Multi-Scale Continuous Gradient Local Binary Pattern for Leaky Cable Fixture Detection in High-Speed Railway Tunnel

YUNZUO ZHANG [ID], ZHOUCHEN SONG [ID], AND WEI GUO [ID]

School of Information Science and Technology, Shijiazhuang Tiedao University, Shijiazhuang 050043, China

Corresponding author: Yunzuo Zhang (zhangyunzuo888@sina.com)

ABSTRACT The feature of leaky cable fixture extracted by Local Binary Pattern (LBP) and its variants in high-speed railway tunnel has the defects of lacking description and high dimension. This paper proposes a new operator named Multi-scale Continuous Gradient Local Binary Pattern (MCG-LBP), which can realize the scale transformation of feature maps and ensure the low dimensionality of descriptors. For MCG-LBP, firstly a bi-directional triplet around the central pixel is presented to indicate the specific direction of gradient in circle neighborhood. Then, an effective dimensionality reduction strategy is introduced to perform successive down-sampling iterations. Finally, the multi-scale joint descriptors are encoded by continuous gradient sequences from different down-sampling maps, and Support Vector Machines is used to classify faulty cable fixtures. The proposed MCG-LBP can elicit a discriminative description through complementary gradient information generated by the combination of different single-scale features. While the low dimensionality of descriptor and no complex parameter to deal with both make it has higher computational efficiency. Experimental results show that the Recall and Precision of MCG-LBP reach 92.6% and 83.5% respectively on cable fixture data set, which is superior to the state-of-the-art methods.

INDEX TERMS MCG-LBP, bi-directional triplet, dimensionality reduction, fault detection, leaky cable fixture, support vector machines.

I. INTRODUCTION

High-speed rail is one of the most economical transportations globally, and it's also a landmark product of modern informatization construction. As the transportation demand increases, it is particularly essential to maintain the regular operation of railway communication systems. The network of mountain tunnel sections is mainly realized by covering leaky cables that are usually hung on the tunnel wall and fixed with specialized fixtures [1]. Fig. 1 shows the leaky cable fixtures which will be seriously affected by air pressure and power waves generated when a high-speed train passes through the tunnel. At the same time, the humid tunnel geological environment will also accelerate the loosening or even detachment of cable fixtures. The detection method of manual positioning and regular inspection one by one is not only

The associate editor coordinating the review of this manuscript and approving it for publication was Farid Boussaid.

restricted severely by environmental and human factors, but also has great safety hazards. With the increasing mileages of railway tunnels, it is obvious that traditional inspection methods can no longer meet actual needs, and automatic inspection has become an inevitable trend of the development in cable fixture detection. To achieve this technology, firstly a high-speed camera needs to be placed on the train to collect an entire picture that is taken during the tunnel section. Then computer vision approaches are used to extract cable fixture characteristics frame by frame to realize inspection work [2]. Currently, the database acquisition part has been implemented already, but the detection part is still mainly in the stage of manual playback video to troubleshoot, which requires a lot of human resources. As for handcraft methods, Shang uses Snake to determine the location of leaky cable, and also operates Haar to extract the characteristics of the fixtures, which are easily disturbed by wall cracks and background noise [3]. To achieve fixture detection, Zhang *et al.*

FIGURE 1. The samples of leaky cable fixture. (a) Tunnel and leaky. (b) Faulty cable fixture and normal cable fixture.

uses Yolo.v4 [4], and Ma uses SSD [5], but their results are not good due to the insufficient data set of faulty cable fixture. Therefore, it is necessary to continuously optimize the feature extraction algorithm, reduce the dependence on the fault data set, and improve the anti-interference ability. Some comprehensive surveys about texture descriptors can be found in [6].

Due to the discriminative quality and ease to train with a small amount of data, Local Binary Pattern (LBP) proposed by Ojala *et al.* can get more favor on cable fixture detection [7]. Moreover, feature extraction and texture classification based on statistical learning of LBP are both widely used in other fields [5], [6]. Such as fault detection [8], medical images analysis [9], face detection and recognition [10], texture classification [11], and many others. Cable fixture fault detection in this paper is just one application of the related research that can make good use of texture analysis by spatial distribution of gray level values. The conventional method to improve detection performance of LBP is conducting deep mining of texture information or making multi-scale features fusion, which can deal with external changes in the imaging conditions like rotation [12], [13], scaling, illumination [14], and noise interference based on threshold scheme [15]. Lots of variants have been proposed after the emergence of LBP. Local Ternary Pattern (LTP) quantizes local difference into three levels [16], which responses sensitively to grey-scale changes and is also easily affected by noise. Center-Symmetric Local Binary Pattern (CS-LBP) as one representative variant of LBP, which is proposed to enhance robustness, and its expansion of central symmetric difference information make computation complexity and feature dimensionality reduce greatly [17]. But CS-LBP ignores the role of central pixel and it is very troublesome to parameters. Completed Local Binary Pattern (CLBP) consisted of three complementary components by introducing signal information, magnitude information, and mean gray level [18]. CLBP operator takes both local structure and textural magnitude into consideration to achieve much better classification accuracy. Then Completed Local Ternary Pattern (CLTP) has subdivided texture features further, it also causes a serious limitation of excessively large dimensions [19]. The operator called Attractive-and-Repulsive Center-symmetric Local Binary Pattern

(ARCS-LBP) establishes the relations between central pixel and center-symmetric pixels pairs to improve the quality of descriptors [20]. To highlight the advantages of gradient components, Local Derivative Pattern (LDP) computes over derivative images in four directions and is coded by comparing the edge and corner responses [21], but none of them is able to achieve feature scaling transformation. Three-Patch Local Binary Pattern (TPLBP) enlarges feature scaling by extending extraction unit from pixel to block to calculate the similarity difference between relevant blocks [22]. Local Directional Ternary Pattern (LDTP) uses mask to expand scaling, and combines the ternary mode and derivative transformation to enhance the expression of edge features [23]. But it will also cause a lot of calculations, especially for the small target in large image. In order to realize the automatic detection of leaky cable fixture, the issues such as incomplete expression, high dimensionality, complex calculation, and other challenges need to be resolved.

In this paper, we propose an effective operator to achieve cable fixture detection, and the main contributions are highlighted as follows:

- The bi-directional triplet (BDT) is presented to express a more specific gradient direction in circle neighborhood, which can also avoid complicated parameter adjustments.

- An effective dimensionality reduction strategy is introduced in the down-sampling process, we perform non-maximum suppression according to the key gradient frequency in Cell unit and obtain the following down-sampling map, which can realize the scale transformation and further improve the computational efficiency of extracting continuous gradient features.

- The continuous gradient feature is proposed to obtain the edge and corner features of cable fixture contour, which counts the combination of relevant continuous gradient sequences in Block unit. By cascading continuous gradient features from multiple down-sampling maps, the final MCG-LBP descriptor is constructed.

Multiple sets of continuous gradient feature extracted from multiple down-sampling maps of different scales are cascaded to obtain the final MCG-LBP descriptor. The outline of this paper is organized as follows: In section II, there is a briefly discuss about the related methods such as LBP, some variant operators and non-LBP operators. Section III describes BDT, dimensionality reduction strategy, and continuous gradient feature in detail. In section IV, extensive experimental results compared with other LBP variant and non-LBP operators are demonstrated. Conclusion is given in Section V.

II. RELATED WORK

In this part, we mainly introduce the feature extraction algorithms such as LBP, CS-LBP and ARCS-LBP. A review of LBP variant and non-LBP descriptors is also given at below.

A. LBP AND CS-LBP

LBP is proposed by Ojala *et al.* originally, and it is likely to capture the sharp difference between the gradation fine texture [7]. A slight change on the encoding method can make this operator have better orientation invariance and rotation invariance, so it is widely used in face recognition and texture classification field. The coding idea of LBP is to compare grayscale difference value between center pixel and P sampling pixels in the circle neighborhood where R is the radius. Then use the Boolean function to calculate a string of binary sequence that contains only 0 or 1. The calculation formula of LBP is defined as follow:

$$LBP_{P,R}(x_c, y_c) = \sum_{i=0}^{P-1} s(g_i - g_c)2^i \qquad (1)$$

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

where (x_c, y_c) is the coordinate of the center pixel c, whereas g_c and g_i ($i = 0, 1, \ldots, P-1$) denote the gray value of center pixel c and sampling pixels in the neighborhood respectively, P is the total number of involved sampling point, and R is the radius of the circle neighborhood.

In order to enhance the performance of LBP descriptor in the spatial direction, CS-LBP is mentioned by Heikkila *et al.* [17]. Instead of LBP operator that encodes the grayscale difference value between sampling points and center pixel, CS-LBP codes the relationship between the gray level difference of pixel pairs that is symmetrical around the center pixel, and the threshold needs to be calculated from experiment. The formula of CS-LBP is defined as follow:

$$CS\text{-}LBP_{P,R,T}(x_c, y_c) = \sum_{i=0}^{(P/2)-1} s(g_i - g_{i+(P/2)})2^i \qquad (3)$$

$$s(x) = \begin{cases} 1, & x \geq T \\ 0, & x < T \end{cases} \qquad (4)$$

where $g_i(i = 0, 1, \ldots, (P/2) - 1)$ and $g_{i+(P/2)}$ both correspond to the gray value of peripheral pixels, and T is the threshold that can be specified by the global average gray value of input picture generally. Obviously, CS-LBP is closely related to gradient operator, as it refers to the gray level difference between pixel pair in a circle neighborhood. The following three aspects are the main advantages of CS-LBP operator: 1) As shown in Fig. 2, under the same conditions of sampling number and sampling radius, the feature vector of CS-LBP operator is more compact, lesser number of comparisons is required, and the feature dimension of the descriptor is also lower [17]. 2) CS-LBP inherits good properties from both texture and gradient operators [24]. 3) CS-LBP descriptor has higher stability on flat regions and performs more robustly than those descriptors that only uses texture or gradient [25].
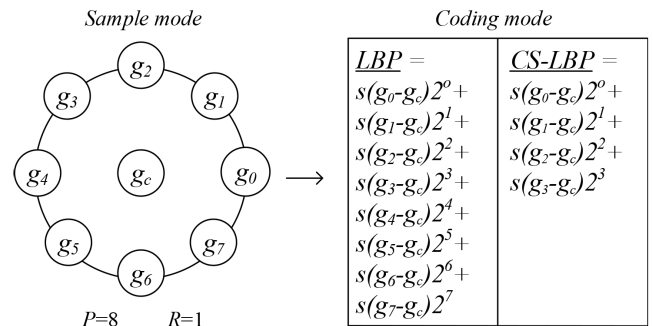

FIGURE 2. Calculating the operators of LBP and CS-LBP under same condition where the number of sampling pixels is 8 and the sampling radius is 1.

B. ARCS-LBP

CS-LBP operator ignores the key role of the center point in texture feature extraction, which leads to an incomplete display of directional characteristics. ARCS-LBP proposed by Y. El merabet is used to improve the incomplete performance [20], which not only considers the center point, but also expands pixel pairs into attractive group and repulsive group in various directions in the neighborhood. The structure model of ARCS-LBP has shown in Fig. 3, which has the same number as four directions as CS-LBP. The two additional constraint conditions of pixel pairs enhance the robustness of this fusion operator on magnitude component in different directions. However, ARCS-LBP has the same number of dimensions as LBP, and both attractive part and repulsive part need to repeatedly calculate the local pixel average values, which results in a huge amount of calculation. When calculating the magnitude difference of a pixel pair, it is also quite difficult to adjust threshold, especially for the images that have obvious brightness differences.
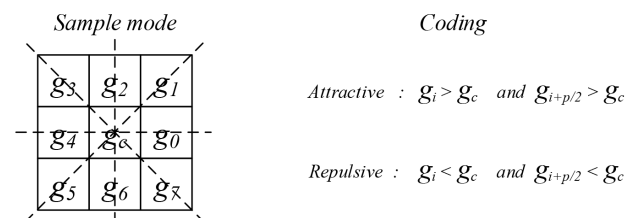

FIGURE 3. The structure of ARCS-LBP, there are vertical direction and horizontal direction, and two diagonal directions.

C. OTHER LBP VARIANT AND NON-LBP OPERATORS

The calculation process of LBP operator is very simple and easy to modify, which make it more suitable to meet different application requirements including texture classification, face recognition, scene classification, and medical image analysis, etc. Pan *et al.* proposed Local Vector Quantization Pattern (LVQP) for texture classification [26]. Chakraborty *et al.* offered Local Directional Gradient Pattern (LDGP) and Local Gradient Hexa Pattern (LGHP) to

make facial image recognition and retrieval [27], [28]. Local Maximum Edge Binary Pattern (LMEBP) and Local Tetra Pattern (LTrP) are proposed to obtain the edge and corner information for fault detection [29], [30]. These deep mining operators are mainly used to extract texture features instead of contour features. Another classic local operator, Histograms of Oriented Gradients (HOG) mainly used to detect target contour's feature by analyzing the gradient magnitude distribution from different orientations statistically [31]. Hence these fusion descriptors, like HOG-LBP and HOG-CLBP descriptors are proposed [32], [33], Principal Component Analysis (PCA) is also used to make further dimensional space reduction on these fusion descriptors [34], such as Linear Discriminant Analysis (LDA) [35], [36], Fisher's Linear Discriminant Analysis (FLD) [37], and Independent Component Analysis (ICA) [38]. Li *et al.* performed the two-dimensional histogram constituted by LBP to make wood defect classification [39]. However, both of the deep mining algorithms and feature fusion algorithms have the limitations in scaling transformation, and it is also difficult to reduce the feature dimension of descriptors. Although deep learning has made outstanding contributions in fault detection during recent years, just like using the enhanced network based on SSD to deal with faulty cable fixtures [40]. There are also many problems with computational complexity and difficulty in parameter adjustment [41], especially when the number of faulty fixture data set is not enough. This leads to the fact that local features based on statistical learning seem to be able to better achieve cable fixture detection [42].

III. PROPOSED METHOD

CS-LBP and ARCS-LBP operators only roughly explain that there are gradient differences in some diameter directions within the circle neighborhood, while none of them point out the specific gradient direction of the pixel pairs. Therefore, this paper proposes a bi-directional triplet model that is based on the fundamental idea about calculating gray difference value of center-symmetric pixel pair. The center-symmetric pixel pair and its center pixel are combined into a triplet, and the specific gradient direction is determined by calculating the ratio of the difference between the center pixel and the pixels on both sides of this triplet. Since the output of this step is a binary sequence and no decimal conversion is performed, the number of channels in resulting feature maps will be the same as the number of sampling points. Then we will implement the scaling transformation of the multi-channel binary feature maps. The final MCG-LBP descriptor is constructed by extracting continuous gradient features from a series of down-sampling maps on different scales, and Fig. 4 shows the overall process of feature extraction. In this section, the involved approaches will be described.

A. THE BI-DIRECTIONAL TRIPLET

For the purposes of convenient presentation, the same sampling conditions as the mentioned LBP and CS-LBP above are selected when calculating BDT value. In the circular
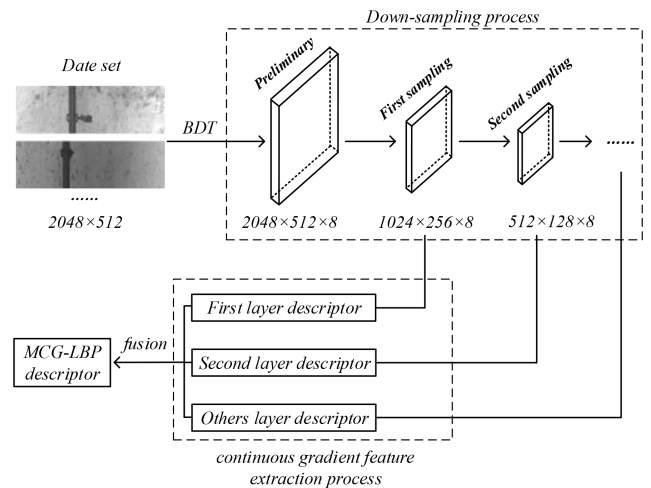

FIGURE 4. The flow chart of extracting MCG-LBP descriptor.

domain, the number of sampling pixels P is 8, and the sampling radius R is 1. Firstly, the bilinear interpolation is used to obtain the gray values of all sampling pixels in a circular neighborhood. In the triplet, the grayscale difference between the center pixel and the two center-symmetric sampling pixels are calculated respectively. Then the ratio of these two differences is compared with the threshold to determine the specific gradient direction. Finally, the gradient directions of all triples in the circular domain are calculated, and the output result of BDT is obtained. The complete calculation process is shown in Fig. 5, and we will explain these four groups of triples in detail.

The calculation of BDT always uses triplet as the basic unit, and the values of u, v, w for each group need to be calculated in advance. For Group0, the result of $-\frac{u_0}{v_0}$ is negative value, so the two count bits of 0 and 4 corresponding to the gradient directions of 0° and 180° are both marked as 0, which means the triplet does not output gradient direction. For Group1, the value of $-\frac{u_1}{v_1}$ is 1.5 that is within the range of threshold T, and the other constraint w_1 is a positive value. So, the count bit of 1 is marked as 1 and the count bit of 5 is marked as 0, which means the output gradient direction of this triplet is 45°. In order to express the calculation process more conveniently, the value of T is specified, but it usually needs to be obtained based on experiments. For Group2, although the value of $-\frac{u_2}{v_2}$ is positive, it is out of the range of T, so the two count bits of this triplet are both marked as 0, which also means that the pixel pair has no gradient direction. For Group3, the value of u_3 is 0, so there is no gradient direction for this pair, and the two count bits are also marked as 0. In summary, the output value of this BDT is 00000010. The complete formula of BDT is defined as follow:

$$BDT(x_c, y_c) = [b_0(m), \cdots b_i(m), \cdots, b_{i+(P/2)}(m)$$
$$\cdots, b_{p-1}(m)] \tag{5}$$
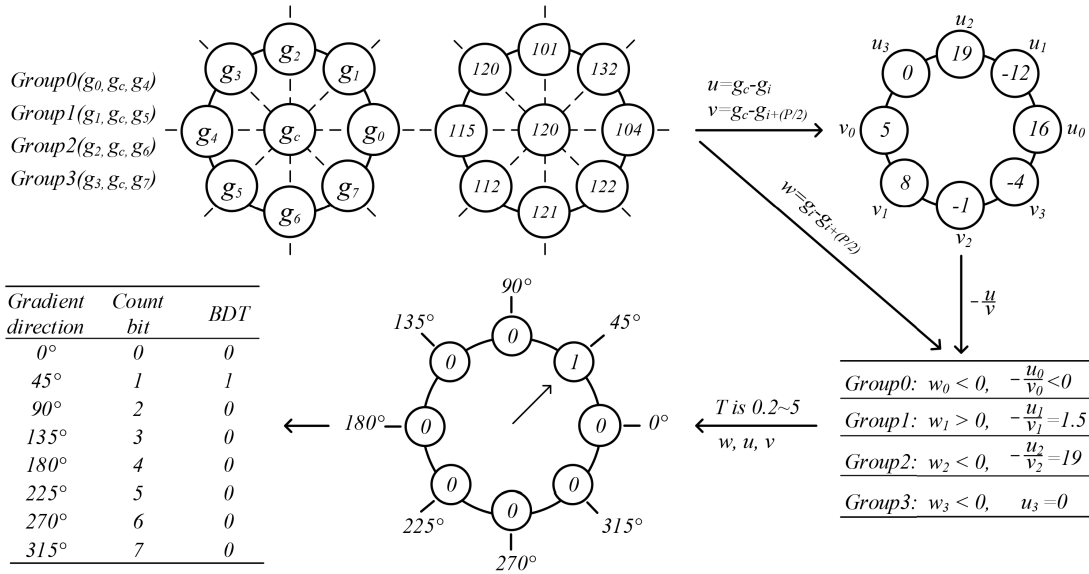$$u = g_c - g_i \tag{6}$$

FIGURE 5. The calculation process of BDT.

$$v = g_c - g_{i+(P/2)} \tag{7}$$

$$w = g_i - g_{i+(P/2)} \tag{8}$$

If $u \neq 0$ and $v \neq 0$:

$$b_i(m) = \begin{cases} 1, & -\dfrac{u}{v} \in T \text{ and } w \geq 0 \\ 0, & otherwise \end{cases} \tag{9}$$

$$b_{i+(P/2)}(m) = \begin{cases} 1, & -\dfrac{u}{v} \in T \text{ and } w < 0 \\ 0, & otherwise \end{cases} \tag{10}$$

Else $u = 0$ or $v = 0$:

$$b_i(m) = \begin{cases} 1, & w > 0 \\ 0, & otherwise \end{cases} \tag{11}$$

$$b_{i+(P/2)}(m) = \begin{cases} 1, & w < 0 \\ 0, & otherwise \end{cases} \tag{12}$$

where the result of $BDT(x_c, y_c)$ is a one-dimensional array whose number of elements is P, and $x_i(m)$ ($i = 0, 1, \cdots, (P/2) - 1$) is the binary value of each bit in array. Repeating above operations on the input image, we can obtain a complete preliminary gradient feature map with eight channels. The value of each pixel in the preliminary feature map uniquely corresponds to a one-dimensional array that is composed of eight binary digits. In short, there are four groups of triplets in the circle neighborhood, and the output of each triplet is only three cases, which means the binary value of each one-dimensional array pixel has 81 (3^4) cases.

B. SCALE TRANSFORMATION AND DIMENSIONALITY REDUCTION

If the continuous gradient feature is extracted directly from a preliminary gradient feature map, the dimensions of the obtained descriptor will be too large. Therefore, an effective dimensionality reduction strategy is proposed to decrease the total number of binary pixel value arrangements. At the same

time, this strategy can also realize the scale transformation by calculating different down-sampling feature maps rather than enlarging the radius of sampling circle or dividing the input image into 2×2 or 4×4. The value range of a single pixel is compressed, and the key gradient is used to represent the trend of overall gradient direction in a Cell unit. The relationship between gradient directions and count bits is shown in Fig. 5, where the count bits distributed from 1 to 8 in a counterclockwise way correspond to eight gradient directions respectively. In Cell unit, the values of all binary array are accumulated one by one according to the same count bit, so that the direction corresponding to the maximum count bit of the accumulated result is the key gradient direction.

In Fig. 6, the Cell is composed of four binary array pixels a, b, c, d whose values are calculated by BDT, and we take the key gradient direction of this Cell as an example. It can be seen from the cumulative result that the values in direction 45° and 90° are maximum, so the corresponding count bits 1 and 2 are marked as 1, and other count bits are marked as 0. The final output coding value of this Cell is 00000110. The calculation formula is defined as follow:

$$X_{ij} = [x_{i0}, x_{i1}, \cdots x_{ij} \cdots, x_{i(P-1)}] \tag{13}$$

$$Y = \sum_{i=0}^{n} X_i$$

$$= [\sum_{i=0}^{n} x_{i0}, \sum_{i=0}^{n} x_{i1}, \cdots \sum_{i=0}^{n} x_{ij} \cdots, \sum_{i=0}^{n} x_{i(P-1)}] \tag{14}$$

$$J = f(Y)$$

$$= f([\sum_{i=0}^{n} x_{i0}, \sum_{i=0}^{n} x_{i1}, \cdots \sum_{i=0}^{n} x_{ij} \cdots, \sum_{i=0}^{n} x_{i(P-1)}]) \tag{15}$$

where X_i is the pixel value of binary array, $i(i = 0, 1, \cdots, n)$ is the serial number of binary array in the Cell, n is the total
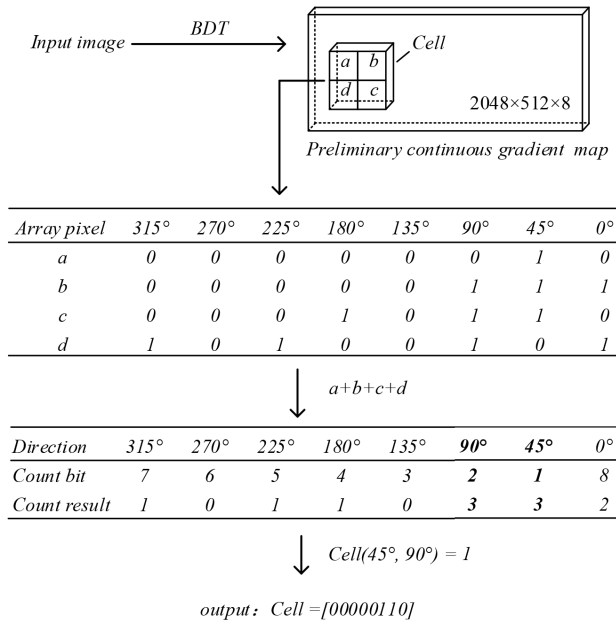
Array pixel table:

Array pixel	315°	270°	225°	180°	135°	90°	45°	0°
a	0	0	0	0	0	0	1	0
b	0	0	0	0	0	1	1	1
c	0	0	0	1	0	1	1	0
d	1	0	1	0	0	1	0	1

$a+b+c+d$

Direction	315°	270°	225°	180°	135°	**90°**	**45°**	0°
Count bit	7	6	5	4	3	**2**	**1**	8
Count result	1	0	1	1	0	**3**	**3**	2

$Cell(45°, 90°) = 1$

output：Cell =[00000110]

FIGURE 6. Calculating the characteristics of the key gradient in a Cell, and the output value is a one-dimensional array.

FIGURE 7. The output feature map of each layer during down sampling process.

Channel number	8	1	2	3	4	5	6	7
Reference channel	×	×	✓	×	✓	×	×	×

Channel number	**reference**	left	right	**reference**	left	right
	2	1	3	**4**	3	5
Sequence length	4	1	2	2	2	2
Coding number	**2**	412		**4**	222	

Output : **2** 412

FIGURE 8. Calculate the characteristics of the continuous gradient in a Block and the channel is a loop from 1 to 8.

number of binary array in the Cell, and $j(j = 0, \ldots, P-1)$ is the count bit which also represents the channel serial number in preliminary gradient feature map and down-sampling feature maps. Y is the accumulation result of all one-dimensional arrays in the Cell, and the function $f(Y)$ outputs the value J that corresponds to the channel with the maximum accumulation. Repeat all above operations on the entire preliminary gradient feature map to obtain the first down-sampling feature map. Note that one Cell can output multiple key gradients, which means that there may be not only one channel that output digital 1. Then perform the same iteration on the first down-sampling feature map to get the second down-sampling feature map.

In down-sampling process, the stride is always 2, and the Cell size is 2×2. The size of down-sampling feature map in each layer corresponds to the size of sub-image segmented by traditional methods like 2×2, 4×4, and 8×8. Non-maximum value suppression is performed according to the frequency of key gradients in a Cell, which is also different from the maximum pooling in deep learning. Fig. 7 is a schematic diagram of the output result for each down-sampling layer. Except that the input image has only one channel, other down-sampling feature maps and the preliminary gradient feature map have eight channels. Through scale transformation, typical contour features such as edges and corners will be preserved, and some isolated noise points or wall cracks will be removed during down-sampling process.

C. THE EXTRACTION PROCESS OF CONTINUOUS GRADIENT FEATURE

Deep mining operators and fusion feature operators of LBP almost only perform texture analysis based on the gray level difference by local sampling points, but ignore the similarity
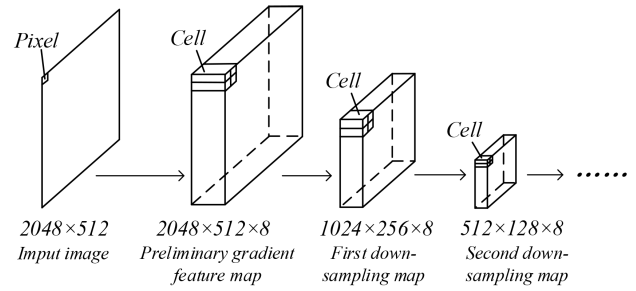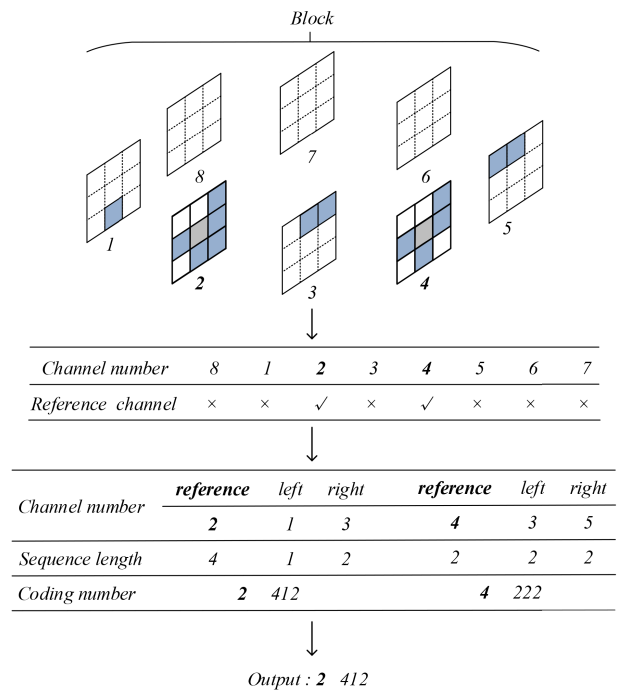
difference between adjacent gradients. Therefore, we propose continuous gradient feature, which can excellently respond to the relevant information between key gradients and related adjacent gradients that appear on multiple down-sampling feature maps. In order to distinguish from the Cell unit in above down-sampling process, a 3×3 window is defined in extraction process, and the unit is named as Block. The number of channels in a Block is also eight, and the stride is set to 1 for feature extraction.

The specific calculation process of extracting continuous gradient features is shown in Fig. 8. Firstly, the reference channels whose central point value is 1 need to be determined in this Block unit. If there is no central point with value 1 in all channels, the Block does not output continuous gradient feature. It can be seen that only channel 2 and channel 4 meet the condition, so these two channels will be used as a reference to extract continuous gradient features independently. The

maximum continuous sequence length composed of eight neighborhood sampling points in reference channel 2 is four, and the maximum continuous sequence length of adjacent channel 1 and channel 3 is one and two respectively. Then the coding number 2412 belonging to reference channel 2 is obtained. In the same way, the coding number of reference channel 4 is 4222. Notice that the maximum continuous sequence length of reference channel 4 is less than three, which means that the reference channel 4 does not output value, so the final output of this block is only 2413. The constraints of calculating continuous gradient features are summarized as the following three points: 1) At least one channel has a center value with 1 in a Block with eight channels. If not, then this Block does not output continuous gradient feature. 2) The maximum continuous sequence length must not be less than three in reference channel, which can ensure the dominance of reference gradient feature. 3) Since the reference channel and the adjacent channel can be converted mutually in different Blocks where the gradient characteristics can be complemented with each other, so when the maximum continuous sequence length of adjacent channels is over three, the length is still marked as 3. The specific sequence information of adjacent channels does not affect its performance when used as reference channel information in other Blocks. Therefore, there are $768(768 = 8 \times 6 \times 4 \times 4)$ encoding typies that can meet the above conditions.

IV. EXPERIMENT AND DISCUSSION

The performance of MCG-LBP operator is validated by a comprehensive comparison with a large number of state-of-the-art variant methods based on LBP and non-LBP operators. The superior quality of our proposed method is verified by quantitative and qualitative evaluations using leaky cable fixture data sets. In order to obtain the best MCG-LBP descriptor, the extraction efficiency and the classification accuracy of each single-layer feature are also compared. Furthermore, all experiments are implemented by using PYTHON 3.7, the CPU used in experimental running environment is Intel(R) i5-4210H, the running memory is 8GB, and the GPU is NVIDIA GTX 850.

A. EXPERIMENTAL SETUP

1) LEAKY CABLE FIXTURE DATA SET

In this section, the leaky cable fixture data set is introduced in detail. As is shown in Fig. 9 (a), a CMOS dual-line camera is used to take the entire picture during the whole train journey, and the size of obtained picture is 2048×512. Since the camera has its own light source and is open on during the entire working time, the influence of illumination on the image is reduced. It can be found clearly that the brightness of different backgrounds varies largely even in the same tunnel section, just like Fig. 9 (b) and (c), which causes the threshold setting to be quite troublesome. Fig. 10 (a), (b), and (c) show that the shapes of normal cable fixtures are relatively simple, especially for their contour characteristics are tend to be the
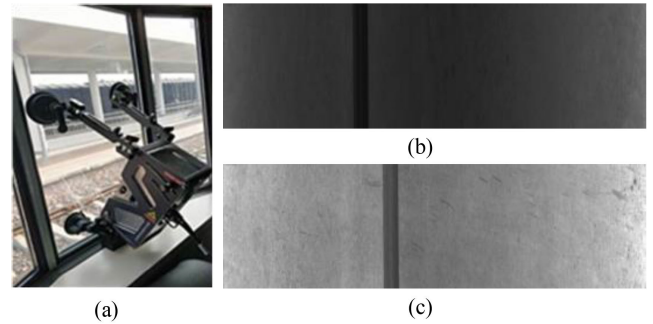

FIGURE 9. (a) is the COMS dual-line camera. (b) and (c) are the backgrounds of leaky cable captured by high-speed camera in tunnel.
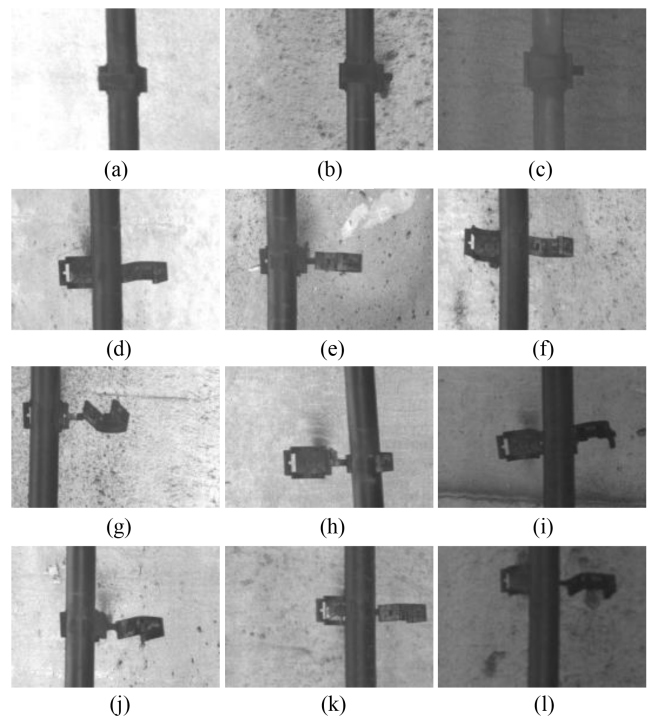

FIGURE 10. Images from (a) to (c) are the samples of normal cable fixture. Images from (d) to (l) are the samples of faulty cable fixture.

same basically. While the shapes of faulty cable fixtures shown in Fig. 10 from (g) to (o) are relatively complicated, it is difficult to define the standard of damage, especially the number of faulty cable fixtures is not large. Therefore, as long as there is a certain difference between the appearance of normal fixture, we regard it as a faulty one. MCG-LBP operator does not require lots of faulty cable fixtures to train the model. In the experiment, there are 900 images of normal fixtures and 195 faulty fixtures. We have selected 70% of the normal fixtures and fault fixtures respectively as the training set, and the remaining 30% of the normal fixtures and fault fixtures are as the test set.

2) THRESHOLD OF BDT

In order to obtain the most suitable threshold of BDT, a total of 50 images with global average grayscale distributed in different intervals are selected. Then the specific corresponding

relationship between the global average grayscale and threshold is analyzed through the imaging effect of leaky cable fixture contour. Gaussian filtering is performed on the input image, and the imaging effects of their respective preliminary gradient direction feature maps under different threshold conditions are compared. To make the process as intuitive as possible, the one-dimensional binary array is converted into a decimal gray value. Although the range of threshold should be as small as possible, it still needs to meet the following two requirements: 1) It must be ensured that the preliminary gradient direction feature map can show the outline of the leaky cable fixture completely and clearly. 2) It can filter the background noise of the tunnel wall to a certain extent.

For the convenience of display, the cable fixture part is cut from entire image. It can be seen from Tab. 1 that the larger range of threshold, the better display effect in low brightness images, while more interference will appear in high brightness images. Although a small range of threshold can effectively remove background noise, its expression of contour features will also be weakened. Therefore, it is necessary to set the adaptive threshold according to the global gray average value, which is defined as follows:

$$T(x) = \begin{cases} 0.1\tilde{\ }10, & x < 80 \\ 0.125\tilde{\ }8, & 80 \leq x < 120 \\ 0.2\tilde{\ }5, & 120 \leq x < 160 \\ 0.25\tilde{\ }4, & 160 \leq x < 200 \\ 0.4\tilde{\ }2.5, & x \geq 200 \end{cases} \quad (16)$$

where x is the global grayscale average value of input image. From Tab. 1, it can be seen that the ratio of gray-scale difference is basically maintained at around 1. For images with high brightness, the imaging effect of different threshold ranges is generally better, so the threshold range can be smaller, which

TABLE 1. Analysis of different gray-threshold.

image					
average grayscale	216	186	133	99	63
0.1~10					
0.125~8					
0.2~5					
0.25~4					
0.4~2.5					
0.5~2					
0.6~1.67					
0.8~1.25					

can filter out more noise. For images with low brightness, the imaging effect with a smaller threshold range will cause the loss of contour feature. To ensure that the contour feature can be displayed completely, the threshold range needs to be larger at this time. Then, the threshold range is set to five different levels, and different average gray levels correspond to different thresholds.

B. EXPERIMENTAL COMPARISON RESULTS

The following two aspects of experimental research are carried out. The first experimental study is to compare between different down-sampling layers to obtain the best fusion descriptor of MCG-LBP, and the other is to compare with LBP variant and non-LBP operators to verify the superiority of MCG-LBP.

1) COMPARISON BETWEEN DIFFERENT DOWN-SAMPLING LAYERS

The down-sampling process will iterate out many feature maps, even if their size will decrease gradually, but compared with directly dividing the feature maps into 4×4 or 2×2, the process still has a certain amount of calculation. Therefore, it is necessary to comprehensively consider computational efficiency and detection accuracy to extract the most cost-effective descriptor. In the experiment, a total of four iterations are performed, then continuous gradient features are extracted from the four down-sampling feature maps respectively. Their histogram features are displayed in Fig. 11, it can be seen that the proportions of dominant features remain unchanged basically, but other features have changed considerably. The subsequent sampling map contains fewer continuous gradient features than previous layer, which is the main reason why only four iterations are performed. In down-sampling progress, the stride of Cell is 2, and in continuous gradient feature extraction process, the stride of Block is 1. Therefore, it can be inferred that the relationship of occupancy time between down-sampling process and feature extraction process is as follow:

$$Time_{(sampling)} \leq \frac{1}{4} Time_{(extraction)} \quad (17)$$

Due to the different strides of these two processes, the feature extraction process takes more time than the down-sampling process, so it is critical to select appropriate sampling maps to extract continuous gradient features.

Average Recall rate, average Precision rate and average Accuracy are used as the evaluation indicators of classification result. Recall refers to the proportion of these fixtures that are judged correctly as faulty in all faulty fixtures, Precision refers to the proportion of these fixtures that are judged correctly as faulty in all fixtures that are judged as faulty, and Accuracy refers to the proportion of these fixtures that are judged correctly in all fixtures. The specific definitions are as follows:

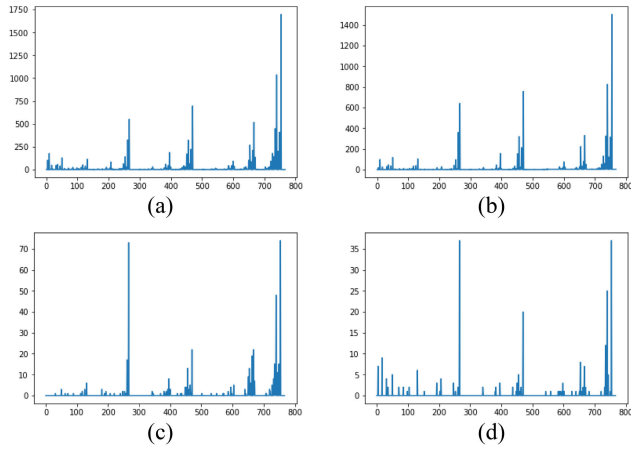$$Recall = \frac{TP}{TP + FN} \quad (18)$$

FIGURE 11. The histogram features of different down-sampling layer. (a) The first sampling map. (b) The second sampling map. (c) The third sampling map. (d) The fourth sampling map.

$$Precision = \frac{TP}{TP + FP} \quad (19)$$

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (20)$$

where *TP* represents the number of faulty fixtures identified as faulty actually, *TN* represents the number of normal fixtures identified as normal, *FP* represents the number of normal fixtures identified as faulty, and *FN* represents the number of faulty fixtures identified as normal. As we can see from Tab. 2, the time occupancy of extracting each layer feature conforms to (17). It also shows the performance of extracting continuous gradient features from each single down-sampling map, including the evaluation indicators of Recall and Precision. The performance of identifying faulty cable fixtures is more in line with actual needs, so Recall is the primary factor. Although it will take less time to extract continuous gradient features from subsequent down-sampling maps, the performance of classification is reduced to a certain extent.

TABLE 2. Comparison of occupied time, recall, and precision for the different single down-sampling map.

Layer	Recall	Precision	Time(s)
1	0.625	0.625	2.864
2	**0.833**	0.769	1.328
3	0.727	**0.889**	0.847
4	0.667	0.857	0.665

There are a total of six cascading types to combine these four single-layer features in pairs, and the detection results are shown in Tab. 3. Among that, 1-2 means to connect the first layer descriptor and the second layer descriptor together, and others are the same. There is not only a big gap between Recall and Precision of these fusion features, but also the time taken by different cascading types is very different, especially for these descriptors including the first layer. In addition, the three-level connections such as 1-2-3 that require more

TABLE 3. Comparison of recall and precision for the fusion feature from different down-sampling map.

Fusion type	Recall	Precision
1-2	0.727	0.889
1-3	0.875	0.778
1-4	0.615	0.727
2-3	**0.923**	0.857
2-4	0.917	0.786
3-4	0.889	**0.941**

time and its feature dimensions will also increase. The Recall of 2-3 reaches 92.3% which is the highest compared to other descriptors. Finally, considering the classification performance and time cost, the fusion type of 2-3 is defined as MCG-LBP, which will be applied in comparison with other variant operators.

2) COMPARISON WITH LBP VARIANT AND NON-LBP OPERATORS

The fusion descriptor of MCG-LBP is compared with the up-to-data methods on cable fixture data set. Some supplementary instructions of these LBP variant and non-LBP operators are as follows:

1) CLBP_$S_{1,8}^{riu2}$/_$M_{1,8}^{riu2}$/C is combined into hybrid distribution where sign component (S), magnitude component (M), and average grayscale threshold (C) are cascaded in sampling neighborhood. The encoding method follows locally rotation invariant uniform pattern (LBP$_{1,8}^{riu2}$), and the feature dimension of each sub-image is 200 ($200 = (2(8 + 2)^2)$) [7]. So the final feature dimension of output descriptor is 3200 ($3200 = 200 \times 4 \times 4$). For the convenience of expression, CLBP is referred to this method specifically, and the calculation of CLTP is the same.

2) As for CS-LBP operator, the input images are divided into 8×8, and the resolution of sub-image is 256×64. Since the feature dimension of each sub-image of CS-LBP is 16, the final feature dimension is expanded to 1024 ($1024 = 16 \times 8 \times 8$).

3) As for HOG operator, the bins of each gradient component are 9, and each Block size is 128×32, which is very similar to dividing image into 16×16. There are 256 windows on the whole input image to count the histogram features, so the dimension is 2304 ($2304 = 256 \times 9$).

4) The length of LDP for a reference pixel is 32 bit which is too large. So, we just calculate it to second order, and the uniform pattern is used to code in derivative space of different directions. The input images are divided into 4×4, and the feature dimension is 3776 ($3776 = 16 \times 4 \times 59$).

5) As for LDGP operator, we just calculate it to second order, and divided the input image into 4×4. The feature dimension is 4096 ($4096 = 16 \times 4 \times 2^6$).

6) The masks of LDTP are the same as [23]. To prevent too many dimensions, we divide the image into 2×2, and the dimension is 4096 ($4096 = 4 \times 2 \times 2^9$).

7) HOG_LBP, HOG_CS-LBP, and HOG_CLBP fusion features refer to concatenating the descriptor obtained at before, and their dimensions need to accumulate directly.

8) For other variant patterns, we perform a 4×4 division with non-overlapping, the resolution of sub-image is 512×128, the sampling radius is 1, and the sampling point is 8.

The volumes of normal fixture set and faulty fixture set are obviously unbalanced, so another evaluation index AUC needs to be added. It is widely used in machine learning two-class classification, which has a greater tolerance for samples unbalance. In order to explain specifically, the following formula needs to be explained firstly:

$$FPR = \frac{FP}{FP + TN} \qquad (21)$$

where TN represents the number of normal fixtures identified as normal actually. The curve drawn with FPR as abscissa and Recall as ordinate is ROC, and the area enclosed by ROC curve and coordinate axis from 0 to 1 is AUC value.

From the comparison results in Tab. 4, it can be seen that the MCG-LBP operator proposed in this paper has significant advantages in faulty cable fixtures classification. Recall and Precision reached 92.3% and 85.7% respectively, and the AUC value is also the highest in the ROC curve shown in Fig. 12. Even though the fusion feature of HOG_LBP can make Recall reach 85.4%, while Precision is only 65.4%, which means that a large number of normal cable fixtures are misjudged as faulty fixtures. As for HOG operator, it is still a problem of low Recall only 64.5%, even if its Precision rate reaches 75.7%, which means that a large number of faulty cable fixtures are misjudged as normal fixtures. On the other hand, from the perspective of feature dimensions, the dimension of MCG-LBP is only slightly higher than that of
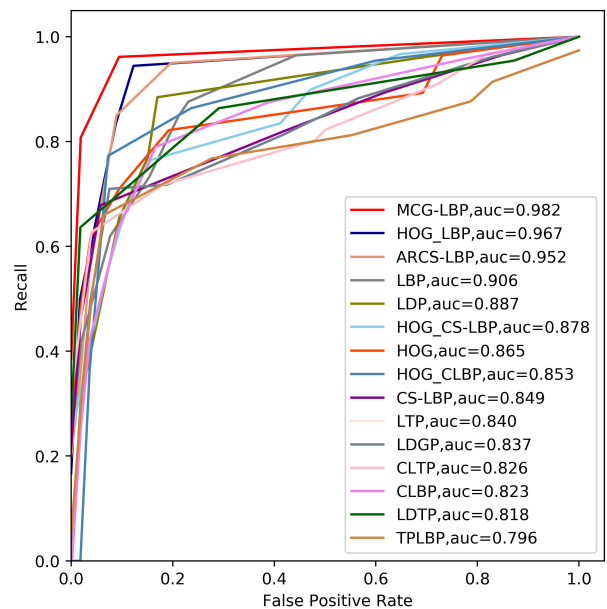

FIGURE 12. The performance of ROC curve with different methods.

CS-LBP, and is far lower than other LBP variant and non-LBP operators, however the detection performance is greatly improved. The size of second down-sampling map is 512×128, and the size of third down-sampling map is 256×64. If we divide input image by 4×4 directly, there are sixteen sub-images with the size of 512×128. Compared with image segmentation, extracting features from the down-sampling image can obviously reduce the computational cost. The main reason is that our dimensionality reduction strategy shrinks the area of extracting features by down-sampling iterations, instead of extracting from sub-images one by one directly, which also shortens the length of final descriptor effectively. It is not difficult to find that even the continuous gradient features obtained from each single layer, and their classification results also have high Recall and Precision rates such as 92.3% and 94.1%. The second and third down-sampling maps not only realize the scale transformation of feature map through iteration, but also contain the key gradient information of cable fixture contour. Therefore, the MCG-LBP descriptor proposed in this paper has obvious advantages.

V. CONCLUSION

This paper proposes MCG-LBP operator for leaky cable fixture detection in the railway tunnel. In order to extract more specific gradient information in the circle neighborhood, the model of bi-directional triplet is introduced. Then, a dimensionality reduction strategy is adopted to implement successive down-sampling iterations, which not only realizes the scale transformation of cable fixture features, but also reduces the calculation for extracting continuous gradient features and greatly increases the computational efficiency. Finally, to further enhance the single layer descriptor, the continuous gradient features extracted from different downs-sampling

TABLE 4. Comparison of recall and precision with other LBP variant and non-LBP operators.

Method	Accuracy	Recall	Precision	Dimension
LBP [7]	0.886	0.745	0.649	4096
CS-LBP [17]	0.856	0.759	0.647	**1024**
CLBP [18]	0.855	0.638	0.618	3200
LTP [16]	0.861	0.718	0.619	8192
LDTP [23]	0.878	0.583	0.701	4096
CLTP [19]	0.863	0.625	0.685	6400
HOG [29]	0.894	0.643	0.818	2304
TPLBP [22]	0.887	0.706	0.770	4096
LDP [21]	0.876	0.716	0.655	3776
LDGP [27]	0.848	0.668	0.621	4096
ARCS-LBP [20]	0.886	0.757	0.665	4096
HOG_CS-LBP	0.864	0.697	0.641	3328
HOG_CLBP [31]	0.894	0.734	0.674	5504
HOG_LBP [30]	0.909	0.854	0.654	6400
MCG-LBP	**0.947**	**0.926**	**0.835**	1536

maps are cascaded to promote classification performance. MCG-LBP has a great improvement in both detection performance and feature dimension, especially decreasing the dependence on the faulty data set, which has superiority in leaky cable fixture detection. The proposed algorithm also provides a reference for the detection of small targets in large images such as airports and subway security inspections, and remote sensing image detection. In further research, combining MCG-LBP with deep learning is a topic worthy of study. Especially to realize a real-time detection system for leaky cable fixtures.

REFERENCES

[1] C. Tian, X, Liu, and L. H. Liu, "Research on the hanging height of GSM-R leaky coaxial cable in the railway tunnel," *J. Railway Eng. Soc.*, vol. 36, no. 10, pp. 82–87, Nov. 2019.

[2] T. Wang, "Development status and prospect of intelligent construction technology for railway tunnel in China," *China Railway.*, vol. 40, pp. 1–9, Jan. 2020.

[3] S. J. Yi, "A vehicle detection system based on visual image of the cable fault," M.S. thesis, Dept. Electron. Eng., Beijing Univ. Chem. Technol., Beiging, China, 2015.

[4] Y. Z. Zhang, P. L. Yang, and W. X. Li, "Detection of leakage cable fixture in tunnel based on feature enhancement SSD," *J. Beijing Jiaotong Univ.*, vol. 45, no. 2, pp. 105–110, 2021.

[5] H. Y. Ma, "Research and realization of tunnel leaky cable fixture fault detetion system based on deep learning," M.S. thesis, Dept. Electron. Eng., Hebei Univ. Sci. Technol., Hebei, China, 2020.

[6] L. Liu, J. Chen, P. Fieguth, G. Zhao, R. Chellappa, and M. Pietikäinen, "From BoW to CNN: Two decades of texture representation for texture classification," *Int. J. Comput. Vis.*, vol. 127, no. 1, pp. 74–109, Jan. 2019.

[7] T. Ojala, M. Pietikäinen, and T. Mäenpää, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971–987, Jul. 2002.

[8] M. Zhu, W. Hu, and N. C. Kar, "Multi-sensor fusion-based permanent magnet demagnetization detection in permanent magnet synchronous machines," *IEEE Trans. Magn.*, vol. 54, no. 11, pp. 1–6, Nov. 2018.

[9] T. H. Rassem, B. Ee Khoo, M. F. Mohammed, and N. M. Makbol, "Medical, scene and event image category recognition using completed local ternary patterns (CLTP)," *Malaysian J. Comput. Sci.*, vol. 30, no. 3, pp. 200–218, Sep. 2017.

[10] M. A. Abuzneid and A. Mahmood, "Enhanced human face recognition using LBPH descriptor, multi-KNN, and back-propagation neural network," *IEEE Access*, vol. 6, pp. 20641–20651, 2018.

[11] N. Alpaslan and K. Hanbay, "Multi-resolution intrinsic texture geometry-based local binary pattern for texture classification," *IEEE Access*, vol. 8, pp. 54415–54430, 2020.

[12] C. Li, Y. Huang, and L. Zhu, "Color texture image retrieval based on Gaussian copula models of Gabor wavelets," *Pattern Recognit.*, vol. 64, pp. 118–129, Apr. 2017.

[13] L. Liu, S. Lao, P. W. Fieguth, Y. Guo, X. Wang, and M. Pietikäinen, "Median robust extended local binary pattern for texture classification," *IEEE Trans. Image Process.*, vol. 25, no. 3, pp. 1368–1381, Mar. 2016.

[14] D. N. L. Yao, A. Bade, N. M. Suaib, and H. A. B. Sulaiman, "Digital image enhancement using enhanced detail and dehaze technique (DDE)," *Adv. Sci. Lett.*, vol. 24, no. 3, pp. 1559–1561, Mar. 2018.

[15] I. Choi and D. Kim, "A variety of local structure patterns and their hybridization for accurate eye detection," *Pattern Recognit.*, vol. 61, pp. 417–432, Jan. 2017.

[16] X. Tan and B. Triggs, "Enhanced local texture feature sets for face recognition under difficult lighting conditions," *IEEE Trans. Image Process.*, vol. 19, no. 6, pp. 1635–1650, Jun. 2010.

[17] M. Heikkilä, M. Pietikäinen, and C. Schmid, "Description of interest regions with center-symmetric local binary patterns," in *Computer Vision, Graphics and Image Processing*. Berlin, Germany: Springer, 2006, pp. 58–69.

[18] Z. Guo and D. Zhang, "A completed modeling of local binary pattern operator for texture classification," *IEEE Trans. Image Process.*, vol. 19, no. 6, pp. 1657–1663, Jan. 2010.

[19] T. H. Rassem and B. E. Khoo, "Completed local ternary pattern for rotation invariant texture classification," *Sci. World J.*, vol. 2014, pp. 1–10, Apr. 2014.

[20] Y. E. Merabet, Y. Ruichek, and A. E. Idrissi, "Attractive-and-repulsive center-symmetric local binary patterns for texture classification," *Eng. Appl. Artif. Intell.*, vol. 78, pp. 158–172, Feb. 2019.

[21] B. Zhang, Y. Gao, S. Zhao, and J. Liu, "Local derivative pattern versus local binary pattern: Face recognition with high-order local pattern descriptor," *IEEE Trans. Image Process.*, vol. 19, no. 2, pp. 533–544, Feb. 2010.

[22] L. Wolf, T. Hassner, and Y. Taigman, "Effective unconstrained face recognition by combining multiple descriptors and learned background statistics," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 10, pp. 1978–1990, Oct. 2011.

[23] I. E. Khadiri, A. Chahi, Y. E. Merabet, Y. Ruichek, and R. Touahni, "Local directional ternary pattern: A new texture descriptor for texture classification," *Comput. Vis. Image Understand.*, vol. 169, pp. 14–27, Apr. 2018.

[24] T.-N. Nguyen, B. Le, and K. Miyata, "Effective feature description using intensity order local binary pattern," in *Proc. Int. Conf. Cyberworlds*, Oct. 2013, pp. 175–182.

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

[26] Z. Pan, H. Fan, and L. Zhang, "Texture classification using local pattern based on vector quantization," *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 5379–5388, Dec. 2015.

[27] S. Chakraborty, S. K. Singh, and P. Chakraborty, "Local directional gradient pattern: A local descriptor for face recognition," *Multimedia Tools Appl.*, vol. 76, no. 1, pp. 1201–1216, Jan. 2017.

[28] S. Chakraborty, S. K. Singh, and P. Chakraborty, "Local gradient hexa pattern: A descriptor for face recognition and retrieval," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 28, no. 1, pp. 171–180, Jan. 2018.

[29] M. Subrahmanyam, R. P. Maheshwari, and R. Balasubramanian, "Local maximum edge binary patterns: A new descriptor for image retrieval and object tracking," *Signal Process.*, vol. 92, no. 6, pp. 1467–1479, 2012.

[30] S. Murala, R. P. Maheshwari, and R. Balasubramanian, "Local tetra patterns: A new feature descriptor for content-based image retrieval," *IEEE Trans. Image Process.*, vol. 21, no. 5, pp. 2874–2886, May 2012.

[31] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2005, pp. 886–893.

[32] Y. Wan, H. H. Li, and K. F. Wu, "Fusion with layered features of LBP and HOG for face recognition," *J. Comput.-Aided Des. Comput. Graph.*, vol. 27, no. 4, pp. 640–650, Apr. 2015.

[33] J. Yao and F. Q. Yu, "Pedestrian detection based on combination of candidate region location and HOG-CLBP features," *Laser Optoelectron. Prog.*, vol. 58, no. 2, pp. 165–172, Feb. 2021.

[34] Y. Zhang, C. Xu, C. Li, H. Yu, and J. Cao, "Wood defect detection method with PCA feature fusion and compressed sensing," *J. Forestry Res.*, vol. 26, no. 3, pp. 745–751, Mar. 2015.

[35] L. Chen, H. Liao, M. Ko, J. Lin, and G. Yu, "A new LDA-based face recognition system which can solve the small sample size problem," *Pattern Recognit.*, vol. 33, no. 10, pp. 1713–1726, 2000.

[36] O. Déniz, G. Bueno, J. Salido, and F. De la Torre, "Face recognition using histograms of oriented gradients," *Pattern Recognit. Lett.*, vol. 32, no. 12, pp. 1598–1603, 2011.

[37] P. N. Belhumeur, J. P. Hespanha, and D. Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 711–720, Jul. 1997.

[38] M. S. Bartlett, J. R. Movellan, and T. J. Sejnowski, "Face recognition by independent component analysis," *IEEE Trans. Neural Netw.*, vol. 13, no. 6, pp. 1450–1464, Nov. 2002.

[39] S. Li, D. Li, and W. Yuan, "Wood defect classification based on two-dimensional histogram constituted by LBP and local binary differential excitation pattern," *IEEE Access*, vol. 7, pp. 145829–145842, 2019.

[40] Y. Z. Zhang, P. L. Yang, and W. X. Li, "Leaky coaxial cable fixture detection based on improved SSD algorithm," *Laser Optoelectron. Prog.*, vol. 57, no. 14, pp. 1–12, Feb. 2020.

[41] C.-X. Ren, Z. Lei, D.-Q. Dai, and S. Z. Li, "Enhanced local gradient order features and discriminant analysis for face recognition," *IEEE Trans. Cybern.*, vol. 46, no. 11, pp. 2656–2669, Nov. 2016.

[42] J. Chen, V. M. Patel, L. Liu, V. Kellokumpu, G. Zhao, M. Pietikäinen, and R. Chellappa, "Robust local features for remote face recognition," *Image Vis. Comput.*, vol. 64, pp. 34–46, Aug. 2017.

ZHOUCHEN SONG received the B.S. degree from the School of Electrical and Engineering, Wuhan Polytechnic University, China, in 2018. He is currently pursuing the M.A.Eng. degree with the School of Information Science and Technology, Shijiazhuang Tiedao University. His research interests include image processing, multiple object tracking, and surveillance video analysis and processing.

YUNZUO ZHANG received the Ph.D. degree from the School of Information and Electronics, Beijing Institute of Technology, China, in 2016. He was a Visiting Scholar with California State University, in 2018. He is currently a Professor with the School of Information Science and Technology, Shijiazhuang Tiedao University, Shijiazhuang, Hebei, China. His research interests include video processing, image processing, and radar signal processing.

WEI GUO received the B.S. degree from the Internet of Things Engineering, Bao Ding University, China, in 2020. She is currently a Graduate Student with the Department of Computer Science and Technology, Shijiazhuang Tiedao University. Her research interests include image processing and surveillance video analysis and processing.

• • •