Dynamic Facial Expression Recognition based on the Fusion of Handcraft Features

Yan Wang, *Member*, *IAENG*, Shaodong Miao, Xiaoying Gu, Yancong Zhou, and Bo Zhang*

***Abstract*—Facial expression recognition becomes an attractive research area in the field of biological recognition, which plays a vital role in a wide range of the application of the medical treatment, security systems, education, and so on. Extracting discriminative facial expression features has been an essential solution to improve the performance in dynamic facial expression recognition tasks. In this paper, we present a novel dynamic facial expression recognition framework that fusing geometric features with semantic features to effectively extract robust features. First, the proposed distance difference method is used to identify the peak frame in an image sequence, and the geometric features are calculated by the proposed geometric feature-based method. Then the geometric features are described semantically to construct a semantic feature set. Finally, we conduct an assessment on these geometric features according to the feature intensity standards and semantic rules to obtain the final fused features. Our method is carried out on two benchmark datasets: CK + (Extended Cohn-Kanade) and MMI (M & M Initiative). Its efficiency is demonstrated by extensive experiments. The best result achieves 98.73% recognition rate as classifying the 6 classes on CK+ and 87.36% on MMI, outperforming the results of state-of-the-arts.**

***Index Terms*—dynamic expression recognition, geometric feature, semantic feature, SVM**

I. Introduction

Facial expressions can intuitively reflect human internal activities and mental states, and convey a variety of emotions in communication. Mehrabian [1] pointed out that in the process of human communication, the expression conveys 55% of the amount of information. Now, facial expression is widely used in the fields of car driving, medical treatment, distance education, and human-computer interaction [2-4].

As the expression is a process of continuous transformation, it is difficult for the static image-based method to represent the changes of an expression. The dynamic sequence-based approach contains both spatial features and temporal features, which can perform expression recognition better. There are texture feature-based approaches including LBP-TOP (Local Binary Pattern with Three Orthogonal Planes)) [5], STM-LBP (Space-Time Motion Local Binary Pattern) [6], STWLD (Space-Time Weber Local Descriptor) [7], and GSP-HOG (Graphic Signal Processing Histogram of Oriented Gradient) [8]. These texture feature-based methods have good rotation invariance and noise resistance. However, there are texture deviations as the frame resolution changing, and the methods also susceptible to the influence of light. Geometric feature-based methods utilize the position or displacement relationship of feature points, such as ASM (Active Shape Model) [9], AAM (Active Appearance Model) [10, 11, 12], EBGM (Elastic Beam Graph Matching) [13], *et al*. These methods have the advantage of good translation rotation invariance, illumination robustness, simple calculation, and fast recognition speed. The deep learning methods can find potential relationships among features. Many deep learning networks have been adopted for recognizing expressions, such as Alexnet [14], VGG (Visual Geometry Group Network) [15], Googlenet [16], and Resnet [17]. Although deep learning achieved impressive results in feature extraction tasks, the training and testing stages of deep learning networks require a large amount of data, and most of the datasets are too small to meet the need. Semantic feature-based methods can extract high-level semantic features through analyzing the low-level visual features. The methods include Bag-of-Words models [18] and axiomatic fuzzy sets [19, 20]. Duan *et al*. [21] proposed a semantic feature-based method that built semantic concepts to describe expression features and establish a semantic rule to express the specific features. The experimental results show the method has better semantic interpretability for expression features and ideal accuracy for facial expression recognition.

Inspired by [21], we present a method fusing handcraft features to effectively extract discriminative information of dynamic expressions. We first perform preprocessing such as rotation correction and normalization on the frame to avoid the interference of angle, scale, and illumination. Then, we present the detection method to identify the peak frame of a sequence. After marking the feature points of the peak frame and neutral frame, we can calculate the geometric features of an expression by the proposed geometric-based method. Finally, we construct a semantic feature set according to the geometric characteristics and analyze the geometric features through the semantic rules to obtain the fused features.

The paper is organized as follows: Section II introduces the preprocessing of facial expression sequence. Section III illustrates the proposed fused handcraft method. Section IV discusses the experimental results. The paper is concluded in section V.

Manuscript received October 30, 2020; revised June 1, 2021.

Yan Wang is a teacher of Tianjin University of Commerce, Tianjin 300134, China (e-mail: ywrenrenren@163.com).

Shaodong Miao is a postgraduate of School of Information, Hebei University of Technology, Tianjin 300401, China (e-mail: 106748693@qq.com).

Xiaoying Gu is a teacher of Tianjin University of Commerce, Tianjin 300134, China (e-mail: guxiaoying2006@163.com).

Yancong Zhou is a Professor of Tianjin University of Commerce, Tianjin 300134, China (e-mail: zycong78@126.com).

Bo Zhang is a Professor of Tianjin University of Commerce, Tianjin 300134, China (phone: 022-26667577; corresponding author, e-mail: fantasy_zb@163.com).

II. Image Sequence Preprocessing

In this section, we first perform face detection and feature point location on every sequence. Then, we detect the peak frame by calculating the distance between each frame and neutral frame in a sequence according to the proposed method. In which, the neutral frame can be got by the artificial way, and the frame taking the maximum distance can be regarded as the peak frame.

It is necessary to locate the facial feature points of a sequence due to the detecting of a face play a vital role in the process of facial expression recognition. We adopt the function facedetect_multiview_reinforce () of the LibfaceDetection face detection method [22] to locate 60 feature points. The coordinates of the 60 feature points can be expressed as Equation (1):

$$\mathbf{X} = ((x_1, y_1), (x_2, y_2), ..., (x_i, y_i), ..., (x_{60}, y_{60}))^{\mathrm{T}} \quad (1)$$

where x_i, y_i are the abscissa and ordinate of the *ith* point, $i \in [1,60]$. The detection result is shown in Figure 1. It can be seen that the method can mark the faces and locate the feature points exactly, including the frame that has a large deflection angle.

We select a total of 11 typical feature points of the eyebrows, eyes, mouth, and nose, that is, the 21*st*, 33*rd*, 36*th*, 39*th*, 43*rd*, 47*th*, 48*th*, 51*st*, 54*th*, 62*nd*, and the 66*th* points to represent an expression. The six lines formed by these points are shown in Figure 2.

The distances of D_1 to D_6 can be calculated as:

$$\begin{aligned} D_1 &= \sqrt{(x_{21}-x_{39})^2+(y_{21}-y_{39})^2} \\ D_2 &= \sqrt{(x_{43}-x_{47})^2+(y_{43}-y_{47})^2} \\ D_3 &= \sqrt{(x_{36}-x_{48})^2+(y_{36}-y_{48})^2} \\ D_4 &= \sqrt{(x_{33}-x_{51})^2+(y_{33}-y_{51})^2} \\ D_5 &= \sqrt{(x_{62}-x_{66})^2+(y_{62}-y_{66})^2} \\ D_6 &= \sqrt{(x_{48}-x_{54})^2+(y_{48}-y_{54})^2} \end{aligned} \quad (2)$$

where x_{ij} denotes the coordinates of the point ij.

The six distances of the *ith* frame are accumulated to get the S_i. Then we subtract S_i from the neural frame S_n to get v_i, and store v_i in the array t, $t = (v_i)$, $i \in N$, N is the number of the frame in a sequence. Finally, we can find the maximum value v_j which represents the peak frame, as shown in Figure 3.

III. The Proposed Method

A. Geometric feature extraction

We extract geometric features on eyebrows, eyes, noses, and mouths where the feature change significantly. The processes of the method are as follows:

(1) Locating geometric regions

We select 30 feature points changing obviously on the regions of eyebrows, eyes, nose, and mouth to form the expression geometric feature set **TR**:

$$\mathbf{TR} = \{\mathbf{tr}_1, \mathbf{tr}_2, ..., \mathbf{tr}_i, ..., \mathbf{tr}_{10}\} \quad (3)$$

where the *ith* geometric feature $\mathbf{tr}_i$ can be written as :

$$\mathbf{tr}_i = X\_S_{i,j} \quad (4)$$

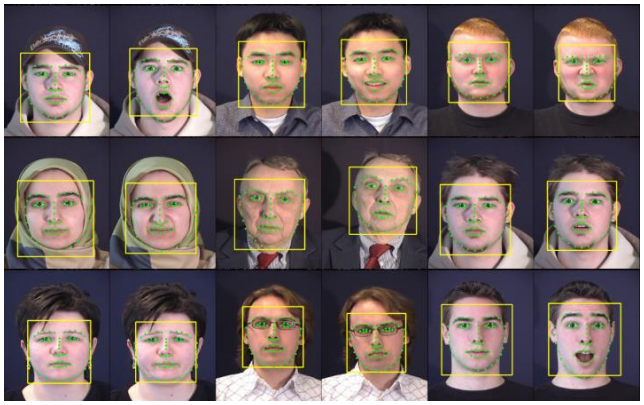

Fig. 1. The diagrams of face detection and feature points location.

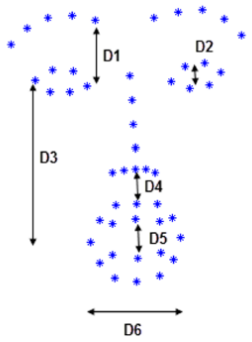

Fig. 2. The typical distances formed by 11 feature points

Fig. 3. Examples of the peak frame detection results.

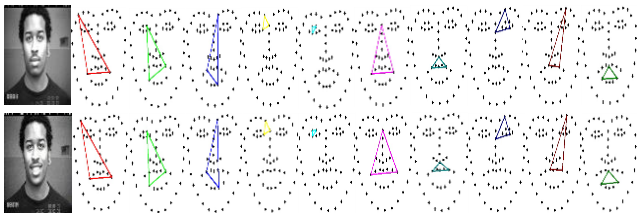

Fig. 4. The located geometric features of emotion "happiness"

where $i \in [0,9]$, $X\_S_{i,j}$ denotes the *jth* ($j = 1,2,3$) vertex of the *ith* feature $\mathbf{tr}_i$. The features of "happiness" are shown as Figure 4.

The geometric regions include left eyebrow and lips; left eyes and lips; right eyes and lips; eyes and eyebrows; only eyes; center point of nose and corner of mouth; nose, lips, and eyebrows; eyes and nose; right eyes and lips; only lips.

(2) Geometric feature definition

Geometric distance feature: The Euclidean distance $d_{i,j}$ between the two vertices of the *ith* geometric triangle is defined as the geometric distance feature:

$$d_{i,j} = \sqrt{(x_{i,a} - x_{i,b})^2 + (y_{i,a} - y_{i,b})^2} \quad (5)$$

where i is the number of geometric triangles, $i \in [0,9]$; j is the number of distance features; a and b are the geometry vertex, $a, b \in [1,3]$.

Geometric angle feature: The internal angle $r_{i,j}$ of a geometry is defined as the geometric angle feature:

$$r_{i,j} = \frac{(x_{i,a} - y_{i,c}) \times (y_{i,a} - y_{i,c}) + (x_{i,b} - x_{i,c}) \times (y_{i,b} - y_{i,c})}{\sqrt{((x_{i,a} - y_{i,c})^2 + (y_{i,a} - y_{i,c})^2) \times ((x_{i,b} - x_{i,c})^2 + (y_{i,b} - y_{i,c})^2)}} \quad (6)$$

where $(x_{i,a}, y_{i,a})$, $(x_{i,b}, y_{i,b})$, and$(x_{i,c}, y_{i,c})$ are the horizontal and vertical coordinates of vertex a, b and c, $i \in [0,9]$, $a, b, c \in [1,3]$; j is the angle number, $j \in [1,3]$.

(3) Geometric feature calculation

The processes of getting geometric features are as follows:

Firstly, we store the distance features and angle features of the neutral frame, that is, the value of $d_{i,1}$, $d_{i,2}$, $d_{i,3}$, $r_{i,1}$, $r_{i,2}$, $r_{i,3}$ are stored to the vector $\mathbf{h}_i$:

$$\mathbf{h}_i = \{d_{i,1}, d_{i,2}, d_{i,3}, r_{i,1}, r_{i,2}, r_{i,3}\} \quad (7)$$

Then, the distance features and angle features of the peak frame, that is, the value of $d_{i,1}$, $d_{i,2}$, $d_{i,3}$, $r_{i,1}$, $r_{i,2}$, $r_{i,3}$ are stored to the vector $\mathbf{w}_i$:

$$\mathbf{w}_i = \{d_{i,1}, d_{i,2}, d_{i,3}, r_{i,1}, r_{i,2}, r_{i,3}\} \quad (8)$$

Finally, the ratio of $\mathbf{w}_i$ and $\mathbf{h}_i$ is put into an array $\mathbf{z}$ which is the geometric feature:

$$\mathbf{z} = \{\mathbf{w}_{i,j} / \mathbf{h}_{i,j}\} \quad (9)$$

In Equation (7), (8) and (9), $i \in [0, 9]$ represents the number of the geometry, $j \in [0, 5]$ denotes the number of feature.

B. Extracting fused handcraft feature

The processes of the fusing handcraft features are: we conduct semantic analysis on geometric features, and then establish the feature intensity according to the decision rules to obtain the fused handcraft features.

(1) Semantic feature description

Semantic features include analysis of the geometry, position, explicit or implicit attributes of the expression.

Definition: $\mathbf{M} = \{m_{a,b}^u \mid 1 \le a \le A, 1 \le b \le B, 1 \le u \le U\}$ is a set, the expression has A features, a is the feature number, a feature has U attributes, u is the attribute number, and an attribute can be divided into B intensity levels, b is intensity level number; $m_{a,b}^u$ is the semantic feature of an expression, and $\mathbf{M}$ is the semantic description set of expressions.

There are 59 semantic descriptions of a frame, which can be represent by $\mathbf{YU}=\{\mathbf{yu}_1, \mathbf{yu}_2, \ldots \mathbf{yu}_i, \mathbf{yu}_{59}\}$, $\mathbf{yu}_i=\{m_{i,1}, m_{i,2}, m_{i,3}\}$, $i \in [1, 59]$, $m_{i,1}$, $m_{i,2}$, and $m_{i,3}$, denote the semantic intensity of small, medium, and large in the *ith* feature.

(2) Decision rules

We set the decision rule set as:

The eigenvalues in the training set $\mathbf{f}$ are sorted in ascending order to obtain the fused eigenvalue set $\mathbf{PF}$:

$$\mathbf{PF} = \{\mathbf{pf}_1, \mathbf{pf}_2, ..., \mathbf{pf}_i, ..., \mathbf{pf}_{59}\} \quad (10)$$

where $\mathbf{pf}_i$ is the eigenvalue of $\mathbf{yu}_i$, $i \in [1,59]$. Each $\mathbf{pf}_i$ can be divided equally to the intensity of small, medium, and large.

(3) Fused handcraft features

We establish the semantic description set for each expression geometric triangle $\mathbf{tr}_i$, as shown in Table 1. Then, we calculate the mean value *TR_AVG* of 6 eigenvalues of $\mathbf{tr}_i$ as Equation (11) and standard deviation *TR_SD* of $\mathbf{tr}_i$ as Equation (12).

$$TR\_AVG = \left(d_{1+i\times6} + d_{2+i\times6} + d_{3+i\times6} + r_{4+i\times6} + r_{5+i\times6} + r_{6+i\times6}\right) / 6 \quad (11)$$

$$TR\_SD = \sqrt{\left(\sum_{j=1}^{3} (d_{j+i\times6} - TR\_AVG)^2 + \sum_{j=4}^{6} (r_{m+i\times6} - TR\_AVG)^2\right) \Big/ 6} \quad (12)$$

where $d_{j+i\times6}$ and $r_{m+i\times6}$ are the *jth* distance feature and the *mth* angular feature respectively, $j \in [1, 3]$, $m \in [4,6]$, $i \in [0,9]$.

The 10 geometric features of each frame are sorted in ascending order according to *TR_SD*. The top 5 geometric features of each expression are regarded as the optimal geometric feature set, as shown in Table 2.

Finally, according to the statistical results of feature intensity in various expressions, the largest number of feature intensity is considered as the intensity of a feature, that is, the fused features of each emotions that are shown as Table 3-8.

TABLE I
THE SEMANTIC FEATURE DESCRIPTION SET

Geometric feature	The semantic feature set
$\mathbf{tr}_0$	$\mathbf{yu}_1 \wedge \mathbf{yu}_2 \wedge \mathbf{yu}_3 \wedge \mathbf{yu}_4 \wedge \mathbf{yu}_5 \wedge \mathbf{yu}_6$
$\mathbf{tr}_1$	$\mathbf{yu}_7 \wedge \mathbf{yu}_8 \wedge \mathbf{yu}_9 \wedge \mathbf{yu}_{10} \wedge \mathbf{yu}_{11} \wedge \mathbf{yu}_{12}$
$\mathbf{tr}_2$	$\mathbf{yu}_{13} \wedge \mathbf{yu}_{14} \wedge \mathbf{yu}_{15} \wedge \mathbf{yu}_{16} \wedge \mathbf{yu}_{17} \wedge \mathbf{yu}_{18}$
$\mathbf{tr}_3$	$\mathbf{yu}_{19} \wedge \mathbf{yu}_{20} \wedge \mathbf{yu}_{21} \wedge \mathbf{yu}_{22} \wedge \mathbf{yu}_{23} \wedge \mathbf{yu}_{24}$
$\mathbf{tr}_4$	$\mathbf{yu}_{25} \wedge \mathbf{yu}_{26} \wedge \mathbf{yu}_{27} \wedge \mathbf{yu}_{28} \wedge \mathbf{yu}_{29} \wedge \mathbf{yu}_{30}$
$\mathbf{tr}_5$	$\mathbf{yu}_{31} \wedge \mathbf{yu}_{32} \wedge \mathbf{yu}_3 \wedge \mathbf{yu}_{33} \wedge \mathbf{yu}_{34} \wedge \mathbf{yu}_{35}$
$\mathbf{tr}_6$	$\mathbf{yu}_{36} \wedge \mathbf{yu}_{37} \wedge \mathbf{yu}_{38} \wedge \mathbf{yu}_{39} \wedge \mathbf{yu}_{40} \wedge \mathbf{yu}_{41}$
$\mathbf{tr}_7$	$\mathbf{yu}_{42} \wedge \mathbf{yu}_{43} \wedge \mathbf{yu}_{44} \wedge \mathbf{yu}_{45} \wedge \mathbf{yu}_{46} \wedge \mathbf{yu}_{47}$
$\mathbf{tr}_8$	$\mathbf{yu}_{48} \wedge \mathbf{yu}_{49} \wedge \mathbf{yu}_{50} \wedge \mathbf{yu}_{51} \wedge \mathbf{yu}_{52} \wedge \mathbf{yu}_{53}$
$\mathbf{tr}_9$	$\mathbf{yu}_{54} \wedge \mathbf{yu}_{55} \wedge \mathbf{yu}_{56} \wedge \mathbf{yu}_{57} \wedge \mathbf{yu}_{58} \wedge \mathbf{yu}_{59}$

TABLE II
THE OPTIMAL GEOMETRIC SET OF VARIOUS EMOTIONS

Emotion	The optimal geometric feature set
Happiness	$\mathbf{tr}_9 \wedge \mathbf{tr}_1 \wedge \mathbf{tr}_5 \wedge \mathbf{tr}_2 \wedge \mathbf{tr}_6$
Anger	$\mathbf{tr}_1 \wedge \mathbf{tr}_2 \wedge \mathbf{tr}_7 \wedge \mathbf{tr}_6 \wedge \mathbf{tr}_3$
Disgust	$\mathbf{tr}_8 \wedge \mathbf{tr}_6 \wedge \mathbf{tr}_2 \wedge \mathbf{tr}_3 \wedge \mathbf{tr}_4$
Sadness	$\mathbf{tr}_2 \wedge \mathbf{tr}_1 \wedge \mathbf{tr}_6 \wedge \mathbf{tr}_3 \wedge \mathbf{tr}_4$
Fear	$\mathbf{tr}_9 \wedge \mathbf{tr}_8 \wedge \mathbf{tr}_1 \wedge \mathbf{tr}_2 \wedge \mathbf{tr}_6$
Surprise	$\mathbf{tr}_1 \wedge \mathbf{tr}_5 \wedge \mathbf{tr}_2 \wedge \mathbf{tr}_3 \wedge \mathbf{tr}_9$

TABLE III
THE FUSED FEATURE SET OF EMOTION "HAPPINESS"

Geometric feature	The fused feature set
$\mathbf{tr}_9$	$m_{54,3} \wedge m_{55,3} \wedge m_{56,3} \wedge m_{57,3} \wedge m_{58,1} \wedge m_{59,1}$
$\mathbf{tr}_1$	$m_{7,2} \wedge m_{8,2} \wedge m_{9,3} \wedge m_{10,3} \wedge m_{11,2} \wedge m_{12,1}$
$\mathbf{tr}_5$	$m_{31,1} \wedge m_{32,3} \wedge m_{3,3} \wedge m_{33,3} \wedge m_{34,1} \wedge m_{35,1}$
$\mathbf{tr}_2$	$m_{13,2} \wedge m_{14,3} \wedge m_{15,3} \wedge m_{16,3} \wedge m_{17,2} \wedge m_{18,1}$
$\mathbf{tr}_6$	$m_{36,2} \wedge m_{37,2} \wedge m_{38,1} \wedge m_{39,3} \wedge m_{40,1} \wedge m_{41,1}$

TABLE IV
THE FUSED FEATURE SET OF EMOTION "ANGER"

Geometric feature	The fused feature set
$\mathbf{tr}_1$	$m_{7,1} \wedge m_{8,1} \wedge m_{9,1} \wedge m_{10,2} \wedge m_{11,3} \wedge m_{12,1}$
$\mathbf{tr}_2$	$m_{13,1} \wedge m_{14,2} \wedge m_{15,1} \wedge m_{16,2} \wedge m_{17,3} \wedge m_{18,1}$
$\mathbf{tr}_7$	$m_{42,1} \wedge m_{43,1} \wedge m_{44,1} \wedge m_{45,3} \wedge m_{46,1} \wedge m_{47,1}$
$\mathbf{tr}_6$	$m_{36,1} \wedge m_{37,1} \wedge m_{38,2} \wedge m_{39,2} \wedge m_{40,2} \wedge m_{41,2}$
$\mathbf{tr}_3$	$m_{19,1} \wedge m_{20,1} \wedge m_{21,1} \wedge m_{22,3} \wedge m_{23,1} \wedge m_{24,1}$

TABLE V
THE FUSED FEATURE SET OF EMOTION "DISGUST"

Geometric feature	The fused feature set
$\mathbf{tr}_8$	$m_{48,1} \wedge m_{49,1} \wedge m_{50,2} \wedge m_{51,2} \wedge m_{52,2} \wedge m_{53,3}$
$\mathbf{tr}_6$	$m_{36,1} \wedge m_{37,1} \wedge m_{38,2} \wedge m_{39,2} \wedge m_{40,2} \wedge m_{41,2}$
$\mathbf{tr}_2$	$m_{13,1} \wedge m_{14,1} \wedge m_{15,1} \wedge m_{16,2} \wedge m_{17,3} \wedge m_{18,1}$
$\mathbf{tr}_3$	$m_{19,1} \wedge m_{20,1} \wedge m_{21,1} \wedge m_{22,3} \wedge m_{23,1} \wedge m_{24,1}$
$\mathbf{tr}_4$	$m_{25,1} \wedge m_{26,1} \wedge m_{27,1} \wedge m_{28,3} \wedge m_{29,3} \wedge m_{30,1}$

TABLE VI
THE FUSED FEATURE SET OF EMOTION "SADNESS"

Geometric feature	The fused feature set
$\mathbf{tr}_2$	$m_{13,2} \wedge m_{14,3} \wedge m_{15,1} \wedge m_{16,1} \wedge m_{17,3} \wedge m_{18,3}$
$\mathbf{tr}_1$	$m_{7,3} \wedge m_{8,3} \wedge m_{9,1} \wedge m_{10,1} \wedge m_{11,3} \wedge m_{12,3}$
$\mathbf{tr}_6$	$m_{36,2} \wedge m_{37,2} \wedge m_{38,1} \wedge m_{39,1} \wedge m_{40,3} \wedge m_{41,3}$
$\mathbf{tr}_3$	$m_{19,2} \wedge m_{20,3} \wedge m_{21,1} \wedge m_{22,2} \wedge m_{23,2} \wedge m_{24,3}$
$\mathbf{tr}_9$	$m_{54,1} \wedge m_{55,1} \wedge m_{56,1} \wedge m_{57,1} \wedge m_{58,3} \wedge m_{59,3}$

TABLE VII
THE FUSED FEATURE SET OF EMOTION "FEAR"

Geometric feature	The fused feature set
$\mathbf{tr}_9$	$m_{54,2} \wedge m_{55,2} \wedge m_{56,3} \wedge m_{57,2} \wedge m_{58,2} \wedge m_{59,2}$
$\mathbf{tr}_8$	$m_{48,3} \wedge m_{49,2} \wedge m_{50,3} \wedge m_{51,3} \wedge m_{52,2} \wedge m_{53,2}$
$\mathbf{tr}_1$	$m_{7,3} \wedge m_{8,3} \wedge m_{9,3} \wedge m_{10,3} \wedge m_{11,2} \wedge m_{12,2}$
$\mathbf{tr}_2$	$m_{13,3} \wedge m_{14,3} \wedge m_{15,2} \wedge m_{16,3} \wedge m_{17,2} \wedge m_{18,2}$
$\mathbf{tr}_6$	$m_{36,3} \wedge m_{37,2} \wedge m_{38,3} \wedge m_{39,3} \wedge m_{40,1} \wedge m_{41,2}$

TABLE VIII
THE FUSED FEATURE SET OF EMOTION "SURPRISE"

Geometric feature	The fused feature set
$\mathbf{tr}_1$	$m_{7,3} \wedge m_{8,3} \wedge m_{9,2} \wedge m_{10,1} \wedge m_{11,1} \wedge m_{12,3}$
$\mathbf{tr}_5$	$m_{31,3} \wedge m_{32,3} \wedge m_{3,1} \wedge m_{33,1} \wedge m_{34,3} \wedge m_{35,3}$
$\mathbf{tr}_2$	$m_{13,3} \wedge m_{14,3} \wedge m_{15,3} \wedge m_{16,1} \wedge m_{17,1} \wedge m_{18,3}$
$\mathbf{tr}_3$	$m_{19,3} \wedge m_{20,3} \wedge m_{21,3} \wedge m_{22,1} \wedge m_{23,3} \wedge m_{24,3}$
$\mathbf{tr}_9$	$m_{54,3} \wedge m_{55,3} \wedge m_{56,1} \wedge m_{57,1} \wedge m_{58,3} \wedge m_{59,3}$

IV. EXPERIMENTS

In this section, we provide a detailed experimental analysis of the proposed method and compare it with some state-of-the-arts.

A. Datasets

In this paper, the CK + (Extended Cohn-Kanade) and MMI (M & M Initiative) datasets are selected for dynamic expression recognition. The ratio of the testing set and training set is divided as 3: 7.

The CK+ dataset [23] was released in 2010. The dataset contains 123 people, a total of 593 expression image sequences. There are divided into 7 categories, namely Happiness, Anger, Disgust, Sadness, Fear, Surprise, and Scorn. The resolution of frames is 640*490. This paper selects all of the sequences in CK+ for the experiment.

The MMI dataset [24] contains a total of 2903 expression sequences, of which 44% are women and 56% are men.

B. Fused handcraft method validation

The validations of the proposed fused handcraft method are performed on a single dataset, a cross-dataset, and hybrid dataset. The single dataset means that an experiment is conducted only on one dataset. The cross dataset means that we select 3/4 data from one dataset as a training set and 1/4 data from the other dataset as a testing set randomly.

We set the parameters for SVM: linear kernel function as kernel function, C_SVC as classifier type, and iteration termination condition 100. In Figure 5-9, SU, FE, HA, SA, DI, AN denote surprise, fear, happiness, sadness, disgust, and anger respectively.

(1) Validation with a single dataset

We conduct experiments on CK+ and MMI 10 times respectively, and the average accuracy is taken as the final result, as shown in Figure 5 and Figure 6.

It can be seen from Figure 5, the results of sadness and anger are the lowest, which are 97% and 96.14%. The recognition rates of other expressions are higher than 99%. The accuracy of surprise, happiness, and disgust can reach 100%. The average recognition result in CK+ is 98.73%.

In Figure 6, the results of fear and sadness are the lowest, which are 70.16% and 96.14% respectively. The recognition rates of surprise, happiness, and disgust are all above 93%. The average recognition rate obtained in MMI is 87.36%.

(2) Validation with the cross dataset

Figure 7 shows the result by setting CK+ as a training set and MMI as a testing set. It can be seen that the recognition accuracy of fear and sadness are the lowest due to these two expressions have the similar movements of shutting up and eyes widening, which lead to misrecognition. The accuracy of the other expressions is higher than 66%, and the accuracy of anger is up to 100%. The average recognition rate is 71.70%.

Figure 8 shows the recognition result by setting MMI as a training set and CK+ as a testing set. It can be seen from the experimental results that the accuracy of anger and fear are lower due to the large differences in the expression movements of CK+ and MMI. The recognition accuracy of other expressions is all above 70%. The accuracy of disgust reaches 90.9% because that the movement of eyebrows and eyes in CK + is more obvious than in MMI. The average recognition rate obtained in this cross dataset is 76.19%.

(3) Validation with the hybrid dataset

The recognition results of the hybrid datasets are shown in Figure 9. We can see that the accuracy of fear is the lowest. This is because that fear may be confused with sadness since they all have the same movements of the mouth stretching up and down. The recognition rate of other expressions is higher than 76%. The average recognition rate is 80.27%.

C. Comparison with state-of-the-arts

To further validate the effectiveness of the proposed method, we compare the performance with states-of-the-arts on several datasets, as shown in Table 9 and Table 10.

In literature [6], STM-LBP is susceptible to noise, which leads to low recognition accuracy. The method in [7] combined STWLD and block optical flow histogram to extract features which can capture the changes of pixels. The GSP-HOG [8] has a large texture deviation as the image resolution changes. In [13], the method only extracted geometric features, neglecting the global features of other facial regions. The LBP-TOP + BOW in [18] and LBP-TOP + CNN in [15] only extracted texture features and ignored the reprehensive geometric features. The 3DCNN-DAP [26] and CNN+CRF [25] have fewer network layers and the performance behave badly. The HDNN [27] processed geometric features without considering other features. The proposed method fulfills the extraction of geometric and semantic information by fusing these features, which are beneficial to the expression recognition task.

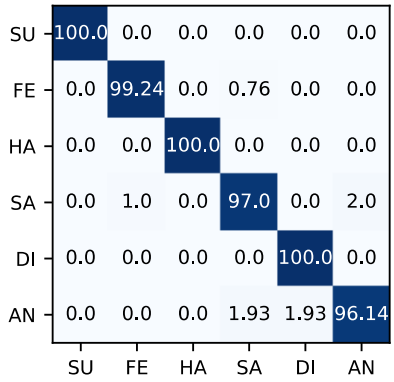

FIG. 5. CONFUSION MATRIX ON CK+

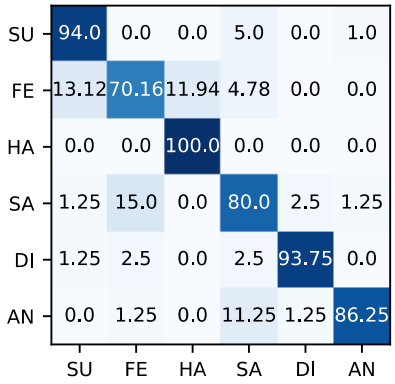

FIG. 6. CONFUSION MATRIX ON MMI

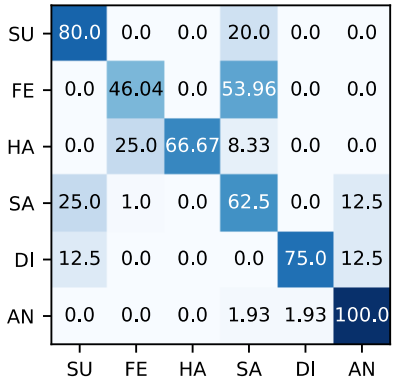

FIG. 7. CONFUSION MATRIX ON CK+ AS TRAINING SET, MMI AS TESTING SET

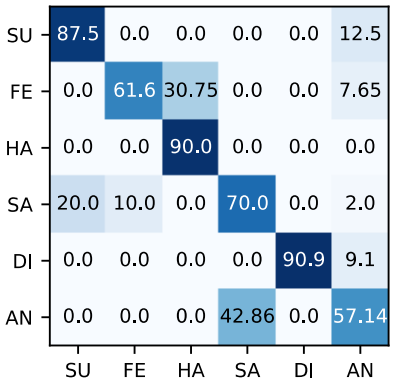

FIG. 8. CONFUSION MATRIX ON MMI AS TRAINING SET, CK+ AS TESTING SET

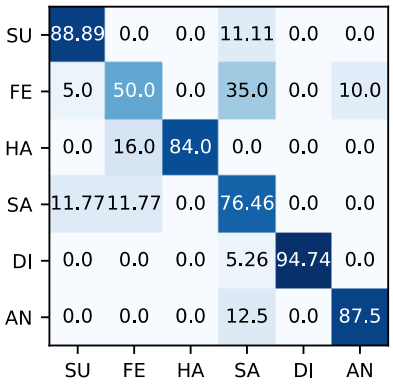

FIG. 9. CONFUSION MATRIX ON HYBRID DATASET

TABLE IX
COMPARISON WITH STATE-OF-THE-ARTS ON MMI (%)

Approaches	Input	Accuracy
STM-LBP[6]	image-based	71.92
STWLD [7]	image-based	71.43
Salient geometric features [13]	image-based	77.22
STC-NLSTM[14]	sequence-based	85.20
CNN+DBN[17]	sequence-based	71.43
3DCNN-DAP[27]	sequence-based	66.40
Proposed method	sequence-based	**87.36**

TABLE X
COMPARISON WITH STATE-OF-THE-ARTS ON CK+ (%)

Approaches	Input	Accuracy
STM-LBP[6]	image-based	95.80
STWLD [7]	image-based	91.60
GSP-HOG[8]	image-based	97.61
6 Distance Feature[10]	image-based	98.00
AAM+OF[11]	image-based	94.17
Salient geometric features [13]	image-based	97.80
LBP-TOP+BOW[18]	sequence-based	97.70
LBP-TOP+CNN[15]	sequence-based	93.76
HDNN[28]	sequence-based	96.46
CNN+CRF[26]	sequence-based	93.04
3DCNN-DAP[27]	sequence-based	92.40
Proposed method	sequence-based	**98.73**

V. CONCLUSIONS

In this paper, we have presented a novel approach for dynamic facial expression recognition via feature fusion

method, which can recognize the expressions with higher accuracy. Both the geometric features and semantic features are fused to identify the expressions by learning features not only the reprehensive locations but also the skin changes from expression sequences. Particularly, the feature fusion framework has been established to identify the dynamic expressions successfully by extracting low dimension robust features from data. In the end, we conduct extensive validation experiments to demonstrate the proposed method. The comparison results show that the proposed approach outperforms the other state-of-the-arts in terms of 6-class facial expression recognition.

REFERENCES

[1] A. Mehrabian, "Communication without words," Psychology Today, vol. 2, no. 4, pp. 53-56, 1968.

[2] R. Sharma, and B. Kaushik, "Facial expression recognition: a survey," International Journal of Computer Applications, vol. 153, no. 10, pp. 32-36, 2016.

[3] C. Henriquez, F. Briceno, and D. Salcedo, "Unsupervised model for aspect-based sentiment analysis in Spanish," IAENG International Journal of Computer Science, vol. 46, no. 3, pp. 430-438, 2019.

[4] D. Yang, A. Alsadoon, P. W. C. Prasad, A. K. Singh, and A. Elchouemi, "An emotion recognition model based on facial recognition in virtual learning environment," Procedia Computer Science, vol. 125, no. 1, pp. 2-10, 2018.

[5] G. Y. Zhao, and M. Pietikainen, "Dynamic texture recognition using Local Binary Patterns with an application to facial expressions," IEEE Transactions on Pattern Analysis & Machine Intelligence, vol. 29, no. 6, pp. 915-928, 2007.

[6] L. Zhao, Z. Wang, and G. Zhang, "Facial expression recognition from video sequences based on Spatial-Temporal Motion Local Binary Pattern and Gabor multiorientation fusion histogram," Mathematical Problems in Engineering, no. 13, pp. 1-12, 2017.

[7] X. Wang, C. Xia, M. Hu, and F. Ren, "Facial expression recognition based on the fusion of spatio-temporal features in video sequences, "Journal of Electronics & Information Technology, vol. 40, no. 3, pp. 626-632, 2018.

[8] H. K. Meena, S. D. Joshi, and K. K. Sharma, "Facial expression recognition using Graph Signal Processing on HOG," IETE Journal of Research, no. 65, pp. 1-7, 2019.

[9] Y. Li, S. Wang, Y. Zhao, and Q. Ji, "Simultaneous facial feature tracking and facial expression recognition," IEEE Transactions on Image Processing, vol. 22, no. 7, pp. 2559-2573, 2013.

[10] F. Z. Salmam, A. Madani, and M. Kissi, "New distances combination for facial expression recognition from image sequences," IEEE International Conference of Computer Systems and Applications, June 2017, pp: 1-6.

[11] H. Shao, Y. Wang, and Y. J. Wang, "Dynamic image sequences expression recognition based on active appearance model and optical flow," Computer Engineering and Design, vol. 38, no. 6, pp. 1642-1647, 2017.

[12] A. Durmuşoğlu, and Y. Kahraman, "Facial expression recognition using geometric features," IEEE International Conference on Systems, Signals and Image Processing, May 2016, pp: 1-5.

[13] D. Ghimire, J. Lee, Z. N. Li, and S. Jeong, "Recognition of facial expressions based on salient geometric features and Support Vector Machines," Multimedia Tools and Applications, vol. 76, no. 6, pp. 7921-7946, 2016.

[14] A. Krizhevsky, I. Sutskever, and G. Hinton, "ImageNet classification with deep convolutional neural networks," Advances in Neural Information Processing Systems, vol. 25, pp. 1097-1105, 2012.

[15] K. Simonyan, and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," Computer Science, vol. 8, no. 15, pp. 1409-1423, 2014.

[16] C. Szegedy, W. Liu, Y. Jia, *P.* Sermanet, and A. Rabinovich, "Going deeper with convolutions," IEEE Conference on Computer Vision &Pattern Recognition, 2015, pp: 1-9.

[17] K. He, X. Zhang, S. Ren, and S. Jian, "Deep residual learning for image recognition," IEEE Conference on Computer Vision & Pattern Recognition, 2016, pp: 770-778.

[18] A.S. Spizhevoy, "Robust dynamic facial expressions recognition using LBP-TOP descriptors and Bag-of-Words classification model," Pattern Recognition & Image Analysis, vol.26, no. 1, pp. 216-220, 2016.

[19] Z. D. Li, Q. L. Zhang, X. D. Duan, and W. Wei, "Semantic knowledge based on fuzzy system for describing facial expression," IEEE Chinese Control Conference, July 2017, pp: 9865-9870.

[20] B. F. Hu, and Y. C. Huang, "A novel facial expression recognition method based on semantic knowledge of analytical hierarchy process," Journal of Image and Graphics, vol.16, no. 3, pp. 420-426, 2011.

[21] X. D. Duan, Z. D. Li, and C. R. Wang, "Multi-ethnic face semantic description and mining method based on AFS," Chinese Journal of Computers, vol. 39, no. 7, pp. 1435-1450, 2016.

[22] https://github.com/ShiqiYu/libfacedetection

[23] P. Lucey, J. F. Cohn, T. Kanade, *J. Saragih*, and I. Matthews, "The extended Cohn-Kanade dataset (CK+): A complete dataset for action unit and emotion-specified expression," Computer Vision & Pattern Recognition Workshops, pp. 94-101, 2010.

[24] M. Pantic, M. Valstar, R. Rademaker, and M. Ludo, "Web-based database for facial expression analysis," IEEE International Conference on Multimedia & Expo., 2005, pp: 317-321.

[25] B. Hasani and M. H. Mahoor, "Spatio-temporal facial expression recognition using Convolutional Neural Networks and Conditional Random Fields," IEEE International Conference on Automatic Face & Gesture Recognition, 2017, pp: 790-795.

[26] M. Liu, S. Li, S. Shan, R. Wang, and X. Chen, "Deeply learning deformable facial action parts model for dynamic expression analysis," Asian Conference on Computer Vision, 2014, pp: 143-157.

[27] D. Feng, and F. Ren, "Dynamic facial expression recognition based on Two-Stream-CNN with LBP-TOP," IEEE International Conference on Cloud Computing and Intelligence Systems, December 2019, pp: 355-359.