Facial Expression Recognition Based on Anomaly Detection and Multispectral Imaging

Kan Hong

Abstract—**This study proposes a facial expression recognition model based on anomaly features and multispectral imaging (AF-MSI) to extract spectral, spatial, anomalous, and temporal features for recognizing facial expressions. To ensure that the extracted anomaly is consistent with the characteristics of a human face and expression, the background of the face required by anomaly extraction is reconstructed using the spatial characteristics and facial spectral signal correlation. Then, improved spectral correlation is employed to reconstruct a sparse coding model based on spectral images, thereby constructing a sparse dictionary. A temporal anomaly feature signal is divided into short intervals with a fixed number of frames. Further, the descriptor and correlation of the short-term sequence features are obtained by the sparse dictionary's base vector. The characteristic signals with good correlation are selected as representative spectral-spatial-anomaly-temporal features for subsequent training and learning. The proposed model is verified experimentally on different datasets and compared with the state-of-the-art methods. The experimental results demonstrate that the proposed facial expression recognition model can achieve an accuracy of over 86% and has a significant advantage over comparable algorithms.**

Index Terms—**Multispectral imaging, facial expression, anomaly detection**

I. Introduction

Emotion recognition has been attracting increasing attention in the field of human–computer interaction since emotion recognition became an important part of artificial intelligence with broad application to medical science and health management [1-3]. Emotion recognition based on facial expression has always been popular due to its complete noncontact and noninvasive features [4-7]. The current feature extraction methods for facial expression recognition (FER) focus on extracting the hand-crafted and learned features. Based on the technique for image data acquisition, the FER methods can be roughly divided into RGB (the Red-Green-Blue) sensor-based methods and other imaging sensor-based algorithms.

Manuscript received March 25, 2024; revised September 12, 2024.

This work was supported by the National Natural Science Foundation of China (Grant No. 61866015), the Science and Technology Project of Education Department of Jiangxi Province (Grant No. GJJ180598), the Natural Science Foundation of Jiangxi Provincial (Grant No. 20224BAB201029), and the Science and Technology Project of the Education Department of Jiangxi Province (Grant No. GJJ180612).

Kan Hong is a lecturer in the School of Software and Internet of Things Engineering, Jiangxi University of Finance and Economics, Nanchang 330000, China. (Phone: +86 13576061058; email: 179739527@qq.com).

A. Hand-crafted and Learned Feature-based FER

The hand-crafted features can be divided into static and temporal expression features according to data properties. The hand-crafted features are initially obtained from static data. Hybrid features, geometric features, sparse representation, genetic algorithm, principal component analysis, and local binary model algorithm coupled with computer vision models have been extensively used in FER [8–11]. The facial structural features [17–19] and geometric differences [20], [21] in geometric features [12–16], [50–54] have also been used to study local binary patterns, local phase quantification, and Gabor wavelets. Recently, the methods for obtaining facial expression features from time-series imaging data have been improved to realize dynamic emotion recognition progress [22–26]. Many methods can be used to track facial changes and movement trends in real time to obtain dynamic emotion recognition features [28–31]. Spatiotemporal features have also been popular because they can obtain facial morphology and identify action units (AUs) [32–37], [48–50]. Moreover, expression dynamics in facial expression sequences, including dynamic appearance descriptors, have also received increasing attention [38-41].

The hand-crafted features can also be divided into appearance and geometrical features according to the feature acquisition region [12–14]. Many subtle changes in static or dynamic features are still challenging to detect and track by hand-crafted feature-based methods, even when geometric features are intuitive and efficient in describing facial expressions. Appearance features, which use the intensity signal of images, particularly global appearance features [1] [43-47], [74], [79], are advantageous in feature detection and noise stability tasks. However, the methods based on these features have poor performance in describing details and might be impacted by various facial characteristics.

Unlike the hand-crafted feature-based methods, the development of artificial intelligence has increased the application of deep learning-based algorithms to the extraction of relevant features of facial expressions in the machine vision field [57]. Deep learning-based models, including recurrent neural networks [58],[59], long short-term memory (LSTM) models [61], and Bayesian networks [62], [64], [65], have been widely used in the FER. In addition, deep learning-based models, such as convolutional neural networks[66],[67], various neural networks [69], and boosted deep belief networks (DBNs) [71], have also been combined with machine learning-based algorithms to extract AUs and expression classification features [71-81]. However, spectral features in the RGB-band time-series data have not been effectively considered in previous works.

B. Imaging Sensor-based FER

The FER uses imaging sensors to acquire features. In addition to usually adopted RGB sensors (e.g., Cer [3], Ldl-Alsg (L-A) [56], Stfrl [112], It-Rbm [113], Lbvcnn [114], Ada-CM[26], and Daugn [116]), thermal and other imaging sensors [82–85] have been increasingly used in recent years. The FER based on thermal imaging relies on the human body temperature to extract features. Some studies have adopted machine vision and deep learning-based methods to extract disordered temperature laws and head motion from specific areas of the human face to identify various facial expressions and emotions [82–85]. Apart from the FER, the features of temperature changes in the forehead, periodic region, and nose of the human face can help to recognize a specific emotion. For instance, emotional stress recognition based on thermal imaging has made encouraging progress in the last 20 years [87–92].

Further, in the last 10 years, spectral imaging technology has emerged in affective computing. Unlike traditional imaging technologies, spectral imaging systems can acquire both spatial and spectral information. This information can provide sufficient spectral characteristics for target recognition, bio-information extraction, and geographic investigation [94–100]. In the Earth observation and remote sensing tasks, anomaly detection is feasible due to the rich spectral information [105], [106]. Moreover, the features indicating facial tissue oxygen saturation can be extracted by hyperspectral and multispectral imaging (MSI) technology to identify human stress states [102–104].

Spectral imaging is based on narrow-band data structure, and it combines spectral technology to detect a target's two-dimensional (2D) geometric spatial information and one-dimensional (1D) spectral information. It also provides continuous and narrow-band image data with high spectral resolution. Spectral imaging has been developing rapidly in recent years and includes grating beam splitting, acousto-optic tunable filter splitting, and prism splitting. The crucial characteristic of spectral imaging is the combination of imaging and spectral detection. When the spatial characteristics of a target are imaged, each spatial pixel is dispersed to form dozens or even hundreds of narrow bands for continuous spectral coverage. Therefore, these data can be described as a three-dimensional (3D) data block. Thus, a sample's spatial and spectral data are integrated into an image set. Spectral imaging has been widely applied to the fields of food safety, medical diagnosis, and remote aerospace sensing. Compared to traditional three-band RGB imaging, spectral imaging technology provides more useful information and references for structural changes caused by facial expressions. This study introduces multispectral technology to FER.

C. Challenges and Contributions

Previous research on the FER has made many important achievements. However, achieving accurate and efficient facial expression faces the following challenges:

(1) Global appearance features are robust to noise and can realize complete facial expression detection. However, these features cannot be used to describe the details of facial expression and might impact facial characteristics. In addition, the method of using these features lacks flexibility in feature selection and fails to obtain an acceptable combination of facial structure and expression features.

(2) Spatiotemporal data obtained by real-time imaging have been combined with deep learning-based algorithms to extract temporal features. However, the spectral features in the RGB-band time-series data cannot be considered effectively in this way, which affects expression recognition results. Moreover, the correlation between the bands cannot be correlated with facial expression features due to the limited number of channels (i.e., there are only three bands).

(3) The intensity variation in traditional RGB imaging also limits the accuracy of FER. In addition, different personal conditions can cause significant uncertainty in facial expression intensity. Therefore, an appropriate selection of band data can improve the recognition rate.

To overcome the aforementioned challenges, this study proposes the AF-MSI model based on anomaly features (AF) and multispectral imaging (MSI). The proposed algorithm uses MSI to acquire multispectral time-series images of the human face with different expressions. Anomaly detection is employed to extract abnormal targets based on significant spectral differences from the surroundings. However, the human face's structure inevitably changes with facial expressions, which provides an opportunity to apply the proposed anomaly detection-based feature extraction method. Then, the anomaly features are combined with facial expression features, spectral correlation, and spatial correlation to determine the relationships between the structure anomaly and expression changes. Furthermore, a representative AF is extracted and introduced into the LSTM model for training and recognition. This study employs an adaptive method for spectral band selection to select the band interval directly based on the correlation of the spectral and spatial signals.

The main contributions of this study can be briefly summarized as follows:

(1) This paper proposes an innovative AF-MSI algorithm based on the MSI and AF. The experimental results show that the proposed algorithm can effectively recognize facial expressions and outperform state-of-the-art algorithms.

(2) Anomalies of the human face are considered the feature objects for FER. Unlike previous studies, this study incorporates all spectro-anomalous spatiotemporal characteristics into the proposed model to acquire information on the spatial structure and expression changes of the human face.

(3) The spectral distance (SD) is used together with the spatial distance to form a representative background (BG) to obtain the spatial and spectral anomalies. Different from traditional anomaly detection methods, in this work, spectral and spatial anomalies are detected by using the initial multiband spectral correlation between the spatial and spectral characteristics of a facial contour;

(4) To incorporate representative facial expression data into network training and improve the separability of extracted features, this study rescreens the AFs according to their spatial-spectral correlations. Thus, the representative AFs can accurately reflect the correlation between the feature texture and facial structure;

(5) This study uses spectral imaging to extract facial anomaly signals as FER features for the first time. The proposed algorithm achieves an average accuracy of over 86%.

The rest of this paper is structured as follows. Section II describes the MSI data acquisition and introduces the proposed algorithm. Section III presents the experimental results. Finally, Section IV draws the main conclusions.

II. Proposed AF-MSI Algorithm

The spectral imaging characteristics can improve the sensitivity of AFs to feature expression. Spectral image information can fully reflect the size, shape, structure, and other external characteristics of a target. An image can reflect changes in some facial structures in a certain wavelength range based on a difference in spectral absorption data, where spectral information fully reflects the differences in the physical structure and morphology of the human face. These characteristics make spectral imaging have advantages in reflecting facial structure and geometric changes. Therefore, compared to the RGB imaging-based AU methods, using spectral imaging AFs can better reflect the changes and differences in the facial structure. Moreover, the proposed algorithm obtains AFs directly from the human face through anomaly detection without requiring any prior information. Therefore, unlike the AUs, the AFs belong to the global appearance features and do not require locating specific positions.

This study aims to find a global feature in the proposed algorithm that can be acquired from the entire face. Therefore, facial AF is used as a target feature. An anomaly detection algorithm separates target (i.e., anomaly) information from the image BG and noise. An example of this type of algorithm is the classical Reed–Xiaoli (RX) anomaly detection algorithm [105],[106], which is a local target detection algorithm. The detection window of this algorithm includes the target and BG windows, but it should be noted that the BG window is much larger than the target window. The RX algorithm assumes that the data space whitens and obeys a Gaussian distribution. The window's mean and variance are analyzed and compared with the target's ones to determine the anomaly value.

Considering the anomaly characteristic and previous work [107], this study proposes to apply anomaly detection to feature extraction of FER. In general, an expression is neither smooth nor a single facial area. When a facial expression changes, this change will inevitably bring changes in facial muscles, and the resulting texture changes in the image may become an AF outside the BG. This study deliberately selects a smooth facial area as an example BG (i.e., a facial area with a single BG). A neutral expression can be considered a single BG that conforms to the Gaussian distribution. A single surrounding BG refers to a BG that comprises a single texture, such as skin with limited contours. Therefore, the state of the entire BG appears smoother than the others.

The change rate of the anomaly intensity in a randomly defined region of interest (ROI) during an expression change (from neutral to happiness) is presented in Fig. 1. As shown in Fig. 1, the facial texture change and the variation in the anomaly intensity in this region after an expression change can exceed 25%. Furthermore, the anomaly intensity change between the seven facial expressions is 10%–35%. The aforementioned experiment illustrates that facial texture variations can be extracted using an anomaly detector for facial expression changes. This texture variation can also reflect changes in expressions. Therefore, the AF can be used to recognize a change in an expression.

Although the AF can reflect changes in expressions, this study does not aim to extract features by changing the physical shape of a specific ROI because this is challenging to do in machine vision. Namely, this work aims to employ a global appearance feature. An anomaly of an entire face can be extracted through anomaly detection from the front side of the face. However, an anomaly detection (classical RX) algorithm assumes that the BG of an anomaly follows a multivariate Gaussian distribution, but the entire facial contour or structure does not present a homogeneous local BG. In addition, the facial texture of different people might also increase the complexity of the BG. Although the research object of this study is the whole face, this study reconstructs a new BG based on the facial structural characteristics of the Candide3 model [101] instead of using the entire face as BG. In the Candide3 model, each triangle is defined as a BG and a basic element of the change in a facial expression. Since there has been no evidence that proves that every local BG is homogeneous, not all BGs are suitable for anomaly detection. Therefore, this study uses the correlation between multispectral bands and local outlier factor (LOF) structure [108] to process each BG.

To use both spatial and spectral characteristics fully, the BG is divided into BG^A and BG^B. Each BG is defined based on its relevant spatial and spectral characteristics. Accordingly, the obtained anomaly can better reflect the facial structure and texture changes in facial expression. Moreover, the BG can also be more homogeneous. The spectral anomaly extracted by the spectral band correlation and LOF structure is defined as BG^A, and its features directly become AFs. The anomaly detector and DBN model are applied to BG^B to obtain the remaining spatial AFs. This part of the features is defined as spatial AF. In this way, the spectral and spatial AFs can be extracted by applying the correlation between the spectral bands and the LOF structure. The definitions of BG^B and BG^A guarantee that anomaly based on the spectral structure can be directly used as an FER feature signal; also, the nonhomogeneous local BG phenomenon can be improved.

The temporal features are obtained after anomaly detection. However, this study does not train and learn these features directly. Many temporal features are obtained, and the amount of temporal data useful for FER is unclear. Therefore, this study extracts the representative Afs; namely, the temporal features are divided into multi-short-term sequence feature intervals with a fixed number of frames. Each interval is further decoded by sparse coding and spectral correlation to achieve a descriptor between the intervals. The representative AFs with strong correlation are obtained by calculating the inner product of the short-term sequence feature intervals. These representative AFs reflect the correlation between the feature texture and facial structure. The representative AFs are then fed to the LSTM model to complete the recognition process. The flowchart of the proposed algorithm is shown in Fig. 2.

A. AF Extraction from MSI Data Cube

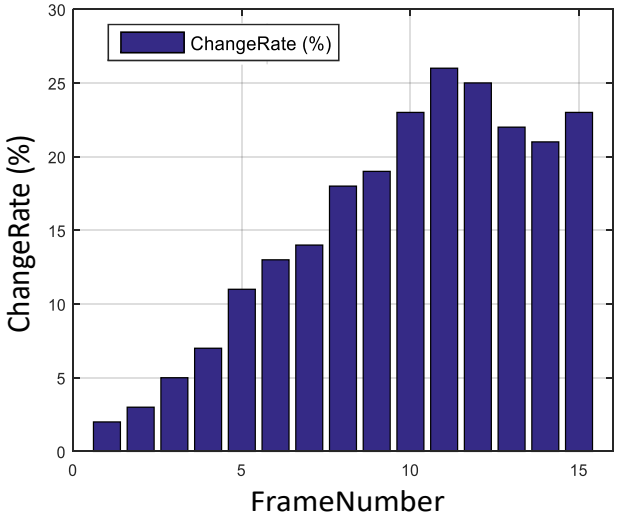

Fig. 1. The changing rate of the anomaly intensity in the ROI region. The frame number denotes the number of image frames when the facial expression changes.

A classical anomaly detector model is the RX algorithm developed using the generalized likelihood ratio test and Mahalanobis distance (MD). This algorithm assumes that the BG pixels around an anomaly target can be modeled by a multivariate Gaussian distribution [105], [106]. Thus, the BG should be initially determined. This study defines a single-frame MSI data cube $M \in R^{L\times m\times n}$, where L denotes the band of an MSI data cube, and m and n define the spatial domain size of a single-band image. The entire MSI data cube can be regarded as an L-dimensional space of $m \times n$ vectors. Therefore, a cube can be defined by

$$M = \left[M_1, M_2, M_3, \ldots M_i, \quad \ldots M_L\right], \tag{1}$$

where M refers to the entire MSI data cube, and M_i represents the collection of pixels for a band.

As discussed previously, the RX algorithm assumes that the BG obeys a multivariate Gaussian distribution, and many algorithms aim to improve the non-homogeneity probability by reducing the BG size. However, this study employs the Candide3 model and divides the face into 168 triangles to consider the structural features of the face. During the experiment conducted in this research, the participant sat still in a high-back chair facing the camera. Although there can be a slight deviation in the fitting of the Candide3 model, the effect of this deviation on the detection result is marginal. Each divided triangle is designed as BG, and the *i*th band signal can be expressed by

$$M_i = \left(\eta_1^i, \eta_2^i, \ldots, \eta_j^i, \ldots \eta_{168}^i\right) \tag{2}$$

where η_j^i is the *j*th triangle of the *i*th band in the MSI data cube; the *j*th triangle data cube (BG_j) can be defined as an $L \times N_j$-dimensional BG matrix; N_j is the total number of pixels of η_j^i.

The *j*th triangle, that is, the *j*th BG, is used as an example to derive our algorithm. Therefore, an MSI cube of BG_j can be defined by

$$BG_j = (\eta_j^1, \eta_j^2, \ldots \eta_j^L) \tag{3}$$

Unlike in traditional MD-based methods, in this work, spatial and spectral anomalies are introduced into anomaly detection. First, an MD-based anomaly is defined as a spatial anomaly. Then, an anomaly calculated using a spectral distance (SD) is regarded as a spectral anomaly to consider the correlation between the spectral bands. Inspired by the LOF model [107], this study determines a spectral anomaly by calculating the SD to determine outliers in the spectral signal. A spatial distance D (*e*, *f*) between any two points *e* and *f* in BG_j and the *k-distance* (KD) are defined. The KD is based on the SD rather than on the spatial distance. Therefore, the SD derived from multiband spectral correlation is defined. Most common correlation calculation methods are based on the Pearson coefficient between two datasets. However, a multispectral dataset represents a group set of multiband datasets with different spatial locations. Therefore, the multivariant correlation method is used to extract the spectral correlation (i.e., the defined SD). This method is based on a zero-lag correlation matrix and adopts techniques based on the random matrix theory [107], [109].

Each point within an MSI cube contains multiband signals of the same wavelength. A group of equal-band correlation matrixes is designed, together with the calculation of spectral correlation (SD) between the selected peripheral p points in BG_j. The spectral signals of p points are defined as $\chi_\kappa(\lambda)$ *(k=1,2,3…p,* λ *=1,2….L)*. Then, the equal-band correlation matrix ψ can be defined by normalizing the band signal as follows:

$$\widetilde{\chi_\kappa}(\lambda) = \frac{\chi_\kappa(\lambda) - \overline{\chi_\kappa(\lambda)}}{\sigma_i(\lambda)} \tag{4}$$

where σ_i is the standard deviation and $\overline{\chi_\kappa(\lambda)}$ represents the average value.

The Pearson correlation coefficient is also used to analyze the parameters, and ψ is defined as follows:

$$\psi_{mn}(\lambda) = \frac{1}{\lambda}\sum_{\lambda=1}^{L} \widetilde{\chi_m}(\lambda)\widetilde{\chi_n}(\lambda) = \langle \widetilde{\chi_m}\widetilde{\chi_n}\rangle_\lambda \tag{5}$$

Since a range $band \in [1, L]$, the equal-band correlation matrix ψ can be expressed by

$$\psi = \frac{1}{L}\tilde{\chi}\tilde{\chi}' \tag{6}$$

where $\tilde{\chi}'$ is the transposition of $\tilde{\chi}$.

The χ_m and χ_n values of different spatial points are included in bivariate measures and are defined by $p(p-1)/2$ independent matrix coefficients ψ_{mn}. Matrix ψ_{mn} represents the cross-correlation between the spectral data of a specific point and that of the other spatial points. In addition, this method is easy to calculate and provides a direct interpretation. The eigenvalue and eigenmatrix provide the joint probability distribution of the similarity magnitude. If spectral signals of all spatial points are irrelevant, then the non-diagonal elements of matrix ψ will equal zero. Conversely, if these signals are consistent, then matrix ψ elements will equal one. Therefore, the size of the maximum eigenvalue of matrix ψ represents the most significant parameter of the correlation. Accordingly, the eigenvalue distribution of the correlation matrix of a spectral signal χ_κ can be expressed by

$$W(\gamma) = \frac{V}{2\pi}\frac{\sqrt{(\gamma_+ - \gamma)(\gamma - \gamma_-)}}{\gamma} \tag{7}$$

where, $\gamma \epsilon [\gamma_-, \gamma_+]$, and V is equal to L/p. Therefore, γ_- and γ_+ can be defined as follows:

$$\gamma_\pm(V) = 1 + \frac{1}{V} \pm \frac{2}{\sqrt{V}} \tag{8}$$

Correlation matrix ψ represents a spectral correlation of a spatial point. Thus, it is required to calculate only the maximum eigenvalue and compare it with γ_+. If the maximum eigenvalue is larger than γ_+, then data are similar

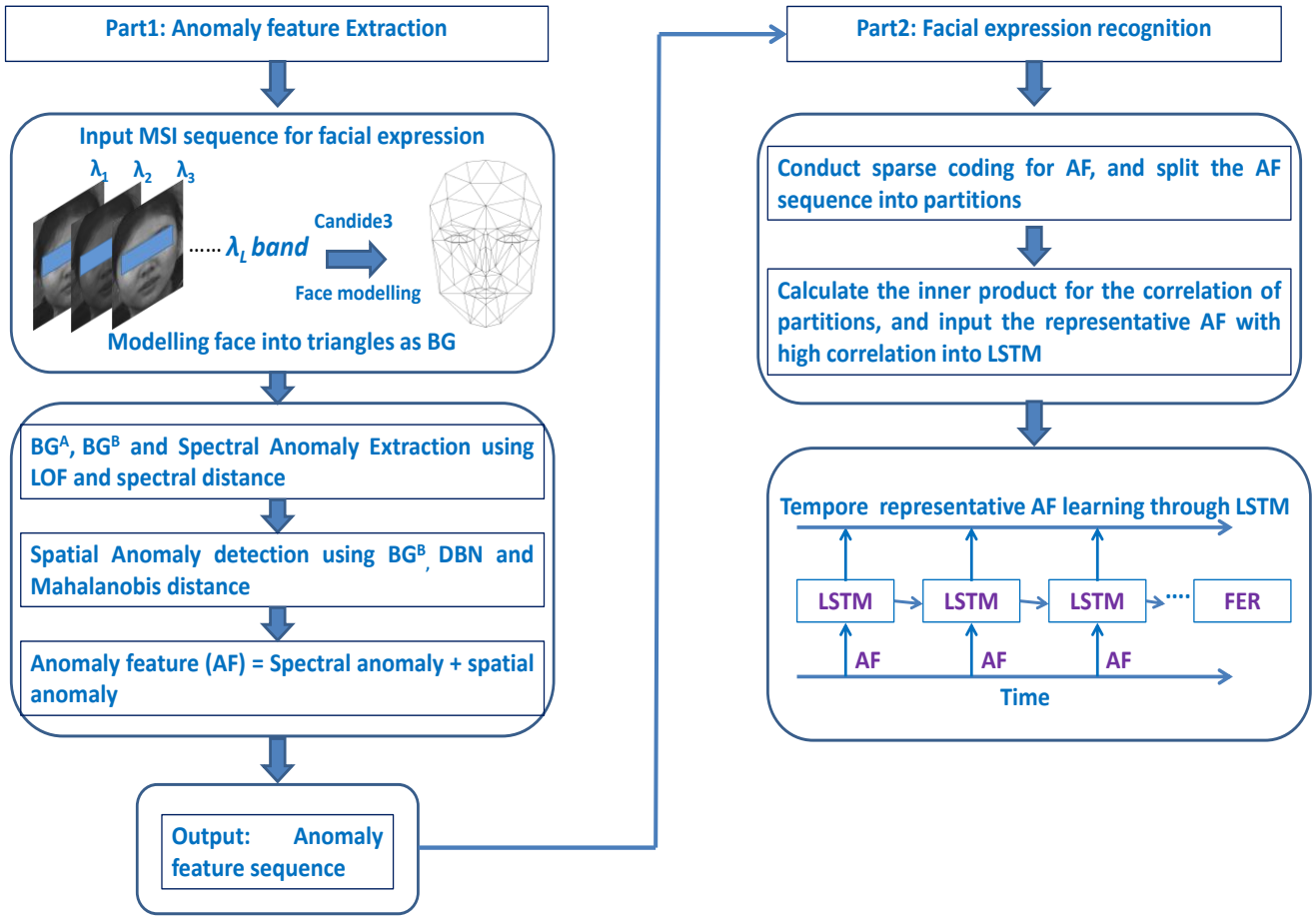

Fig. 2. The flowchart of the AF-MSI algorithm.

and relevant. To this end, this study defines $\gamma_{\max}/\gamma_{+}$ as an SD. When data are relevant, the SD value is larger. Therefore, by using the SD in combination with the LOF model, this study derives a spectral anomaly (BG_j^A) and a spatial anomaly (BG_j^B). In Fig. 3, the blue point is the *e* point, and the surrounding yellow point (e.g., *f* point) is the target point for the SD. Each point in the diagram represents a spectral sequence signal and consists of multiple band data with the same spatial position. It should be noted that the SD is not only calculated at points *e* and *f*. The multivariate correlation method is used to calculate the correlation of multiple points. The SD of point *e* and a peripheral point (PP) are calculated, and the distance of the *k*-nearest points of point *e* is defined as a *k*-distance (KD). The LOF is used to calculate the Euclidean distance, but the SD is calculated based on multivariate correlation. Therefore, the correlation of several object points can be calculated simultaneously.

The specific steps of the BG detection algorithm, presented in Algorithm 1, are as follows:

Step 1: The number of *S* peripheral object points is input into the algorithm to obtain SD; V_S and SD_S are respectively defined as follows:

$$V_S = (PP_1, PP_2, \ldots PP_S, e) \quad (9)$$

$$SD_S = SD(V_S) \quad (10)$$

Step 2: As peripheral points are added successively to V_S until the SD calculation result becomes larger than one, the point recently added to the calculation is regarded as the closest point to the *e* point spectrum, with the one-distance.

Step 3: The spectral data of the other points are used to calculate the SD until the current SD becomes greater than the previous SD. The second closest point to the *e* point spectrum denotes the most recently added point, with the two-distance.

Step 4: Step 3 is repeated until the *k*-nearest point is calculated, and the corresponding SD denotes the KD.

Step 5: The point within the KD of *e* point is set as a *k*-nearest neighbor set $\omega_k(e)$. The reachability distance (RD) of *e* point to an object *f* in $\omega_k(e)$ is defined as follows:

$$RD_k(e,f) = max\{KD(f), SD(e,f)\} \quad (11)$$

where *SD*(*e*, *f*) is the SD between points *e* and *f*.

The local reachability density (LRD) and LOF of point e are defined as follows:

$$LRD_k(e) = 1/\left(\frac{\sum_{f\in\omega_k(e)} RD_k(e,f)}{|\omega_k(e)|}\right) \quad (12)$$

$$LOF_k(e) = \frac{\sum_{f\in\omega_k(e)} LRD_k(f)}{|\omega_k(e)|} / LRD_k(e) \quad (13)$$

If the LOF value is close to one, then the neighborhood density of point *e* is nearly the same, and point *e* is defined as a dense point. In contrast, the density of point *e* is lower than that of its neighborhood points when the LOF is greater than 1. Thus, point e is more likely to be defined as an outlier.

Step 6: The spectral anomaly is calculated using spectral correlation as BG_j^A, and the remaining points in BG_j are defined as BG_j^B.

Algorithm 1 illustrates the details of the BG detection process. The anomaly detection based on the MD of BG_j^B is conducted to obtain spatial anomalies. This study extracts BG_j^A directly from BG_j though the SD calculation. However, this might affect the representative spatial feature structure, so representative features are extracted in BG_j^B. In DBNs, a layer-by-layer approach can be used to extract high-level

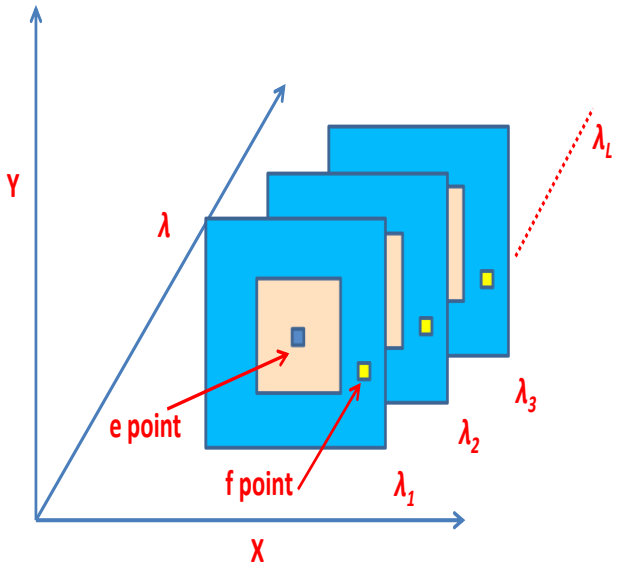

Fig. 3. Schematic of *e* point SD.

features. Therefore, before MD calculation BG_j^B, discriminant and spatial representative features are extracted using the DBN.

Furthermore, the DBN represents a deep nonlinear transformation. The proposed algorithm can capture the potential features in the spectral domain and transform a spectral signal from high to low to reduce the number of data dimensions. The DBN conducts layer-by-layer unsupervised learning, and the network is gradually deepened using one-layer learning of network structure.

In this study, a two-tier DBN and BG_j^B are used as a basic framework and input, respectively. As previously mentioned, the DBN denotes layer-by-layer unsupervised learning. Thus, the output of each hidden layer represents the input of the next hidden layer. The representative features obtained after BG_j^B completing one layer of a DBN (with dimension reduction) BG_j^A can be combined into the DBN model. Accordingly, the feature extraction of the second layer is completed to preserve the spectral structure of feature representation and composite for the spatial structure loss caused by the BG separation. The output features of the first layer can be expressed as follows:

$$h_1 = g\left(\beta_\mu + \sum_{\mu=1}^{\mathrm{T}} W_{\mu\nu} BG_j^B\right) \quad (14)$$

where $W_{\mu\nu}$ is the weight between the input data and a hidden unit, T is the total number of visible units, and β_μ represents the bias value.

Next, BG_j^A is added to the first output, and the second output expression is as follows:

$$h_2 = g\left(b_o + \sum_{\theta=1}^{\Omega} W_{\theta\tau} h_1 + \sum_{\vartheta=1}^{\mathrm{H}} W_{\vartheta\zeta} BG_j^A\right) \quad (15)$$

where $W_{\theta\tau}$ and $W_{\vartheta\zeta}$ denote the weights of the hidden unit, and b_o is the bias; Ω and H define the total number of units.

After h_2 is obtained, the MD is used to extract relevant spatial AFs directly. Thus, the MD can be expressed as follows:

$$D = (h_2 - \mu)^T C_{I.\times I.}^{-1} (h_2 - \mu) \quad (16)$$

where μ is the mean vector, C represents the covariance matrix of the BG, and L is the number of the image's spectral bands.

Algorithm 1: Algorithm for the detection of BG_j^A

Input: BG_j

Output: BG_j^A

1. Calculate the SD for e point and S number of PP.
V_S= (PP_1, PP_2, . . . PP_S, e)
SD_S=SD(V_S)

2. Distance check for *1-distance*
(a) If *SD_S*>1, then *SD_S is 1-distance.*
(b) Else, go to step 1 and add one more PP into *V_S*

3. Distance check for *2-distance*
(a) Add one more *PP* into *V_S*, *SD_{S+1}=SD(V_{S+1})*
(b) If *SD_{S+1}>SD_S, then SD_{S+1}* is 2-distance
(c) Else, go to step 3 and add one more PP into *V_{S+1}*

4. Repeat Step 3 to calculate KD.

5. Calculate the RD, LRD, and LOF with KD.

6. Calculate the spectrum singular point, and determine spectral anomaly BG_j^A

Spatial AFs are obtained by calculating the MD, and the AF represents the collection of spatial and spectral AFs. This study introduces the calculation method of a triangle AF, and the same method is employed for the other triangles. In addition, the AFs of all time-series MSI data cubes are extracted. Therefore, the final output feature is a time-series feature, and an AF cube is defined as $f(\varpi) \in R^{L\times\epsilon\times\iota}$, where ϖ is the number of frames.

B. Facial Expression Recognition Using AF

After obtaining facial anomaly information, this study focuses on expression recognition. The next step determines which part of a temporal anomaly can better reflect changes in local facial expression features. An anomaly MSI cube is a time-series signal, and an anomaly is a global feature. Therefore, it is expected to extract the most useful frame in the time series as a signal for the LSTM, thus reducing signal confusion (many expressions have similar processes). Identifying the features with good similarity (especially between the spectra) before training the samples is beneficial to accuracy. This operation can be considered a representative expression, an AF extraction process to obtain representative AFs. This can also reduce the computation complexity of the proposed algorithm.

The AF cube of the time series is divided into a short-term sequence of feature intervals every five frames to reflect feature information and anomaly shape structure rapidly. Each frame has an MSI AF cube $f(\varpi)$, where ϖ is the frame, and the *i*th short-term sequence feature interval is defined by

$$AF^i = \{f(5i+1), f(5i+2), \ldots f(5i+5)\} \quad i=0, 1,2 \ldots (17)$$

where *i* represents the sequence number of a short time-series feature interval.

All spectral signals can be composed by a set of vector bases in a short time-series AF cube to calculate the spectral descriptor between different AFs (AF^i). The base vector set is obtained by sparse coding and can be defined as a

dictionary U. The correlation between AF^i can be calculated by finding the inner product of two short time series to reflect the similarity between two changes in the facial structure. If the similarity is large, the short time-series data can reflect the expression feature. The inner product of the two short time-series data AF^i and AF^{i+1} can be obtained by

$$\langle AF^i, AF^{i+1}\rangle = \sum_{\forall m,n} \int_{5i+1}^{5i+10} \varphi_{im} U_i U_{i+1} \varphi_{(i+1)n}\, dt \quad (18)$$

where φ_{im} and $\varphi_{(i+1)n}$ are coefficient vectors, and the AF descriptor can be spanned by a set of bases (dictionary U).

Thus, the inner product of the AF descriptor can be found. The correlation of the short time-series feature intervals is employed to solve a base vector set U and guarantee their independence. The AF training set is obtained as follows: $X = [x_1, x_2, .., x_N]$. Dictionary learning is iteratively performed under sparse constraints, and the sparse representation vector set of image elements is solved as $A = [a_1, a_2, .., a_N]$. Finally, the preliminary model is defined as follows

$$arg\ min \frac{1}{2} \|X - UA\|_F^2 + \varphi \sum_{i=1:N} \|a_i\|_1 \quad (19)$$

Since an anomaly is closely related to the facial structure and the spectral correlation changes, spectral correlation is added to the model, and a local spectral correlation constraint is introduced. The sparse coding process is sparse, and similar coding methods can be used for local expression features with strong spectral correlation. Therefore, the proposed model can be rewritten as follows

$$arg\ min \frac{1}{2} \|X - UA\|_F^2 + \varphi \sum_{i=1:N} \|\Upsilon_i \odot a_i\|_1 \quad (20)$$

where U is the dictionary, $u_i \in U$ represents the words, $\odot$ refers to the multiplication of elements, and $\Upsilon_i \in R^{k\times l}$ is a local adapter; Υ_i is defined as $\Upsilon_{ik} = SD(x_i, u_i)$.

In this study, the SD is used to define the correlation distance, that is, the spectral correlation constraint is introduced. Next, $\frac{1}{2}\|X - UA\|_F^2$ limits the reconstruction error, and $\sum_{i=1:N} \|\Upsilon_i \odot a_i\|_1$ guarantees that similar coding methods can be used for features with similar spectral relationships. These methods ensure spectral correlation and reflect changes in the facial structure and features caused by changes in expressions.

When the dictionary is obtained, the correlation between intervals can be calculated by the inner product as follows:

$$\langle AF^i, AF^{i+1}\rangle = \varphi_i^{\mathrm{T}} \Phi(\varpi) \varphi_{i+1} \quad (21)$$

$$\Phi(\varpi) = \langle U^i, U^{i+1}\rangle \quad (22)$$

The distance between every two short time-series datasets can be obtained by calculating the inner product between the corresponding AF descriptors. Therefore, a set of short-term sequence characteristic intervals with strong correlation is obtained, and the extracted data are defined as a representative AF. When the required representative AF is obtained, it is defined as a spectra-anomalous spatiotemporal representative feature (SSRF) per the AF nature. Then, the representative AF is fed to the LSTM model for FER. The LSTM represents an advanced variant model of recurrent neural networks (RNNs). Namely, it inherits the characteristics of most RNN models and improves the long-term dependence problem and vanishing gradient. The LSTM, as a nonlinear model, can be used as a complex nonlinear element to construct neural networks with a deeper depth. Moreover, the LSTM is suitable for solving problems related to time-series data. Although a feedforward network, such as a convolutional neural network (CNN), can have performance advantages over the LSTM, the potential of the LSTM in long-term and highly complex classification tasks is better than that of CNNs. In addition, the LSTM model can more closely represent and simulate the cognitive process of human behavior, logical development, and neural organization than the CNN models. Therefore, this study selects the LSTM model to complete the FER and classification tasks.

C. Band Selection

The band parameter selection process represents a type of feature selection. The most distinguishable spectral bands can be enhanced with feature selection. In general, the spectral response features of objects can distinguish certain categories of objects easily. The popular band selection methods include joint entropy (JE) and the optimal index factor (OIF). However, the space-time complexity of the calculation of these methods is relatively high, which limits their practical applications.

Considering the aforementioned limitations, Liu [111] proposed an adaptive band selection method after fully studying the OIF and other methods. Both spatial and spectral correlations of each band are fully considered in this method. This study selects this model for band selection, which can be expressed as follows:

$$I_j = S_i / \left[(R_{i-1,i} + R_{i,j+1})/2\right] \quad (23)$$

where S_i is the standard deviation of the ith band; $R_{i-1,i}$ and $R_{i,i+1}$ represent correlation coefficients of the ith band and its neighboring bands or the correlation coefficients between the ith band and any two bands; I_j is the exponential size of the ith image.

When $R_{i-1,i}$ and $R_{i,i+1}$ are set to the correlation coefficient of the ith band and its neighboring bands, the index value of the three-band combination is obtained. When $R_{i-1,i}$ and $R_{i,i+1}$ are used as correlation coefficients of the ith band and any two bands, the index value of a single band is obtained. Therefore, this method can obtain the combination results of three bands and be extended to any N-band scenario as needed.

III. Experimental Results

The AF-MSI algorithm was applied to the experimental data to verify its ability to identify facial expressions. The advantages of the SSRF extracted by the recognition algorithm were also analyzed.

A. Datasets

In this study, 500 healthy volunteers were recruited through online forums and newspapers. Their ages ranged from 18 to 76 years old, with an average value of 37.2 years and a standard deviation of 17.2. The volunteers' ethnicities included Chinese, Indian, Caucasian, and African. All the volunteers involved signed their consent for participation in the experiment.

A multispectral imager covering the visible and near-infrared bands within 450 nm–800 nm was used in the experiment. The system consisted of a Tamron lens, a Brimrose acoustic optic tunable filter imaging spectrograph,

a fast-imaging camera, and a computer. The image size could be adjusted according to a specific requirement for the frame rate. Not all band data were required, and the most stable band sensitive to emotion was selected as the band of interest. Prior to selecting the band of interest, facial expression signals were collected in different bands in the ranges of 450 nm–500 nm, 500 nm–550 nm, 550 nm–600 nm, 600 nm–650 nm, 650 nm–700 nm, 700 nm–750 nm, and 750 nm–800 nm. The spectral acquisition interval was set to 5 nm. Data were collected separately for each band range, and data for 11 bands were obtained. All obtained MSI data cubes denoted temporal data due to the MSI's mature rapid imaging technology.

The volunteers were invited to the laboratory where the experiment was conducted. First, the volunteers were given five minutes to adapt to the laboratory's internal environment. Then, they were asked to sit in a well-illuminated room with the MSI system placed in front of them. Afterward, film clips containing emotional scenes with audio exposed subjects to real-life scenarios and elicited subjective and physiological changes were used in the experiment. The participants were exposed to emotional films to elicit their subjective and physiological changes. The experiment involved seven categories of facial expressions elicited by the exposure. At the start, the volunteers showed seven main expressions: happiness, anger, sadness, disgust, surprise, fear, and neutrality.

The MSI system recorded the entire sequence of volunteers' expressions in real-time for more than 25 min. A sufficient number of movies were prepared for each participant to elicit his/her full expressions. Each expression was monitored for three to four minutes by the MSI system, and the corresponding data were collected. The movies were played continuously and simultaneously with the imaging system. When a movie was played, the software controller recorded the facial expression data. At the same time, the imaging control software kept track of the time corresponding to the facial expression data and saved these data as an attachment to the MSI data. Thus, all the MSI data were separated. The frame rate of the imaging system was set to 30 Hz, and more than 45,000 images were captured for each participant. All datasets were processed independently; 70% of the collected image sequences were used for training, and the remaining 30% were used for testing. The experimental environment is presented in Fig. 4, and the same trial setup was employed in the previous emotion recognition study conducted by the authors [29].

B. Implementation Configuration

The collected data were purposely expanded to prevent overfitting using different techniques, including horizontal flipping of sequence frames, rotating images at specific angles, and scaling the frames with various scaling factors. In this study, the initial learning rate of the LSTM was set to 0.07 for the multispectral image dataset, and the maximum number of training epochs was set to 50. The training of the model took approximately 110 min on a NVIDIA Tesla K40 GPU. All datasets were processed independently; 70% of the collected image sequences were used for training, and the remaining 30% were used for testing. The experimental results of the AF-MSI model are shown in Table I, where the confusion matrices of seven expression states are presented.

C. Performance Comparison with State-of-the-Art Methods

The experimental results of the proposed algorithm, presented in Table I, showed that the proposed AF-MSI model was effective. The seven expressions had an average recognition accuracy of 86%. The proposed algorithm was compared with other facial expression algorithms for objective verification. However, it should be noted that the proposed algorithm has been the first algorithm that uses a multispectral dataset and AF to recognize expressions, and most existing FER algorithms are based on visible band imaging data (RGB datasets). Therefore, the proposed algorithm could not be verified on a public dataset. Fortunately, the MSI dataset constructed in this study ranged from 450 nm to 800 nm, which included the entire visible band range. Thus, the comparison with recent methods, particularly RGB-band imaging-based dynamic image methods and learned methods, CWCST [46], Ldl-Alsg (L-A) [56], CSRL [51], It-Rbm [113], MTAC [52], Ada-CM[26], and AGRA [117] methods, was conducted on the MSI dataset. Since the proposed algorithm is based on the global appearance feature, it was compared with recent hand-crafted and learned-feature methods. The experiment included seven categories (nonbinary classification). The F1 score was selected as a comparison parameter, and the F1 score results of different algorithms are presented in Table II. Compared to the dynamic image-based and deep learning methods, the proposed algorithm performed better and obtained higher F1 scores in seven expression states (Table II); it also had an evident advantage in the average F1 score (Table II).

In addition to the experiment on the self-developed MSI dataset, the proposed AF-MSI model (using the MSI dataset) was compared with the above algorithms on the RGB datasets, including the MMI, DFEW, Oulu-CASIA, Affwild2, and AFEW datasets [118],[119]. The MMI dataset was obtained from 32 participants using designated facial expressions collected under laboratory conditions. The dataset contained 2,900 videos and 740 images with a 720 × 576 pixel resolution. The DFEW is a new large-scale database of dynamic facial expressions captured in natural environments, consisting of over 16,000 video clips from thousands of movies. The Oulu-CASIA dataset included data collected from 80 participants under different lighting conditions in a laboratory using near-infrared and visible light. The emoji tags included: happy, sad, surprised, scared, disgusted, and angry. This dataset contained 2,880 video sequences with a 320 × 240-pixel resolution. The AffWild dataset was composed of facial expression videos retrieved from YouTube. The AffWild2 dataset added 260 videos to the AffWild dataset. The content of the AFEW dataset included video clips containing expressions edited from movies. The AFEW dataset denoted a dynamic facial expression dataset, consisting of 1,749 video clips. Even though the experimental environment and datasets were different, the encouraging detection results, displayed in Tables III to VII, demonstrated that the proposed AF-MSI model's recognition rate exceeded those of the other state-of-the-art algorithms. Therefore, the experimental results proved the advantage of the proposed algorithm.

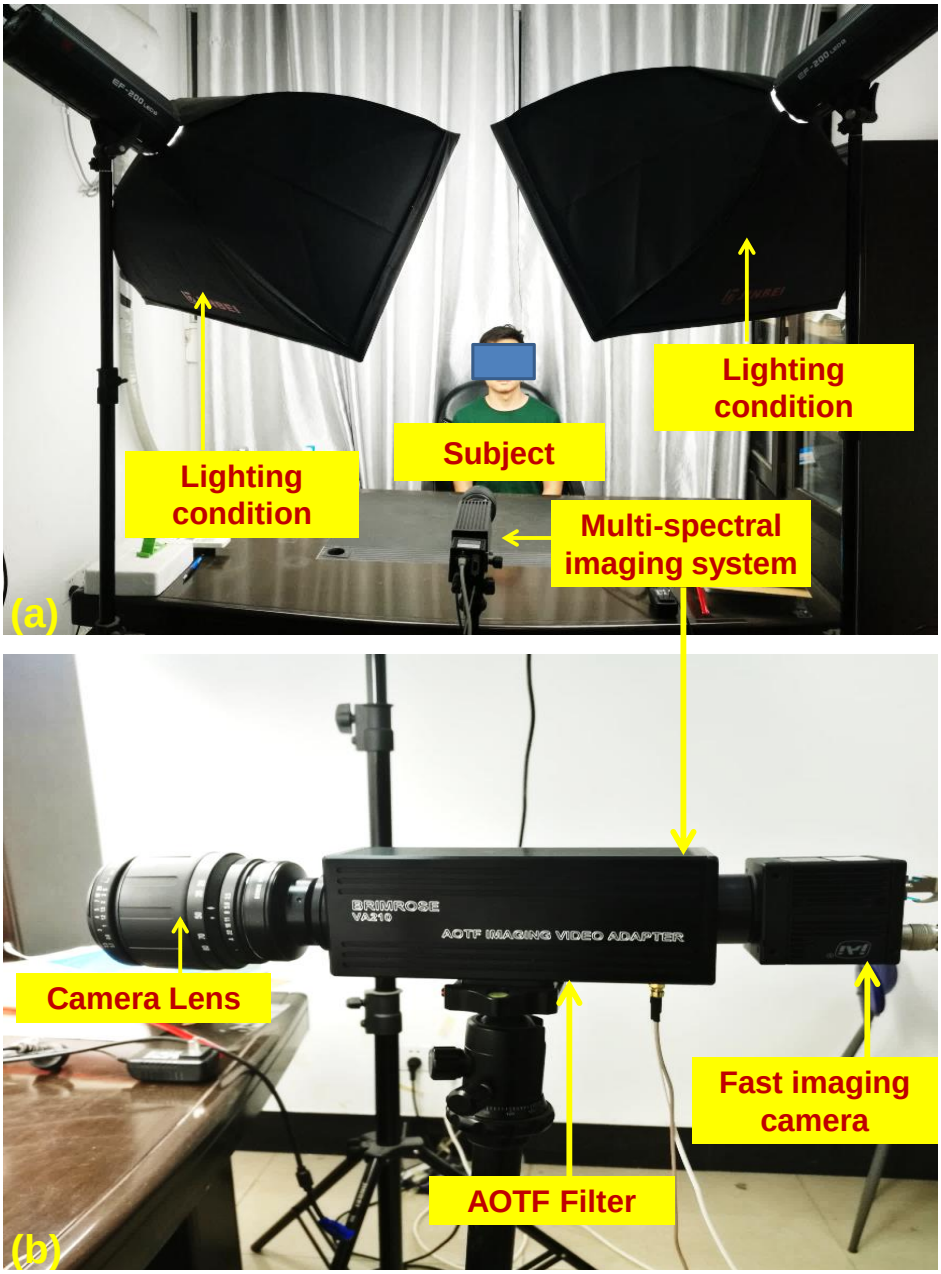

Fig. 4. (a) Experimental environment; (b) the MSI system.

TABLE I
The confusion matrices of the experimental result of the AF-MSI model. The results are expressed in percent.

	Happiness	Anger	Sadness	Disgust	Surprise	Fear	Neutral
Happiness	86	0	4.67	0	6	2	1.33
Anger	3.33	84.67	2	4	6	0	0
Sadness	4	0	86.67	0	2.67	6.67	0
Disgust	0	6	5.33	87.33	0	1.33	0
Surprise	7.33	0	3.33	4.67	83.33	0	1.33
Fear	1.33	3.33	2.67	0	0	88.67	4
Neutral	4	0	0.67	0	3.33	2.67	89.33

TABLE II
The F1 score of different algorithms for facial expression. The last row contains each algorithm's average F1 score for seven facial expressions.

	IT-RBM	CWCST	CSRL	AGRA	ADA-CM	MTAC	L-A	AFMSI
Happiness	0.721	0.708	0.756	0.788	0.705	0.727	0.797	0.835
Anger	0.688	0.673	0.783	0.773	0.792	0.735	0.784	0.872
Sadness	0.80	0.684	0.801	0.796	0.717	0.785	0.795	0.844
Disgust	0.735	0.681	0.772	0.816	0.814	0.727	0.807	0.891
Neutral	0.681	0.627	0.792	0.835	0.696	0.736	0.762	0.911
Fear	0.761	0.685	0.790	0.809	0.742	0.764	0.792	0.880
Surprise	0.785	0.689	0.773	0.808	0.795	0.787	0.785	0.827
All emotion	0.730	0.680	0.780	0.800	0.780	0.750	0.790	0.865

TABLE III
The comparison results of recognition accuracy (%) of different methods on the MMI dataset.

APPROACH	FER ACCURACY (%)
It-Rbm	82.22
CWCST	72.44
CSRL	83.57
AGRA	79.10
Ada-CM	76.69
MTAC	77.24
L-A	70.31
AFMSI (USING MSI DATA)	86.57

TABLE IV
The comparison results of recognition accuracy (%) of different methods on the DFEW dataset.

APPROACH	FER ACCURACY (%)
It-Rbm	75.99
CWCST	68.05
CSRL	64.29
AGRA	66.94
Ada-CM	70.11
MTAC	62.74
L-A	70.85
AFMSI (USING MSI DATA)	86.57

TABLE V
The comparison results of recognition accuracy (%) of different methods on the OULU-CASIA dataset.

APPROACH	FER ACCURACY (%)
It-Rbm	67.51
CWCST	63.18
CSRL	66.48
AGRA	72.73
Ada-CM	69.97
MTAC	70.89
L-A	63.94
AFMSI (USING MSI DATA)	86.57

TABLE VI
The comparison results of recognition accuracy (%) of different methods on the AFFWILD2 dataset.

APPROACH	FER ACCURACY (%)
It-Rbm	69.25
CWCST	60.79
CSRL	59.64
AGRA	64.12
Ada-CM	65.46
MTAC	64.86
L-A	58.79
AFMSI (USING MSI DATA)	86.57

TABLE VII
The comparison of recognition accuracy (%) of different methods on the AFEW dataset.

APPROACH	FER ACCURACY (%)
It-Rbm	70.95
CWCST	62.73
CSRL	56.88
AGRA	56.43
Ada-CM	51.05
MTAC	59.94
L-A	55.87
AFMSI (USING MSI DATA)	86.57

D. Uncertainty Handling Performance Evaluation

This study also analyzed the parameters of the proposed algorithm, including the influence of skin color, viewing angle, and band selection. The recognition rate for the participants with different skin colors, namely Caucasian, Black, Mongoloid, and Morena, are presented in Fig. 5. Among them, Caucasian accounted for 30%, and Black accounted for 25%; Mongoloid and Morena accounted for 25% and 20%, respectively.

The experimental results showed no significant difference in the recognition rate of seven facial expressions for different skin colors. In addition, the results of the t-test showed no statistical difference in the recognition rate for different skin colors. The null hypothesis that the experimental results for these different skin colors (accuracy rate) could be different was not supported (i.e., $p > 0.6$ in all experimental data).

In the band selection, a large index indicated that the information on the corresponding band combination image was abundant and highly representative. Given that the number of bands N was arbitrary, the number of bands could be an interval of bands. This study used the expression signals in different bands: 450 nm–500 nm, 500 nm–550 nm, 550 nm–600 nm, 600 nm–650 nm, 650 nm–700 nm, 700 nm–750 nm, and 750–800 nm. Considering the correlation of spectral and spatial signals, the band interval was directly selected instead of a single wavelength. Thus, the data source was selected from the seven band intervals, where N was set to the band interval.

The final calculations showed that the index of 550–600 nm was the largest for all band intervals. Therefore, the band signals in the band interval 550 nm–600 nm were selected as a spectral data source. After the band interval was obtained, the spectral signals with different wavelengths were used to verify the proposed AF-MSI algorithm's results, and the most sensitive interval was selected for each facial expression recognition.

The experimental results verified the advantage of the 550–600 nm band, as shown in Fig. 6. The results also showed that the maximum recognition rate was obtained in the 550–600 nm range. Thus, the band interval selection is of significant importance to the algorithm's effectiveness.

The sensitivity of the parameter angle, which represented the viewing angle between the face and the imaging system, was also analyzed. The stability of the proposed algorithm was examined for different viewing angles. The value range of the viewing angle was set as {0, 15, 30, 45, 60, 75, 90}, where the angle of zero referred to a face viewing the imaging system from the front, whereas the angles from 15° to 90° represented the viewing angle. The results are presented in Table VIII, where it can be seen that the proposed algorithm's recognition rate increased when the face and the imaging system were at an angle of less than 45°. The best observation angle was 45°, which corresponded to an accuracy of 89%. When the angle exceeded 45°, the recognition rate decreased and finally reached 83% at 90°. Therefore, the viewing angle did not affect the stability of the proposed algorithm.

E. Ablation Study

Ablation studies were performed to verify the contribution of each branch of the AF-MSI algorithm and the key hyperparameters proposed in this study. The proposed algorithm was implemented in several steps, from the initial MSI experiment to the AF extraction and the subsequent

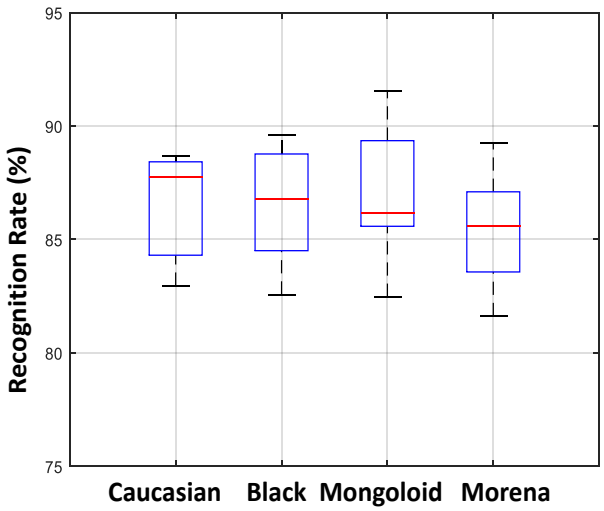

Fig. 5. The boxplot of seven FER rates of different skin colors.

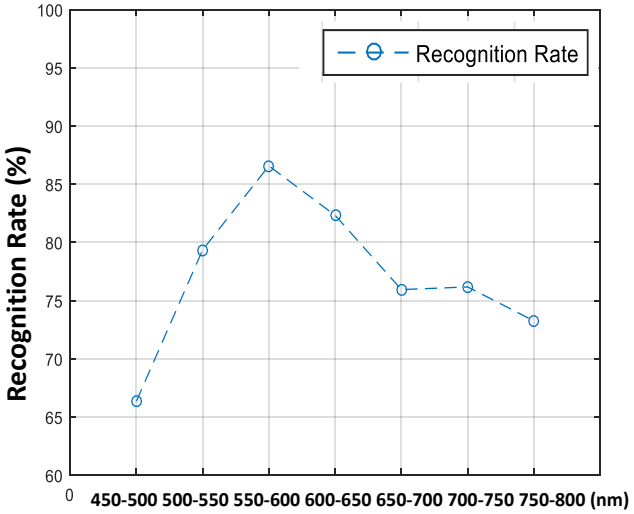

Fig. 6. The recognition rate of different band intervals. The band ranges from 450 nm to 800 nm, and every 50 nm is an interval of interest.

representative AF selection, which denoted the core parts of the proposed algorithm. Therefore, the effects of the main steps on the final recognition results were examined. The input features of the proposed FER model were divided into the following situations: MSI + MD-based AF, MSI + spectral & spatial AF, and MSI +representative spectral & spatial AF. This setting was used to demonstrate the influence of different input features on the proposed algorithm's result. Different features were input into the AF-MSI model for FER calculation, as shown in Table IX. The result indicated that when the MD algorithm was used to extract anomalies to identify facial expressions, the accuracy was 65.46%. For the MSI + spectral & spatial AF and the MSI + representative spectral & spatial AF, the algorithm's accuracy was 79% and 86%, respectively. Thus, the proposed algorithm's branches had a positive effect on the experimental results.

Next, the proposed AF-MSI's ability to handle uncertain samples was investigated. Specifically, comparative testing was conducted in synthetic uncertainty scenarios. Two basic methods, the LSTM and the DenseNet, and two state-of-the-art methods, the CWCST and the Ada CM [26][46] were used in this analysis. This study synthesized 10%, 20%, and 30% samples in the training sets of the MSI dataset with a random category other than their original labels. The experimental results, presented in Table X, showed that

TABLE VIII
Average recognition accuracy versus the viewing angle

Angle	Accuracy (%)
0	86.57
15	87.35
30	89.76
45	89.13
60	86.81
75	84.44
90	83.65

ABLE IX
The comparison results of the average accuracy of the input features calculated by the AF-MSI model

Input Feature	Accuracy (%)
MSI+MD based AF	65.17
MSI+ spectral & spatial AF	79.74
MSI + representative AF	86.57

the AF-MSI performed better than the LSTM and DenseNet methods. In addition, as the proportion of uncertainty increased, the performance of the AF-MSI decreased less compared to the other methods, indicating the effectiveness of the proposed spectral features. Due to these advantages, the proposed AF-MSI could achieve competitive performance, performing the best in all uncertain experiments.

IV. Conclusion

With the application of remote sensing anomaly target recognition, anomaly detection has been able to identify a target with a significant difference between spectral image and BG information using the MD-based principle. Unlike traditional MD-based anomaly detection methods and considering that a recognition target is to extract the feature of expression change, this study focuses on the structural changes and features of the face when extracting an anomaly. Therefore, spectral and spatial correlations between facial structures are used to extract spectral and spatial AFs. In a small BG divided by Candide3, the multivariate correlation method and LOF model are used to calculate the multiband correlation of spectral signals. Then, the spectral AF from the BG is extracted, and the MD is used to extract spatial AF. The remaining BG information is learned through DBN training to retain the spatial structure information.

This study does not aim to separate this feature from the spatial region, which can introduce signal confusion (many expressions have similar processes). Identifying the features with good similarity (especially between spectra) and extracting a representative AF before training the samples is beneficial. The time-series AFs are divided into short time-series AFs with a specific frame interval. The descriptors of short time-series AFs are constructed using the dictionary as a base vector. In addition, the inner product of AFs is calculated to determine the correlation between short time series. The extracted SSRF with high correlation is fed to the LSTM model for final recognition.

TABLE X
Encountering synthetic uncertain samples

Approach	Uncertainty	FER Accuracy (%)
LSTM		57.42
DenseNet		58.33
CWCST	10%	61.26
Ada-CM		72.94
AFMSI		82.17
LSTM		55.29
DenseNet		57.84
CWCST	20%	61.03
Ada-CM		70.98
AFMSI		81.77
LSTM		53.36
DenseNet		56.49
CWCST	30%	60.28
Ada-CM		69.09
AFMSI		80.05

The FER mostly uses methods based on hand-crafted and learned features. These methods are commonly combined in current recognition algorithms. However, the process of extracting geometric facial features and tracking changes is challenging. Therefore, this study uses the global appearance feature when extracting recognition features. The proposed method is practical, and the AF-MSI algorithm is proposed for these reasons.

The results of the FER experiment show that the AF-MSI algorithm can achieve an accuracy of 86%. The proposed method has better results than the currently used methods based on global appearance and learned features. The experimental results demonstrate that skin color does not affect the proposed algorithm's performance. Moreover, it is shown that the viewing angle does not affect the stability of the proposed algorithm. Furthermore, it is demonstrated that the main steps of the proposed algorithm have a positive effect on the final recognition result. The adaptive band selection method is employed, and the selected band interval achieves the maximum recognition rate.

To the best of the authors' knowledge, the proposed AF-MSI algorithm has been the first algorithm that employs MSI technology to recognize facial expressions and the anomaly detection methods to extract the FER features. Positive and applicable results are obtained by extracting spatial-spectral representative anomaly-temporal features. This study introduces a useful and efficient method for FER.

The proposed approach can achieve good results, but additional research should be conducted. Future work could expand the scope of the experiments and consider additional practical scenarios. For instance, the imaging system could be placed in a supermarket, and the proposed algorithm could be used to identify the purchase intention and expectations of customers. In addition, real-time emotional monitoring could also be beneficial to health management. Further, the proposed algorithm could be applied to the security field because subtle expression signals can provide useful screening information for security purposes. Finally, future work could further miniaturize and customize the imaging equipment for the application scenarios.

Furthermore, the proposed algorithm could be embedded into an imaging system for real application. In this way, the whole system would have a smaller size and a lower cost. The accuracy of the proposed algorithm could be further improved for specific applications. In particular, the correlation between the spectral signal and the specific facial structure requires further quantification. Although future work plans are relatively challenging, the proposed method holds considerable promise for practical applications.

REFERENCES

[1] G. Littlewort, et al., "The computer expression recognition toolbox (CERT)," in Proc. IEEE Int. Conf. Workshops Automatic Face Gesture Recognition., 2011, pp. 298–305.

[2] K. Anderson and P. W. Mcowan, "A real-time automated system for the recognition of human facial expressions," IEEE Trans. Syst. Man Cybern. Part B: Cybern., vol. 36, no. 1, pp. 96–105, Feb. 2006.

[3] S. h. lee, "Collaborative expression representation using peak expression and intra class variation face images for practical subject-independent emotion recognition in videos," Pattern Recogn., vol. 54, pp. 52–67, 2016.

[4] Ciprian adrian corneanu., et al.: Survey on RGB, 3D, Thermal, and Multimodal Approaches for Facial Expression Recognition: History, Trends, and Affect-Related Applications. IEEE Transactions on pattern analysis and machine intelligence. VOL38(8), 121-141 (2016)

[5] Siyue Xie, et al.: Facial Expression Recognition with Two-branch Disentangled Generative Adversarial Network. IEEE Transactions on Circuits and Systems for Video Technology. (Early Access 2020)

[6] Bilal taha, et al.: Learned 3D Shape Representations Using Fused Geometrically Augmented Images: Application to Facial Expression and Action Unit Detection. IEEE Transactions on Circuits and Systems for Video Technology. Vol30(9), 2900 - 2916 (2020)

[7] Ercheng Pei, et al.: Monocular 3D Facial Expression Features for Continuous Affect Recognition. IEEE Transactions on Multimedia. (Early Access, 2020)

[8] W. H. Al-arashi, H. Ibrahim, and S. A. suandi, "Optimizing principal component analysis performance for face recognition using genetic algorithm," Neurocomputing, vol. 128, pp. 415–420, Mar. 2014.

[9] M.-W. Huang, Z.-W. Wang, and Z.-L. Ying, "A new method for facial expression recognition based on sparse representation plus LBP," in Proc. Int. Congr. Image Signal Process., 2010, pp. 1750–1754.

[10] Z. Wang, Q. Q. Ruan, and G. Y. An, "Facial expression recognition using sparse local fisher discriminant analysis," Neurocomputing, vol. 174, pp. 756–766, Jan. 2016.

[11] Xiaowei Zhang, et al.: Emotion Recognition from Multimodal Physiological Signals Using a Regularized Deep Fusion of Kernel Machine. IEEE Transactions on Cybernetics (Early Access), pp:1 – 14 (2020)

[12] M. K. Bhowmik, et al.: Enhancement of robustness of face recognition system through reduced Gaussianity in log-ICA. Expert Syst. Appl. Vol 116(1), 96-107(2019)

[13] I. Kotsia and I. Pitas, "Facial expression recognition in image sequences using geometric deformation features and support vector machines," IEEE Trans. Image Process., vol. 16, no. 1, pp. 172–187, Jan. 2007.

[14] Y. Chang, M. Vieira, M. Turk, and L. Velho, "Automatic 3D facial expression analysis in videos," in Proc. 2nd Int. Conf. Anal. Model. Faces Gestures, 2005, pp. 293–307.

[15] Meng-Ju Han, et al.: Robotic Emotional Expression Generation Based on Mood Transition and Personality Model. IEEE Transactions on Cybernetics. Vol43(4) , 1290 – 1303 (2013)

[16] Kamlesh Mistry, et al.: A micro-ga embedded PSO feature selection approach to intelligent facial emotion recognition. IEEE Transactions on Cybernetics. Vol47(6), 1496 - 1509 (2017)

[17] T. Ahonen, A. Hadid, and M. Pietikainen, "Face description with local binary patterns: Application to face recognition," IEEE transactions on pattern analysis and machine intelligence. vol28(12), 2037–2041(2006)

[18] Weicheng Xie, et al.: Adaptive Weighting of Handcrafted Feature Losses for Facial Expression Recognition. IEEE Transactions on Cybernetics. 1-14 (2019 Early Access)

[19] Lin Zhong, et al.: Learning Multiscale Active Facial Patches for Expression Analysis. IEEE Transactions on Cybernetics. Vol45(8), 1499 – 1510 (2015)

[20] G. Donato, M. S. Bartlett, J. C. Hager, P. Ekman, and T. J. Sejnowski, "Classifying facial actions," IEEE transactions on pattern analysis and machine intelligence., vol. 21, no. 10, pp. 974–989, Oct. 1999.

[21] I. Kotsia and I. Pitas, "Facial expression recognition in image sequences using geometric deformation features and support vector machines," IEEE Trans. Image Process., vol. 16, no. 1, pp. 172–187, Jan. 2007.

[22] S. Koelstra, M. Pantic, and I. Patras, "A dynamic texture-based approach to recognition of facial actions and their temporal models," IEEE transactions on pattern analysis and machine intelligence., vol. 32, no. 11, pp. 1940–1954, Nov. 2010.

[23] I. Cohen, N. Sebe, F. G. Gozman, M. C. Cirelo, and T. S. Huang, "Learning Bayesian network classifiers for facial expression recognition both labeled and unlabeled data," in Proc. IEEE Conf. Comput. Vision Pattern Recog., 2003, pp. I-595–I-601.

[24] P. S. Aleksic and A. K. Katsaggelos, "Automatic facial expression recognition using facial animation parameters and multistream HMMs," IEEE Trans. Inf. Forensics Security, vol. 1, no. 1, pp. 3–11, Mar. 2006.

[25] Dan Zeng, et al.: Face2Exp: Combating Data Biases for Facial Expression Recognition. In the proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR2022).

[26] Hangyu Li, et al.: Towards Semi-Supervised Deep Facial Expression Recognition with An Adaptive Confidence Margin. In the proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR2022)

[27] Shasha Wang.: A Face Recognition Method based on Lightweight Neural Network and Multi Hash Recognition Degree Weighting. IAENG International Journal of Applied Mathematics. Vol54(3), Pages 581-586 (2024)

[28] D. Smeets, P. Claes, J. Hermans, D. Vandermeulen, and P. Suetens, "A comparative study of 3-D face recognition under expression variations," IEEE Trans. Syst., Man, Cybern. C, Appl. Rev., vol. 42, no. 5, pp. 710–727, Sep. 2012.

[29] Kan Hong*.: Spatial-spectral-temporal framework for emotion recognition. IEEE Access. Vol8(1), 1303 - 1315 (2020)

[30] J. Wang, D. Yang, W. Jiang, and J. Zhou, "Semisupervised incremental support vector machine learning based on neighborhood kernel estimation," IEEE Trans. Syst., Man, Cybern., Syst., vol. 47, no. 10, pp. 2677–2687, Oct. 2017.

[31] W. Yang, X. Sun, and Q. Liao, "Cascaded elastically progressive model for accurate face alignment," IEEE Trans. Syst., Man, Cybern., Syst., vol. 47, no. 9, pp. 2613–2621, Sep. 2017.

[32] M. Lyons, S. Akamatsu, M. Kamachi, and J. Gyoba, "Coding facial expressions with gabor wavelets," in Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition, 1998, pp. 200–205.

[33] S. Jaiswal and M. F. Valstar, "Deep learning the dynamic appearance and shape of facial action units," in Proceedings of the Winter Conference on Applications of Computer Vision (WACV), 2016.

[34] S. Du, Y. Tao, and A. M. Martinez, "Compound facial expressions of emotion," Proceedings of the National Academy of Sciences, vol.111, no. 15, pp. E1454–E1462, 2014.

[35] Y. Tong, W. Liao, and Q. Ji, "Facial action unit recognition by exploiting their dynamic and semantic relationships," IEEE Transactions on Pattern Analysis and Machine Intelligence, no. 10, pp. 1683– 1699, 2007.

[36] Xi Zhao，et al.: Automatic 2.5-D Facial Landmarking and Emotion Annotation for Social Interaction Assistance. IEEE Transactions on Cybernetics. Vol46(9), 2042 – 2055 (2016)

[37] Cunling Bian, et al.: Spontaneous facial expression database for academic emotion inference in online learning. IET Computer Vision. Volume 13, Issue 3 p. 329-337 (2019)

[38] G. Zhao and M. Pietikainen, "Dynamic texture recognition using local binary patterns with an application to facial expressions," IEEE Trans. Pattern Anal. Mach. Intell., vol. 29, no. 6, pp. 915–928, Jun. 2007.

[39] B. Jiang, M. Valstar, B. Martinez, and M. Pantic, "A dynamic appearance descriptor approach to facial actions temporal modeling," IEEE Transactions on Cybernetics., vol. 44, no. 2, pp. 161–174, Feb. 2014.

[40] Jingting Li., et al.: CAS(ME)3: A third generation facial spontaneous micro-expression database with depth Information and High Ecological Validity. IEEE transactions on pattern analysis and machine intelligence., Vol45(3), 120-131(2023)

[41] Xianye Ben, et al.: Video-based facial micro-expression analysis: A survey of datasets, features and algorithms. IEEE transactions on pattern analysis and machine intelligence. Vol44(9), 121-133 (2022)

[42] M. Pantic and I. Patras, "Dynamics of facial expression: Recognition of facial actions and their temporal segments from face profile image sequences," IEEE Trans. Syst., Man., Cybern. B., Cybern., vol. 36, no. 2, pp. 433–449, Apr. 2006.

[43] C. Shan, S. Gong, and P. W. Mcowan, "Facial expression recognition based on local binary patterns: A comprehensive study," Image Vision Comput., vol. 27, no. 6, pp. 803–816, 2009.

[44] M. J. Lyons, J. Budynek, and S. Akamatsu, "Automatic classification of single facial images," IEEE transactions on pattern analysis and machine intelligence., vol. 21, no. 12, pp. 1357–1362, Dec. 1999.

[45] Qian Dong., et al.: Multi-Scale Attention Learning Network for Facial Expression Recognition. IEEE Signal Processing Letters. Vol 30(1), 1732-1736 (2023)

[46] Yingjian Li., et al.: Cross-Domain Facial Expression Recognition via Contrastive Warm up and Complexity-Aware Self-Training. IEEE Transactions on Image Processing. Vol 32(1), 5438 - 5450 (2023)

[47] Chunlei Li., et al: FG-AGR: Fine-Grained Associative Graph Representation for Facial Expression Recognition in the Wild. IEEE Transactions on Circuits and Systems for Video Technology. Vol 34(2), 882 - 896 (2023)

[48] Ling Lo., et al.: Modeling Uncertainty for Low-Resolution Facial Expression Recognition. IEEE Transactions on Affective Computing. Vol 15(1), 198 - 209 (2024)

[49] Rongkang Dong., et al: Bi-Center Loss for Compound Facial Expression Recognition. IEEE Signal Processing Letters. Vol 31(1), 641 - 645 (2024)

[50] Yi Zhang., et al.: Facial Prior Guided Micro-Expression Generation. IEEE Transactions on Image Processing. Vol 33(1), 525 - 540 (2023)

[51] Dongliang Chen., et al.: Cross-Domain Sample Relationship Learning for Facial Expression Recognition. IEEE Transactions on Multimedia. Vol 26(1), 3788 - 3798 (2023)

[52] Yang Liu., et al.: Uncertain Facial Expression Recognition via Multi-Task Assisted Correction. IEEE Transactions on Multimedia. Vol 26(1), 2531 - 2543 (2023)

[53] Wei Zhang., et al.: Detecting Facial Action Units From Global-Local Fine-Grained Expressions. IEEE Transactions on Circuits and Systems for Video Technology. Vol 34(2), 983 - 994 (2024)

[54] Shangfei Wang., et al.: Pose-Aware Facial Expression Recognition Assisted by Expression Descriptions. IEEE Transactions on Affective Computing. Vol 15(1), 241 - 253 (2024)

[55] H. Jung, S. Lee, J. Yim, S. Park, and J. Kim, "Joint fine-tuning in deep neural networks for facial expression recognition," in Proc. IEEE Int. Conf. Comput. Vis., 2015, pp. 2983–2991.

[56] Shikai Chen, et al.: Label Distribution Learning on Auxiliary Label Space Graphs for Facial Expression Recognition. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 1275-1278 (2020)

[57] Anima Majumder , et al.: Automatic Facial Expression Recognition System Using Deep Network-Based Data Fusion. IEEE Transactions on Cybernetics. Vol48(1), 103-114 (2018)

[58] N. Fragopanagos and J. G. Taylor, "Emotion recognition in human–computer interaction," Neural Netw., vol. 18, no. 4, pp. 389–405, 2005.

[59] Tong Zhang, et al.: Spatial–Temporal Recurrent Neural Network for Emotion Recognition. IEEE Transactions on Cybernetics. Vol49(3), 839 - 847 (2019)

[60] Wenrui Shen., et al.: Person Re-Identification Algorithm Based on Improved ResNet. IAENG International Journal of Applied Mathematics. Vol54(5), 894-901 (2024)

[61] M. Wcollmer, M. Kaiser, F. Eyben, B. Schuller, and G. Rigoll, "LSTM-modeling of continuous emotions in an audiovisual affect recognition framework," Image Vision Comput., vol. 31, no. 2, pp. 153–163, 2013.

[62] I. Cohen, N. Sebe, L. Chen, A. Garg, and T. S. Huang, "Facial expression recognition from video sequences: Temporal and static modelling," Comput. Vision Image Understanding, vol. 91, pp. 160–187, 2003.

[63] P. S. Aleksic and A. K. Katsaggelos, "Automatic facial expression recognition using facial animation parameters and multistream HMMs," IEEE Trans. Inf. Forensics Security, vol. 1, no. 1, pp. 3–11, Mar. 2006.

[64] C. Wu, S. Wang, and Q. Ji, "Multi-instance hidden Markov model for facial expression recognition," in Proc. IEEE Int. Conf. Autom. Face Gesture Recog., 2015, pp. 1–6.

[65] R. Walecki, O. Rudovic, V. Pavlovic, and M. Pantic, "Variable state latent conditional random fields for facial expression recognition and action unit detection," in Proc. IEEE Int. Conf. Autom. Face Gesture Recog., 2015, pp. 1–8.

[66] M. Ranzato, J. Susskind, V. Mnih, and G. Hinton, "On deep generative models with applications to recognition," in Proc. IEEE Conf. Comput. Vision Pattern Recog., 2011, pp. 2857–2864.

[67] M. Liu, S. Li, S. Shan, and X. Chen, "AU-aware deep networks for facial expression recognition," in Proc. IEEE Int. Conf. Autom. Face Gesture Recog., 2013, pp. 1–6.

[68] Kaibiao Lin., et al.: Multi-View Block Matrix-Based Graph Convolutional Network. Engineering Letters. Vol32(6), 1073-1082 (2024)

[69] Hongying Meng, et al.: Time-Delay Neural Network for Continuous Emotional Dimension Prediction from Facial Expression Sequences. IEEE Transactions on Cybernetics. Vol46(4), 916 – 929 (2016)

[70] Felipe Osorio-Arteaga., et al.: Adaptive Neural Network Identification for Robust Multivariable Systems. IAENG International Journal of Applied Mathematics. Vol54(1), 68-76 (2024)

[71] P. Liu, S. Han, Z. Meng, and Y. Tong, "Facial expression recognition via a boosted deep belief network," in Proc. IEEE Conf. Comput. Vision Pattern Recog., 2014, pp. 1805–1812.

[72] Y.-L. Tian, T. Kanade, and J. F. Cohn, "Recognizing action units for facial expression analysis," IEEE Trans. Pattern Anal. Mach. Intell., vol. 23, no. 2, pp. 97–115, Feb. 2001.

[73] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, "Active shape models—Their training and application," Comput. Vision Image Understanding, vol. 61, no. 1, pp. 38–59, 1995.

[74] G. Littlewort, M. S. Bartlett, I. Fasel, J. Susskind, and J. Movellan, "Dynamics of facial expression extracted automatically from video," in Proc. IEEE Conf. Comput. Vision Pattern Recog. Workshops, 2004, p. 80.

[75] I. Kotsia and I. Pitas, "Facial expression recognition in image sequences using geometric deformation features and support vector machines," IEEE Trans. Image Process., vol. 16, no. 1, pp. 172–187, Jan. 2007.

[76] S. Berretti, B. B. Amor, M. Daoudi, and A. Del Bimbo, "3D facial expression recognition using sift descriptors of automatically detected keypoints," Visual Comput.: Int. J. Comput. Graphics, vol. 27, no. 11, pp. 1021–1036, 2011.

[77] P. Lemaire, M. Ardabilian, L. Chen, and M. Daoudi, "Fully automatic 3D facial expression recognition using differential mean curvature maps and histograms of oriented gradients," in Proc. IEEE Int. Conf. Autom. Face Gesture Recog., 2013, pp. 1–7.

[78] A. Dapogny, K. Bailly, and S. Dubuisson, "Dynamic facial expression recognition by joint static and multi-time gap transition classification," in Proc. IEEE Int. Conf. Autom. Face Gesture Recog., 2015, pp. 1–6.

[79] W. Gu, C. Xiang, Y. Venkatesh, D. Huang, and H. Lin, "Facial expression recognition using radial encoding of local Gabor features and classifier synthesis," Pattern recognition., vol. 45, no. 1, pp. 80–91, 2012.

[80] B. Hern_andez, G. Olague, R. Hammoud, L. Trujillo, and E. Romero, "Visual learning of texture descriptors for facial expression recognition in thermal imagery," Comput. Vision Image Understanding, vol. 106, no. 2, pp. 258–269, 2007.

[81] Naima Otberdout, et al.: Dynamic Facial Expression Generation on Hilbert Hypersphere with Conditional Wasserstein Generative Adversarial Nets. IEEE Transactions on pattern analysis and machine intelligence. Vol44(2), 356-368 (2022)

[82] Y. Yoshitomi, N. Miyawaki, S. Tomita, and S. Kimura, "Facial expression recognition using thermal image processing and neural network," in Proc. 6th IEEE Int. Workshop Robot Human Commun., 1997, pp. 380–385.

[83] I. Costea, et al.: Analysis of facial expressions using thermal imaging. 2016 39th International Spring Seminar on Electronics Technology (ISSE), 2016

[84] P. Liu and L. Yin, "Spontaneous facial expression analysis based on temperature changes and head motions," in Proc. IEEE Int.Conf. Autom. Face Gesture Recog., 2015, pp. 1–6.

[85] Shangfei Wang, et al.: Thermal Augmented Expression Recognition. IEEE Transactions on Cybernetics. Vol48(7), 2203 – 2214 (2018)

[86] Zenith Purisha.: Computed Tomography Reconstruction from Undersampled Data: An Application to Biomedical Imaging. IAENG International Journal of Applied Mathematics. Vol54(1), 25-32 (2024)

[87] I. Pavlidis , et al. ,Human behavior: seeing through the face of deception, Nature 415 (6867) (2002) 35–36 .

[88] I. Pavlidis , J. Dowdall , et al. , Interacting with human physiology, Comput. Vision Image Understanding 108 (2007) 150–170 .

[89] I. Pavlidis , et al. , Fast by nature—how stress patterns define human experience and performance in dexterous tasks, Sci. Rep. 2 (2012) 1–9 .

[90] Kan Hong*, Guodong Liu, Wentao Chen, S Hong.: Classification of the emotional stress and physical stress using signal magnification and canonical correlation analysis. Pattern Recognition. Vol1(77), 140-149 (2018)

[91] Kan Hong*.: Classification of emotional stress and physical stress using facial imaging features. Journal of optical technology. Vol83(8), 508-517 (2016)

[92] Kan Hong., et al.: Real-time stress assessment using thermal imaging. The Visual Computer. Vol32(11), 1369–1377 (2015)

[93] G. Shankar., et al.: A Numerical Investigation of Thermal and Mass Exchange of Blood Along Porous Stenosis Arterial Flow with Applied Magnetic Field. IAENG International Journal of Applied Mathematics. Vol54(3), 532-541(2024)

[94] N. Noora., et al.: Detecting field cancerization using a hypersepctral imaging system. Lasers in surgery and medicine. Vol45, 410–417 (2013)

[95] Robert Pike., et al.: A Minimum Spanning Forest-Based Method for Noninvasive Cancer Detection With Hyperspectral Imaging. IEEE Transactions on Biomedical Engineering. Vol63(3), 653 – 663 (2016)

[96] Jie Lei., et al.: Spectral–Spatial Feature Extraction for Hyperspectral Anomaly Detection. IEEE Transactions on Geoscience and Remote Sensing. Vol57(10), 8131 – 8143 (2019)

[97] Himar Fabelo., et al.: In-Vivo Hyperspectral Human Brain Image Database for Brain Cancer Detection. IEEE Access. Vol7(1), 39098 – 39116 (2019)

[98] Nanjun He., et al.: Feature Extraction With Multiscale Covariance Maps for Hyperspectral Image Classification. IEEE Transactions on Geoscience and Remote Sensing. Vol57(2), 755 – 769 (2018)

[99] Xuelong Li, Yue Yuan, Qi Wang.: Hyperspectral and Multispectral Image Fusion via Nonlocal Low-Rank Tensor Approximation and Sparse Representation. IEEE Transactions on Geoscience and Remote Sensing. 1 - 13 (2020 Early Access)

[100] Xuelong Li, Yue Yuan, Qi Wang.: Hyperspectral and Multispectral Image Fusion Based on Band Simulation. IEEE Geoscience and Remote Sensing Letters. Vol17(3), 479 – 483 (2020)

[101] J. Ahlberg, candide-3 -- an updated parameterized face, Report No. Lith-isy-r-2326, Dept. of Electrical Engineering, Linköping University, Sweden, 2001.

[102] T. Chen , Peter Yuen , Kan Hong., et al.: Remote sensing of stress using Electro-optics imaging technique. Proceedings of the SPIE, Vol7486, 601-612 (2009)

[103] T. Chen, et al.: "Detection of Psychological Stress Using a Hyperspectral Imaging Technique". IEEE transactions on affective computing. Vol5(4), pp. 391-405, (2014)

[104] Kan Hong*, et al.: Detection of Physical Stress using Multispectral Imaging. Neurocomputing. Vol239, 116-128 (2019)

[105] I. S. Reed and X. Yu, "Adaptive multiple-band CFAR detection of an optical pattern with unknown spectral distribution," IEEE Transactions on Acoustics, Speech and Signal Processing, vol.38, no. 10, pp. 1760 – 1770, 1990.

[106] S. Khazai, S. Homayouni, A. Safari, and B. Mojaradi, "Anomaly Detection in Hyperspectral Images Based on an Adaptive Support Vector Method," IEEE Geoscience and Remote Sensing Letters, vol. 8, no. 4, pp. 646-650, 2011.

[107] Kan Hong, et al.: Facial expression recognition based on anomaly feature. Optical review. Vol29(1), 178–187 (2022)

[108] M. M. Breunig, H. P. Kriegel, R. T. Ng, J. Sander. LOF: Identifying Density-based Local Outliers. Sigmod, 2000.

[109] M. Muller, G. Baier, C. Rummel, K. Schindler, and U. Stephani, "A multivariate approach to correlation analysis based on random matrix theory," in Seizure Prediction in Epilepsy: From Basic Mechanisms to Clinical Applications, B. Schelter, J. Timmer, and A. S. Bonhage, Eds. New York, USA: Wiley, 2008, pp. 209–226.

[110] Chacvez P S, Berlin G L, Sowers L B.: Statistical method for selecting landsat MSS ratios. Journal of Applied Photographic Engineering, 1982, Vol1(8): 23- 30.

[111] Chunhong Liu, et al. A New Method of Hyperspectral Remote Sensing Image Dimensional Reduction. Journal of Image and Graphics, 2005, 10 (2): 218- 222.

[112] D. H. Kim, W. Baddar, J. Jang, and Y. M. Ro, "Multi-objective based spatio-temporal feature representation learning robust to expression intensity variations for facial expression recognition," IEEE Transactions on Affective Computing, vol. 10, no. 2, pp. 223–236, 2019.

[113] S. Wang, Z. Zheng, S. Yin, J. Yang, and Q. Ji, "A novel dynamic model capturing spatial and temporal patterns for facial expression analysis," IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 1–1, 2019.

[114] S. Kumawat, M. Verma, and S. Raman, "Lbvcnn: Local binary volume convolutional neural network for facial expression recognition from image sequences," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2019, pp. 0–0.

[115] K. Zhang, Y. Huang, Y. Du, and L.Wang, "Facial expression recognition based on deep evolutional spatial-temporal networks," IEEE Transactions on Image Processing, vol. 26, no. 9, pp. 4193–4203, 2017.

[116] Yang Liu., et al.: Facial Expression Recognition via Deep Action Units Graph Network Based on Psychological Mechanism. IEEE Transactions on cognitive and developmental systems. VOL12(2), 311-322 (2020)
[117] Tianshui Chen., et al.: Cross-Domain Facial Expression Recognition: A Unified Evaluation Benchmark and Adversarial Graph Learning. IEEE Transactions on Pattern Analysis and Machine Intelligence. Vol44(1), 9887 - 9903 (2022)
[118] Mohan Karnati., et al.: Understanding Deep Learning Techniques for Recognition of Human Emotions Using Facial Expressions: A Comprehensive Survey. IEEE Transactions on Instrumentation and Measurement. Vol72(1), 965-976 (2023)
[119] Shan Li., et al.: Deep Facial Expression Recognition: A Survey. IEEE Transactions on Affective Computing. Vol13(3), 1195 – 1215 (2022)