IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received September 25, 2020, accepted October 9, 2020, date of publication October 20, 2020, date of current version October 29, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3032406

Person-Independent Facial Expression Recognition Based on Improved Local Binary Pattern and Higher-Order Singular Value Decomposition

YING HE AND SHUXIN CHEN
Department of Computer Science and Technology, Ren'ai College of Tianjin University, Tianjin 300000, China
Corresponding author: Shuxin Chen (shuxinfriend@126.com)

This work was supported in part by the National Natural Science Foundation of China under Grant U2031142, and in part by the National Science Foundation for Young Scientists of China under Grant 11803013.

ABSTRACT The recognition rate of person-independent facial expression is generally not high, which limits the practical application of facial expression recognition. Aiming at this problem, this paper analyzes the reasons for the low recognition rate of person-independent facial expression, and proposes a recognition algorithm of person-independent facial expression based on improved LBP (Local Binary Pattern) and HOSVD (Higher-Order Singular Value Decomposition). The algorithm has the following contributions of facial expression recognition framework. In the stage of facial expression feature extraction, the transient features extracted by LDP(Local Directional Pattern) and the persistent features extracted by CBP(Centralized Binary Pattern) are integrated to improve the discrimination of facial expression features. Moreover, in the stage of facial expression classification and recognition, the traditional nearest neighbor classification is changed into k-nearest neighbor pre-classification, and the regional energy calculated by HOSVD is used to determine the similarity of two images for secondary classification. Finally, in the extended Cohn-Kanade dataset and Oulu-CASIA NIR&VIS facial expression database, the theoretical analysis and experimental results show that the method has better recognition effect for solving the problem of person-independent facial expression recognition.

INDEX TERMS Facial expression recognition, higher-order singular value decomposition, person-independent, local directional pattern, centralized binary pattern.

I. INTRODUCTION

With the rapid development of artificial intelligence, facial expression recognition has become one of the research hotspots. It has great application value in intelligent monitoring [1], [2], intelligent medicine [3], and online teaching effect analysis [4] and so on.

In recent years, scholars have made many improvements in the expression recognition algorithm to improve the recognition rate of facial expression. At present, the recognition rate has also been greatly improved. But in some experimental data, the training data contains or is equal to the test data, which represents the same expression of the same person has been seen before, such as [5]–[7].It's hard to apply in a real environment. When the training set and the test set are different expression data, the recognition rate is greatly reduced. This situation greatly restricts the practical application of facial expression recognition.

For person-independent facial expression recognition algorithm, Wang and Ahuja [8] and Sun and Liu [9] proposed to decompose the facial expression image by High-Order Tensor Model, and extract the expression features irrelevant to the face appearance, so as to improve the recognition rate. Jiang *et al.* [10] extracted the facial expression texture features of the main facial organs by weighted LBP (Local Binary Pattern), and then combined with sparse

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

representation classification to achieve expression recognition. Hu [11] used High-Order Tensor to separate facial expression features, and combined with expression subspace multi-classifiers integration for classification. Tan and Zhang [12] proposed to calculate the weighted distance of face similarity to remove the differences of expression features caused by individual differences. However, the final recognition effect of these person-independent facial expression recognition algorithms is still not very ideal.

The main difficulty of person-independent facial expression recognition is that the appearance of human face affects the accurate acquisition of expression features. This will greatly affect the effect of expression recognition.

This paper has three contributions of facial expression recognition framework. The first contribution is a new fusion feature extraction method. We find that the single use of texture feature extraction algorithm cannot extract expression features very effectively. And the fusion of whole face transient features extracted by Local Directional Pattern (LDP) and local persistent features extracted by Centralized Binary Pattern (CBP) can enhance the saliency and robustness of expression features.

The second contribution is that in the stage of classification and recognition, this paper proposes a new classification and recognition algorithm, which combines pre-classification with accurate classification. Because the nearest neighbor algorithm (NN) only takes a best matched image as the classification result, which is easy to cause misjudgment. Therefore, the k-NN proposed in this paper extracts the nearest k images as the pre-classification result.

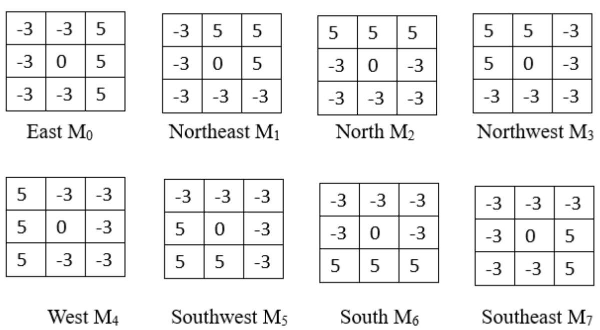

FIGURE 1. **Kirsch operators in eight directions.**

The third contribution is the accurate image classification by high-order singular value decomposition (HOSVD). The images are transformed into transform domain, and the regional energy of two images is calculated to determine the similarity of two images.

In this paper, a face expression recognition algorithm based on improved LBP and HOSVD is proposed. Firstly, the whole face, eyes, eyebrows and mouth regions are segmented by Viola Jones algorithm [13].Secondly the local persistent expression features are obtained by CBP (Centralized Binary Pattern), and the whole transient facial expression features are obtained by LDP (Local Directional Pattern). Then these two features are fused into facial expression features for recognition. Through the improved nearest neighbor algorithm [14] - k nearest neighbor algorithm, the most similar k personal facial expression images are pre-classified. Finally, in order to further distinguish the differences of expression patterns, HOSVD is used for secondary classification, and the regional energy is calculated to determine the similarity of the two images. Thus, k facial expression image can be accurately recognized and the closest classification of facial expression image can be obtained.

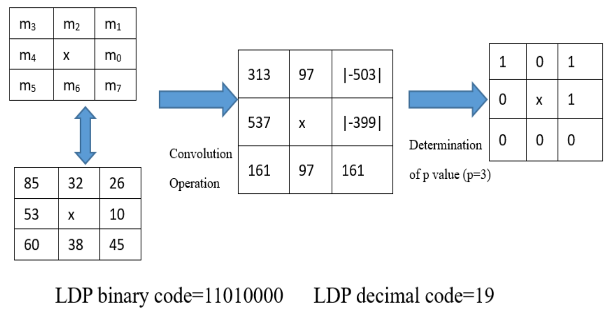

FIGURE 2. **The example of LDP coding process.**

II. FACIAL EXPRESSION FEATURE EXTRACTION

A. FACIAL EXPRESSION FEATURE EXTRACTION BASED ON LDP

LBP (Local Binary Pattern) [15]–[17] operator is usually selected when extracting facial expression texture features, because it can measure and extract local texture information in gray image. In order to improve the anti-noise ability of LBP, Jabid and Kabir [18] proposed a new feature extraction method based on Local Directional Pattern (LDP) operator. LDP operator not only inherits all the advantages of Local Binary Mode operator, but also is more robust to illumination and noise.

The basic idea of LDP is as follows. For a pixel x in the image, there is a local 3×3 neighborhood around it, and the pixel positions in the eight directions are $m_i(i = 0, 1, 2, \ldots, 7)$. The pixel is convoluted with eight Kirsch operators $M_0, M_1, M_2, \ldots M_7$ (as shown in Figure 1). Then the convolution values corresponding to 8 directions are obtained, and the absolute values of convolution values are sorted from large to small. And $p(p < 8)$ maximum absolute values is taken as its main feature with assigning the corresponding p pixel positions as 1 and assigning the remaining $8 - p$ position as 0. According to the pixel position m_0 to m_7, LDP binary code is obtained, and then converted to decimal code, that is, LDP value. The example of LDP coding process is shown in Figure 2.

Different from LBP using gray value comparison, LDP uses convolution to calculate gradient intensity. Because gradient intensity is more stable than gray value, LDP has stronger noise resistance and is not sensitive to light changes.

B. FACIAL EXPRESSION FEATURE EXTRACTION BASED ON CBP

Facial expression information is mainly concentrated in eyebrows, eyes and mouth, so it is very important to extract the texture features of these regions accurately. In order to further improve the discriminative power of the existing LBP operators, Fu and Wei [19] proposed the Centralized Binary Pattern (CBP) operator. CBP operator considers the role of the central pixel and gives it the highest weight to improve the discrimination.

The CBP code value is defined as

$$CBP(M, R) = \sum_{m=0}^{\frac{M}{2}-1} s(g_m - g_{m+(\frac{M}{2})})2^m + s(g_c - \frac{1}{M+1}(\sum_{m=0}^{M-1} g_m + g_c))2^{\frac{M}{2}} \quad (1)$$

where the symbol function $s(x)$ is defined as

$$s(x) = \begin{cases} 1 & when\ |x| \geq D \\ 0 & when\ |x| < D \end{cases} \quad (2)$$

In (1), M and R refer to neighborhood number and radius respectively. g_c is the center pixel, $g_m(m = 1, 2, 3, \ldots, 7)$ is the neighborhood pixel value. In (2), D is the threshold constant.

Different from the traditional LBP, CBP can capture the gradient information better by comparing the difference between the nearest neighbor points and the center pixels, so it can realize the capture of subtle texture by comparing the difference between the "neighbor point pairs". Moreover, CBP is not sensitive to white noise due to the change of the symbol function of the existing LBP Operator.

C. FUSION FEATURES

In common facial expression recognition, the recognition rate is very high when the characters in the test set appear in the training set, but the recognition rate will drop sharply when the training set does not contain the characters in the test set. This is because the essential facial features are not effectively obtained, which include facial appearance features.

After observation, facial expression features can be divided into persistent expression features and transient expression features. Persistent expression features mean that when people express a certain expression, the mouth, eyes and eyebrows change significantly, and have a greater impact on the expression. Transient expression features, including the changes of cheek and facial wrinkles, can also reveal the expression state to a certain extent. Therefore, in order to enhance the discrimination of expression features, we should integrate these two parts of features.

At present, expression texture feature extraction algorithms mainly improve the anti-noise ability of features and enhance the robustness of features, so as to improve the expression recognition rate. However, it is not considered that single selection of whole face features or local areas will lose expression feature information. Although the whole face expression feature extraction includes transient expression feature, it also includes face appearance feature. While the mouth, eyes and eyebrows contain the main persistent expression features with ignoring the change of the whole face lines.

Therefore, the persistent expression features are mainly concentrated in the mouth, eyes and eyebrows, and texture detail variation features need to be extracted. The transient expression features are expressed in the whole face area, and the changes of facial lines need to be captured. The key point of extracting transient expression features is to extract features with higher robustness under the influence of noise and illumination changes. Therefore, LDP is more suitable for extracting transient expression features. The key point of extracting persistent expression features is to extract subtle texture changes. Therefore, CBP is more suitable for extracting subtle expression changes.

This paper combines the whole face expression features extracted by LDP and the local expression features of eyebrow, eye and mouth extracted by CBP. Thus, the robustness of facial expression recognition is improved.

In this paper, the LDP value of the whole facial expression image is obtained by Section II (A). The literature [20] has proved that the best expression recognition rate can be obtained when $p = 3$, so the value of p in this paper is 3. According to the face region obtained by Viola Jones algorithm, the original LDP coding image is clipped and the LDP coded image of the face area is obtained. Then the characteristic histogram of the LDP coded image is calculated. The dimension of the LDP characteristic histogram is 56 dimensions The LDP expression feature extraction process is shown in Figure 3.

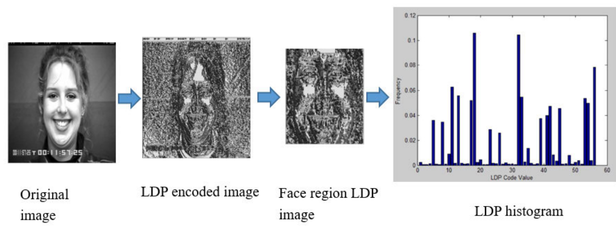

FIGURE 3. LDP expression feature extraction process.

In this paper, CBP coding map of facial expression image is obtained according to (1) and (2).Then the eyes (including eyebrows) and mouth regions are obtained by Viola Jones algorithm, and the original CBP coding image is clipped to obtain the corresponding CBP images of eyes (including eyebrows) and mouth regions. Finally, CBP characteristic histogram of two regions is calculated, and the histograms of two regions are connected in sequence to form CBP local expression features with a total of 64 dimensions. The CBP expression feature extraction process is shown in Figure 4.

Finally, the whole face expression LDP features F_{LDP} with 56 dimensions and local expression CBP features F_{CBP} with 64 dimensions were obtained. Then the two parts of features

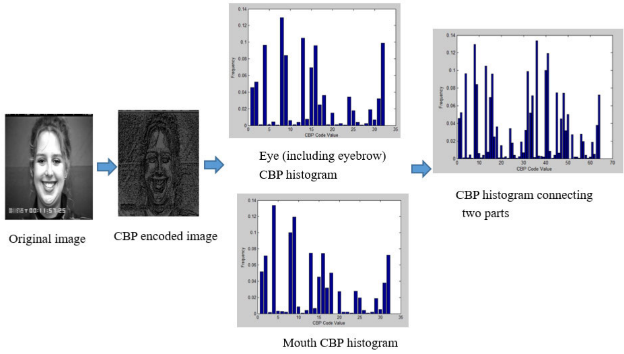

FIGURE 4. CBP expression feature extraction process.

are connected together to obtain a 120 dimensional facial expression feature F_V, which is defined as

$$F_V = \{F_{LDP}, F_{CBP}\}. \tag{3}$$

III. EXPRESSION CLASSIFICATION

The traditional expression recognition algorithms based on nearest neighbor classification use the nearest neighbor classifier to calculate the relative distance between the test expression image sample and the training image sample in the feature space. Since the distance represents the similarity between expression images, the category of the expression image with the minimum distance is selected as the classification result.

If only one best matching facial expression image is regarded as the final matching result, the chance of single facial expression image may lead to misjudgment. In order to increase the reliability and stability of classification, this paper proposes an improved nearest neighbor algorithm-k nearest neighbor algorithm to pre-classify facial expression images. Then, High Order Singular Value Decomposition (HOSVD) is used for accurate classification and recognition.

A. EXPRESSION PRE-CLASSIFICATION BASED ON K-NEAREST NEIGHBOR

In this paper, the k-nearest neighbor algorithm uses the feature distance as the best matching measure to find the k facial expression images with the smallest distance, so as to find the k best matching expression images. The feature distance is defined as

$$D(f, q) = \sum_{i=1}^{n} (f_i - q_i)^2. \tag{4}$$

where f_i is the i-th feature of facial expression image f, q_i is the i-th feature of the facial expression image q, and n is the feature dimension.

In this paper, the feature is obtained according to (3). And the test facial expression image is matched with all facial expression images in the training set according to (4), and then the facial expression images in the training set are arranged from small to large according to the matching distance. Finally, k-nearest neighbor images with the smallest distance are selected as the result of pre-classification. The pseudocode of k-nearest neighbor algorithm is shown as Table 1.

TABLE 1. Pseudocode of K-nearest neighbor algorithm.

k -nearest neighbor algorithm
Input: Training set T_r, Test data T_e, k
Output: k nearest image
1. for each feature $i \in T_r$
2. Calculate all the distance dis from the test image according to (4).
3. end for
4. for each distance $j \in dis$
5. Sort from small to large.
6. end for
7. The minimum k images of training set are selected as pre-classification set.

B. EXPRESSION CLASSIFICATION BASED ON HOSVD

1) TENSOR AND HIGHER ORDER SINGULAR VALUE DECOMPOSITION

In multilinear algebra, tensors can be represented by multidimensional arrays [21]. The order tensor can be expressed as $D \in R^{I_1 \times I_2 \times I_3 \ldots \times I_N}$. When $n = 1$, it is a vector; when $n = 2$, it is a matrix; when $n \geq 3$, it is a tensor. n-order tensor has n data dimensions, which are also called n modes.

Mode-n matrix means that the tensor is expanded into a matrix according to the n-th mode. The tensor $D \in R^{I_1 \times I_2 \times I_3 \ldots \times I_N}$ performed mode-n matrix operation, then the tensor elements are $(i_1, i_2, i_3, \ldots, i_n)$ mapped to matrix elements (i_n, j). Among them

$$j = 1 + \sum_{\substack{k=1 \\ k \neq n}}^{N} (i_k - 1) J_k, \quad J_k = \prod_{\substack{m=1 \\ m \neq n}}^{k-1} I_m \tag{5}$$

For example, the tensor $U \in R^{4\times3\times2}$ can be regarded as a cube as shown in Figure 5.

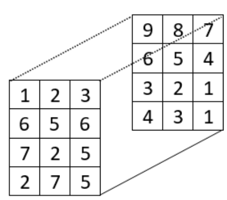

FIGURE 5. Three dimensional tensor.

$U \in R^{4\times3\times2}$ mode-1 matrix is:

$$U^1 = \begin{bmatrix} 1 & 2 & 3 & 9 & 8 & 7 \\ 6 & 5 & 6 & 8 & 5 & 4 \\ 7 & 2 & 5 & 3 & 2 & 1 \\ 2 & 7 & 5 & 4 & 3 & 1 \end{bmatrix}. \tag{6}$$

$U \in R^{4\times3\times2}$ mode-2 matrix is:

$$U^2 = \begin{bmatrix} 1 & 6 & 7 & 2 & 9 & 6 & 3 & 4 \\ 2 & 5 & 2 & 7 & 8 & 5 & 2 & 3 \\ 3 & 6 & 5 & 5 & 7 & 4 & 1 & 1 \end{bmatrix}. \tag{7}$$

$U \in R^{4\times3\times2}$ mode-3 matrix is:

$$U^3 = \begin{bmatrix} 1 & 6 & 7 & 2 & 2 & 5 & 2 & 7 & 3 & 6 & 5 & 5 \\ 9 & 6 & 3 & 4 & 8 & 5 & 2 & 3 & 7 & 4 & 1 & 1 \end{bmatrix}. \tag{8}$$

Mode-n product [22] means that the tensor is multiplied by the matrix in the n-th mode, and the n-th order dimension is required to be equal to the number of rows of the matrix, and the product result is still a tensor. The tensor $D \in R^{I_1\times I_2\times I_3 \ldots \times I_N}$ mode-n product with $U \in R^{J_n\times I_n}$ is:

$$(D \times_n U)_{i_1,i_2,\ldots,i_{n-1},j_n,i_{n+1},\ldots i_N} = \sum_{i_n=1}^{I_n} d_{i_1,i_2,i_3,\ldots i_{n-1},i_n,i_{n+1\ldots},i_N} \bullet u_{j_n,i_n}. \tag{9}$$

For example, the tensor $U \in R^{4\times3\times2}$ mode-1 product with

$$V = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \end{bmatrix}.$$

The result is also a tensor.

The result $M = U \times_1 V \in R^{2\times3\times2}$ is as shown in Figure 6.

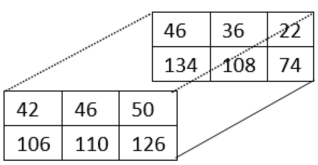

FIGURE 6. Mode-1 product result.

HOSVD (Higher Order Singular Value Decomposition) [23] is a higher order generalization of matrix singular value decomposition. It can decompose the tensor into the mode-n product of a kernel tensor and several orthogonal matrix. For example, the singular value decomposition of the tensor $D \in R^{I_1\times I_2\times I_3 \ldots I_n \ldots \times I_N}$ is as follows

$$D = S \times_1 U_1 \times_2 \ldots \times_i U_i \ldots \times_N U_N \tag{10}$$

where $S \in R^{I_1\times I_2\times I_3\times \ldots \times I_N}$ is a core tensor, and U_i is an $I_i \times I_i$ orthogonal subspace matrix.

2) EXPRESSION CLASSIFICATION BASED ON HOSVD

This paper proposes a secondary expression classification algorithm based on HOSVD. The algorithm matches the similarity between the k face expression images and the test image again by HOSVD, and finds the most similar expression image with the test image, so as to determine the expression category of the test image.

This is the algorithm framework which is shown as Figure 7. Firstly, the test expression image is matched with each facial expression image in the pre-classification set. And the eyes, eyebrows and mouth regions of these expression image are extracted by Viola Jones algorithm. Then, the region cut out from the expression image is reassembled into a new local expression image, and the LDP feature extraction is performed to obtain the local LDP encoded image. Then, the local LDP encoded image is divided into blocks, and the corresponding image blocks constitute tensor, and the tensor is decomposed by HOSVD to get the decomposition coefficient. Then the region energy of the two decomposition coefficients is calculated, and the similarity of the two images is determined according to the region energy. Finally, according to the similarity, the probability of k expression images in the pre-classification set is calculated, and the category of the expression image with the largest probability is the category of the test expression image.

The specific steps of the proposed algorithm are as follows.

a. Let the test expression image be F, and the expression image in the pre-classification set is $G_i (i = 0, 1, \ldots, k)$. According to Viola Jones algorithm, the eye, eyebrow and mouth regions of these expression images are segmented, and these regions are spliced into new local expression images F' and G'_i. The LDP features of F' and G'_i are extracted, and the LDP encoded images are F'' and G''_i respectively.
for $i = 1, 2, \ldots, k$
b. *for* $p = 1, 2, \ldots, R$
for $q = 1, 2, \ldots, C$
   - F'' and G''_i are divided into $R \times C$ corresponding $M \times N$ image sub blocks, and R represents the number of image sub blocks in horizontal direction, C represents the number of image sub blocks in vertical direction. Then $R \times C$ corresponding $M \times N \times 2$ image sub blocks are formed into $R \times C$ sub tensors which are $M \times N \times 2$, denoted as $B_j, j = 1, 2, \ldots, R \times C$.
   - the sub tensor B_j is decomposed by HOSVD according to (11), and the sub block coefficient is obtained according to (12), which represents the characteristics of the image.

$$B_j = S_j \times_1 U_j \times_2 V_j \times_3 W_j \tag{11}$$

$$\bar{S}_j = S_j \times_3 W_j \tag{12}$$

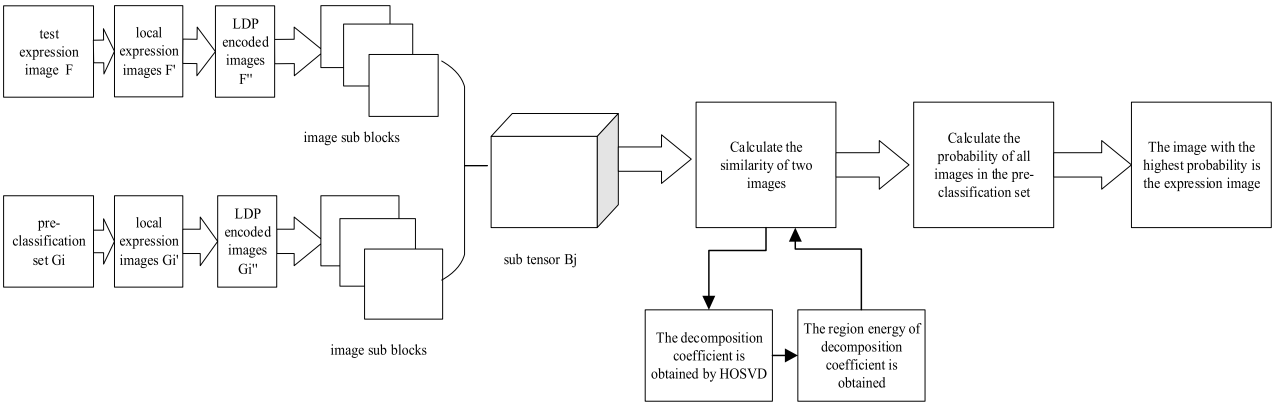

FIGURE 7. Flow chart of expression secondary classification based on HOSVD.

- According to the sub block coefficient $\bar{S}_j$ obtain the regional energy E_i :

$$E_j(l) = \sum_{m=1}^{M}\sum_{n=1}^{N} |\bar{S}_j(m, n, l)|, \quad l = 1, 2 \quad (13)$$

If the region energy values of the two coefficients are closer, the more redundant information between the two image blocks is, the more similar the images are.

- According to the region energy, the similarity value R of the two images is obtained:

$$R = |\frac{E_i(1)}{E_i(2)} - 1| \quad (14)$$

c. According to (14), the similarity value of k facial expression images and test images in the pre- classification set is calculated: $R_i(i = 1, 2, \ldots, k)$. Suppose the type of expression is $c(c = 1, 2, 3, \ldots, 6)$, calculate the probability P_c of belonging to each expression:

$$p_c = \sum_{i=1}^{k} f(R_i) \quad (15)$$

$$f(R_i) = \begin{cases} R_i & if\ c = i \\ 0 & if\ c \neq i \end{cases} \quad (16)$$

d. The maximum probability of expression category belongs to the category of facial expression test image. This algorithm is the second screening of k expression image ranges in the pre-classification set. It mainly transforms the expression image into high-order tensor space for the second accurate screening, and finally obtains the corresponding categories by using probability statistics.

Obviously, the recognition rate of this method is affected by the number k of nearest neighbors in k-NN algorithm. According to the analysis of experimental results, when the number k of nearest neighbors is 4, we can achieve higher recognition rate and better system robustness.

3) ALGORITHM PERFORMANCE ANALYSIS

For an expression image to be classified, the number of pre-classification set is k and the number of blocks in region calculation is $R \times C$. Compared with the traditional k-nearest neighbor classification algorithm, the proposed classification algorithm based on HOSVD needs more steps of secondary classification. The secondary classification algorithm based on HOSVD mainly includes the calculation of block energy. The energy time complexity of image block area is $O(kRC)$. Due to the small values of k, R and C, the time complexity of the proposed classification algorithm is slightly higher than that of the traditional k-nearest neighbor classification algorithm, but the recognition rate has been greatly improved. This is also the direction of further improvement of the algorithm.

IV. EXPERIMENT ANALYSIS

In this section, two publicly available databases, the extended Cohn Kanade dataset (CK+) and the Oulu-CASIA NIR&VIS facial expression database are used for person-independent facial expression recognition.

Oulu-CASIA NIR&VIS facial expression database is composed of 80 facial expression videos. Each person's expression video contains six typical expressions (happiness, sadness, surprise, anger, fear, disgust). All the videos are collected from calm to the corresponding peak state of expression. These objects are photographed by two imaging systems (near infrared) and VIS (visible light) under three different lighting conditions: normal indoor lighting, weak light and dark lighting. The research topic of this paper is facial expression recognition under visible light, so the image used is the expression picture under visible light normal illumination, and take a peak expression picture of each expression video of each person. In Oulu-CASIA NIR&VIS facial expression database, a total of 480 facial expression images were selected. Among them, there are 80 people, and each of them has 6 facial expression images, representing 6 typical expressions respectively. So there are 80 images of each expression kind. And six facial expression images of the database under visible light are shown in Figure 8.

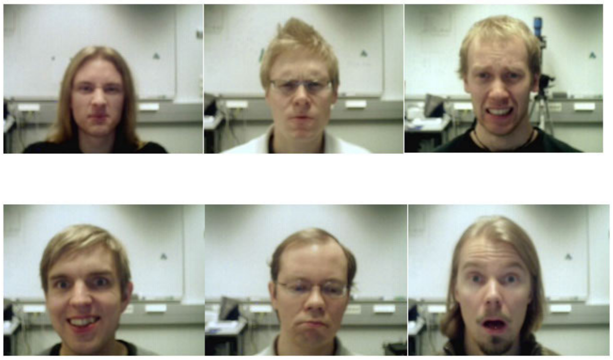

FIGURE 8. Six expressions of Oulu CASIA NIR NIR&VIS facial expression database.

In Oulu-CASIA NIR&VIS facial expression database, the experiment randomly selects all the expression images of 18 people as the test set, and the expression images of the remaining 62 people as the training set images, so as to ensure that the training set does not contain the data in the test set. The experiment is conducted five times, and the recognition rate is taken as the average value of five times. In the secondary classification of HOSVD, the eyes and eyebrows regions were scaled to 44 $\times$ 100, and the mouth area were scaled to 44 $\times$ 64. After horizontally stitching these two regions into 44 $\times$ 164 expression images, the stitched image is divided into 4 (44/11) $\times$ 2 (164/41) image sub-blocks.

The extended Cohn Kanade (CK +) dataset is based on Cohn Kanade dataset and was released in 2010. The dataset includes 123 subjects and 593 image sequences. The method proposed by Patrick *et al.* [24] is used to extract the corresponding expression images, and a total of 279 expression images are selected. Among them, there are 45 angry expression images, 59 disgusting expression images, 25 fear expression images, 69 happy expression images, 28 sad expression images and 83 surprise expression images. Six facial expression images of the database are shown in Figure. 9.

FIGURE 9. Six expressions of CK+ dataset.

In the experiment of CK+ dataset, 6 facial expressions of each kind were randomly selected for testing, and all the remaining expressions were trained. The experiments ensured that the training set does not contain the data in the test set. Because less test data is selected each time than Oulu-CASIA NIR&VIS facial expression database, the experiment was carried out ten times, and the average value of the ten times was taken as the experimental results. Since the image resolution in the CK+ dataset is large, which affects the recognition speed, all images are normalized to 320 $\times$ 145 in the image pre-processing stage. In the secondary classification of HOSVD, the eyes and eyebrows regions were scaled to 46 $\times$ 100, and the mouth area were scaled to 46 $\times$ 46. After horizontally stitching these two regions into 46 $\times$ 146 expression images, the stitched image is divided into 2 (46/23) $\times$2 (146/73) image sub-blocks.

A. COMPARISON OF CLASSIFICATION ALGORITHMS

In order to compare the nearest neighbor classification algorithm with the classification algorithm in this paper, CBP and LDP feature extraction method are both used in the feature extraction stage, and the traditional nearest neighbor algorithm and the classification algorithm in this paper are compared in the classification and recognition stage. The experimental results on Oulu CASIA NIR&VIS facial expression database are shown in Figure10.The experimental results on the CK + dataset are shown in Figure 11.

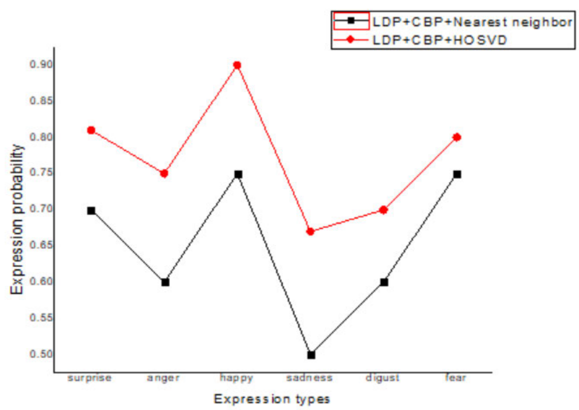

FIGURE 10. Comparison of recognition rate between nearest neighbor algorithm and HOSVD secondary classification algorithm on Oulu CASIA NIR&VIS facial expression database.

It can be seen from Figure 10 and Figure 11 that, the k-NN pre-classification and the HOSVD secondary classification recognition algorithm proposed in this paper are compared with the traditional nearest neighbor algorithm in the Oulu CASIA NIR &VIS facial expression database and the CK+ dataset. The recognition rate of each kind of expression has been improved. The specific data results are shown in table 2 and table 3. In Oulu CASIA NIR &VIS expression database, the average recognition rate increased from about 64.5% to about 77.5%, of which the most improved recognition rate

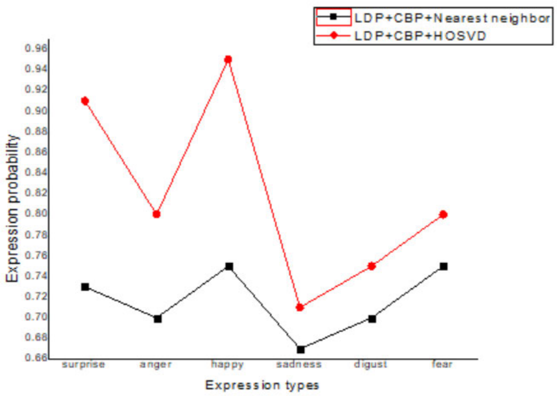

FIGURE 11. Comparison of recognition rate between nearest neighbor algorithm and HOSVD secondary classification algorithm on CK+ dataset.

TABLE 2. Comparison of recognition rate between nearest neighbor algorithm and HOSVD secondary classification algorithm on Oulu CASIA NIR&VIS facial expression database.

Expression types	Number of test samples / Recognition number	
	LDP+CBP+ Nearest neighbor	LDP+CBP+HOSVD
surprise	90/62	90/73
anger	90/55	90/68
happy	90/67	90/80
sadness	90/46	90/62
disgust	90/52	90/64
fear	90/67	90/71
Average recognition rate	64.5%	77.5%

was happy and sad, happy expression increased by 15%, sad expression increased by 18%. And in the CK+ dataset, the average recognition rate increased from 71.6% to about 82%, of which the most improved recognition rate was surprise and happy, the recognition rate of happy expression increased by 27%, and the recognition rate of happy expression increased by 20%.

This shows that using HOSVD as a decomposition tool can effectively extract the structural features of LDP encoded images and find the similarity of the two images more accurately. The effectiveness of the proposed algorithm based on HOSVD is proved. In the Oulu CASIA NIR &VIS facial expression database and CK + dataset, the recognition rate of surprise, happiness and fear is relatively high, because the expression degree of these three expressions is relatively exaggerated and easier to identify.

In addition, the recognition rate of the two algorithms in CK + dataset is higher than that of Oulu CASIA NIR

TABLE 3. Comparison of recognition rate between nearest neighbor algorithm and HOSVD secondary classification algorithm on CK+ dataset.

Expression types	Number of test samples / Recognition number	
	LDP+CBP+Nearest neighbor	LDP+CBP+HOSVD
surprise	60/44	60/54
anger	60/43	60/49
happy	60/45	60/57
sadness	60/40	60/43
disgust	60/42	60/45
fear	60/45	60/48
Average recognition rate	71.6%	82%

&VIS facial expression database. This is because in Oulu CASIA NIR &VIS facial expression database, different people sometimes express the same expression differently. However, in the CK + dataset, everyone has basically maintained the same expression. Taking disgust expression as an example. The expression images in Oulu CASIA NIR &VIS facial expression database are shown in Figure 12, and those in CK + dataset are shown in Figure 13.

FIGURE 12. Disgust expression in Oulu CASIA NIR &VIS expression database.

FIGURE 13. Disgust expression in CK+ expression dataset.

B. EXPERIMENTAL ANAYSIS OF PARAMETER k

Because the recognition rate of this method is affected by the number of neighbors k in k-NN algorithm, we compare the effect of different number of neighbors $k$on the recognition

TABLE 4. Experimental results of different neighbor numbers k on Oulu CASIA NIR&VIS facial expression database.

Neighbor numbers	k=2	k=3	k=4	k=5	k=6
Recognition rate	55.6%	62.9%	75%	62.9%	54.13%

TABLE 5. Experimental results of different neighbor numbers k on CK+ dataset.

Neighbor numbers	k=2	k=3	k=4	k=5	k=6
Recognition rate	70.75%	75.88%	81%	75.88%	70.75%

rate in the Oulu CASIA NIR &VIS facial expression database and CK + dataset respectively. The experimental results of Oulu CASIA NIR&VIS facial expression database are shown in Table 4, and those of CK + dataset are shown in Table 5.

It can be seen from table 4 and table 5 that there is the same rule in the Oulu CASIA NIR&VIS facial expression database and the CK + dataset. The rule is when the value of the number of neighbors k is gradually taken from 2 to 4, the recognition rate increases gradually. This is because that the pre-classification set increase gradually, and the probability of containing correct facial expression images is gradually increasing. When k is 4, higher expression recognition rate can be obtained, but when k is greater than 4, the expression recognition rate gradually decreases. This is because the wrong facial expression image is added when k is greater than 4, and the correct recognition rate of facial expression reduces gradually.

C. COMPARISION OF RECOGNITION OF RATES OF DIFFERENT EXPRESSION FEATURE EXTRACTION METHODS

In order to further illustrate the impact of feature selection on expression recognition, this article will select three methods in the feature extraction stage. The first method is extracting the local features of eyes, eyebrows and mouth only by LDP. The second method is extracting the local features of eyes, eyebrows and mouth only by CBP. The third method is extracting the local features of eyes, eyebrows and mouth by CBP and whole facial expression features by LDP. The comparison results of the three methods in Oulu CASIA NIR &VIS facial expression database are shown in Table 6. The comparison results in CK + expression dataset are shown in Table 7.

It can be seen from table 6 and table 7 that CBP is better than LDP to extract the local features of eyes, eyebrows and mouth in Oulu CASIA NIR &VIS facial expression database and CK + expression dataset. This suggests that CBP can extract local features better and has higher discrimination ability. When CBP is used to extract the local features of the eyes, eyebrows and mouth and LDP is used to extract the whole facial expression feature, the recognition rate is higher than only extracting the local features by LDP or CBP. It suggests that both whole facial expression features and local expression features will affect the recognition rate, and the fusion of whole facial expression features and local expression features can achieve higher recognition rate.

TABLE 6. Comparison of expression recognition rate of feature extraction algorithms on Oulu CASIA NIR&VIS facial expression database.

Algorithm	Facial Expression database	Recognition rate
LDP + HOSVD secondary classification algorithm	Oulu CASIA NIR&VIS facial expression database	58%
CBP + HOSVD secondary classification algorithm	Oulu CASIA NIR&VIS facial expression database	62.8%
CBP+LDP + HOSVD secondary classification algorithm	Oulu CASIA NIR&VIS facial expression database	75%

TABLE 7. Comparison of expression recognition rate of feature extraction algorithms ON CK+ dataset.

Algorithm	Facial Expression database	Recognition rate
LDP + HOSVD secondary classification algorithm	CK+ dataset	63.33%
CBP + HOSVD secondary classification algorithm	CK+ dataset	76.67%
CBP+LDP + HOSVD secondary classification algorithm	CK+ databset	81%

TABLE 8. Compare experiments with state-of-the-art methods.

Algorithm	Facial Expression database	Recognition rate
AAM+CNN+LBP+ Random forest [5]	CK+ dataset	75%
LBP+ CNN classification [6]	CK+ dataset	78%
PCA+Random forest[7]	CK+ dataset	70%
The algorithm in this paper	CK+ dataset	82%

D. COMPARED WITH OTHER ALGORITHMS

Due to the universality of CK + expression dataset, this paper compare the performance of this method with state-of-the-art methods in the person-independent facial expression recognition on the CK+ dataset, as shown in Table 8.

As can be seen from table 8, the recognition rate of person-independent facial expression is generally not high, but the average recognition rate of this method is better than other expression recognition methods. This is mainly due to the fusion of persistent and transient features, which not only uses the persistent expression features, but also highlights the transient expression features. In addition, HOSVD is used for secondary classification and recognition. The algorithm makes full use of the results of k-NN algorithm, and combines with HOSVD for accurate similarity matching to improve the recognition performance, thus obtaining higher expression recognition rate and better system robustness.

V. CONCLUSION

This paper analyzes and studies the person-independent facial expression recognition. An expression recognition algorithm based on improved LBP and HOSVD is proposed. The algorithm has the following characteristics:

1) By comparing the expression features extracted by CBP and LDP, the persistent expression feature by CBP and the transient facial expression feature by LDP are fused. This not only makes use of the persistent face expression features, but also highlights the transient expression features.
2) The k-NN is used for pre-classification and HOSVD is used for secondary classification. In this way, HOSVD decomposition tool is used to make up for the deficiency of k-NN recognition and improve the recognition performance.
3) Through the experiment of Oulu CASIA NIR&VIS facial expression database and CK+ dataset, the experimental results prove that the algorithm has better overall generalization performance and higher average recognition rate.

REFERENCES

[1] Y. Wang, "Study on fatigue driving recognition method based on facial expression," M.S. thesis, Dept. Sci. & Tech., Xi'an Univ, Xi'an, China, 2019.

[2] T. Y. Ma, "Research on driver fatigue state recognition method based on facial expression features," M.S. thesis, Dept. Auto. Eng. Tsinghua Univ, Beijing, China, 2012.

[3] M. T. Li and S. H. Chen, "Research progress on automatic recognition system of neonatal pain expression," *JCM*, vol. 54, no. 11, pp. 1644–1647, Nov. 2019.

[4] H. Li, Y. Li, and Z. J. Zhou, "Analysis of teaching effect based on facial expression in classroom environment," *Modern Distance Edu. Res.*, vol. 4, pp. 97–103, Jul. 2017.

[5] Y. W. Li, W. Y. Zheng, and N. Lin, "Facial expression recognition based on AAM, CNN and LBP features," *Comput. Eng. Des.*, vol. 39, no. 7, pp. 3436–3440, Sep. 2017.

[6] D. P. Jiang, B. Yang, and L. Zhou, "Facial expression recognition based on LBP convolutional neural network," *Comput. Eng. Des.*, vol. 38, no. 12, pp. 1971–1977, Dec. 2018.

[7] Z. Y. Ou, "Expression recognition method based on robust PCA feature and random forest," *Comput. Eng. Des.*, vol. 39, no. 2, pp. 580–584+595, Feb. 2018.

[8] H. Wang and Ahuja, "Facial expression decomposition," in *Proc. 9th IEEE Int. Conf. Comput. Vis.*, Oct. 2003, pp. 958–965.

[9] B. Sun and Y. N. Liu, "Facial expression feature extraction based on tensor analysis," *Comput. Eng. App.*, vol. 52, no. 20, pp. 145–148+226, Jun. 2015.

[10] X. G. Jiang, B. Feng, and Z. F. Li, "Weighted sum sparse classification method for nonspecific expression," *Comput. Syst. App.*, vol. 23, no. 7, pp. 190–194, Jul. 2014.

[11] B. F. Hu, "Facial expression recognition based on multi classifier ensemble of expression subspace," *Comput. App*, vol. 31, no. 3, pp. 736–740, Mar. 2011.

[12] H. Tan, Y. Zhang, H. Cheri, Y. Zhao, and W. Wang, "Person-independent expression recognition based on person-similarity weighted expression feature," *J. Syst. Eng. Electron.*, vol. 21, no. 1, pp. 118–126, Feb. 2010.

[13] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," presented at the IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., CVPR, Kauai, HI, USA, Dec. 2001.

[14] T. Cover and P. Hart, "Nearest neighbor pattern classification," *IEEE Trans. Inf. Theory*, vol. IT-13, no. 1, pp. 21–27, Jan. 1967.

[15] L. S. Yao, G. M. Xu, and B. Fang, "Video expression recognition method based on LBP and SVM," *J. Shandong Uni. Tech. (Nature Sci. Ed.)*, vol. 34, no. 4, pp. 67–72, Apr. 2020.

[16] G. S. Zhang, G. Y. Ge, and R. H. Zhu, "Facial expression recognition based on LBP features and deep learning," *Comput. Mea. And Control*, vol. 28, no. 2, pp. 174–178, Feb. 2020.

[17] S. M. Wang and Y. H. Liang, "Facial expression feature extraction method based on improved LBP," *J. Mea. Sci. Instru.*, vol. 10, no. 4, pp. 342–347, Dec. 2019.

[18] T. Jabid, M. H. Kabir, and O. Chae, "Local directional pattern (LDP) – a robust image descriptor for object recognition," in *Proc. 7th IEEE Int. Conf. Adv. Video Signal Based Surveill.*, Aug. 2010, pp. 482–487.

[19] X. F. Fu and W. Wei, "Facial expression recognition based on multi-scale centralized binary pattern," *Control Theory Appl.*, vol. 26, no. 6, pp. 629–633, Jun. 2009.

[20] Y. M. Zhang and W. Zhang, "Facial expression recognition based on data fusion," *J. Sichuan Uni. (Eng. Sci. Ed.)*, vol. 48, no. 6, pp. 160–164, Oct. 2016.

[21] C. Y. Zhang, "Research on tensor based image fusion method," M.S. thesis, Dept. Com. Sci., Tec., Jiangnan Univ., Wuxi, China, 2015.

[22] Y. He, "Facial expression recognition based on multi-feature fusion and HOSVD," presented at the 3rd Inf. Technol., Netw., Electron. Automat. Control Conf. (ITNEC), Chongqin, China, Jan. 2019.

[23] G. Bergqvist and E. Larsson, "The higher-order singular value decomposition: Theory and an application [lecture notes," *IEEE Signal Process. Mag.*, vol. 27, no. 3, pp. 151–154, May 2010.

[24] L. Patrick, J. F. Cohn, and T. Kanade, "The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression," presented at the IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Workshops, San Francisco, CA, USA, Jun. 2010.

YING HE was born in Suining, Sichuan, China, in 1984. She received the B.S. degree in computer science and technology and the master's degree in computer software and theory from Sichuan Normal University, in 2007 and 2010, respectively.

Since 2010, she has been teaching with the Computer Science and Technology Department, Ren'ai College of Tianjin University. She is the author of two books and six articles. Her research interests include artificial intelligence, graphics and images, pattern recognition, computer simulation, big data analysis, and so on.

SHUXIN CHEN was born in Harbin, Heilongjiang, China, in 1978. She received the B.S. degree in computer science and technology from Qiqihar University, in 2001, the M.S. degree in thermal engineering from the Harbin University of Technology, in 2010, and the Ph.D. degree in optical engineering from the National Key Laboratory of School of Science, Harbin Engineering University, in 2019.

From 2001 to 2015, she was with the Computer Center, Qiqihar University. She was a Professor, in 2018. She is the author of three books, more than 30 articles, and 12 inventions. Her research interests include computer simulation, big data analysis, astronomical spectrum processing, and so on.