Hariyono, J., and Jo, K.-H. (2015). Evaluation of background subtraction algorithms for video surveillance. In *Frontiers of Computer Vision (FCV), 2015 21st Korea-Japan Joint Workshop on*, pages 1–4. IEEE.

Sheikh, Y., Javed, O., and Kanade, T. (2009). Background subtraction for freely moving cameras. In *Computer Vision, 2009 IEEE 12th International Conference on*, pages 1219–1225. IEEE.

Silva, C., Bouwmans, T., and Frélicot, C. (2015). An extended center-symmetric local binary pattern for background modeling and subtraction in videos. In *International Joint Conference on Computer Vision,(VISAPP)*.

Sobral, A. and Vacavant, A. (2014). A comprehensive review of background subtraction algorithms evaluated with synthetic and real videos. *Computer Vision and Image Understanding*, 122:4–21.

St-Charles, P.-L., Bilodeau, G.-A., and Bergevin, R. (2015). Subsense: A universal change detection method with local adaptive sensitivity. *Image Processing, IEEE Transactions on*, 24(1):359–373.

Stauffer, C. and Grimson, W. E. L. (1999). Adaptive background mixture models for real-time tracking. In *Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on.*, volume 2. IEEE.

Wang, Y., Jodoin, P.-M., Porikli, F., Konrad, J., Benezeth, Y., and Ishwar, P. (2014). Cdnet 2014: An expanded change detection benchmark dataset. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2014 IEEE Conference on*, pages 393–400. IEEE.

Xue, G., Song, L., Sun, J., and Wu, M. (2011). Hybrid center-symmetric local pattern for dynamic background subtraction. In *Multimedia and Expo (ICME), 2011 IEEE International Conference on*, pages 1–6. IEEE.

Zivkovic, Z. (2004). Improved adaptive gaussian mixture model for background subtraction. In *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*, volume 2, pages 28–31. IEEE.