

Orthogonal-Moment-Based Attraction Measurement With Ocular Hints in Video-Watching Task

Minqiang Yang^{ID}, Member, IEEE, Xiang Feng^{ID}, Rong Ma^{ID}, Xiongying Li, and Chengsheng Mao^{ID}, Member, IEEE

Abstract—Pupil dilation and eye movements are closely related to human emotional and cognitive processes. Visual stimulus, especially video clips, is widely used in computer-assisted experimental paradigms as emotional inducers. However, the level of attraction as a critical factor to such visual stimulus still needs comprehensive investigation. This article conducts a novel study of attraction assessment with pupil diameter and eye movements. We collected high temporal resolution ocular variation data from 50 subjects while they viewed a variety of emotional video stimuli. Besides, this article proposes two orthogonal-moment-based feature extraction methods for emotion classification, i.e., Legendre moment and Krawtchouk moment. The results of experiments show that our proposed feature sets achieve better classification performance compared with conventional time- or frequency-domain feature sets. The accuracy of predicting attraction level reached 87.4% and 91.0% when new features were used alone and combined with conventional features, respectively. Compared with the conventional features with an accuracy of 86.6%, our proposed features can improve the accuracy by 4.4%. This study conducts a ground investigation of quantitative attraction assessment for video clips, which might provide reference for paradigm design of affective computing research.

Index Terms—Attraction measurement, emotional movies, ocular, orthogonal moments.

I. INTRODUCTION

ANALYSIS of pupil dilation signals and eye movements to extract emotional and cognitive states has been conducted for several decades to predict various emotional and cognitive states, including stress, memory, and other cognitive processes [1], [2], as well as a range of emotions [3], [4], [5], [6]. However, due to the complexity of the neural system that regulates eye activity and state, it is never straightforward for computer-based systems to sense and interpret emotions

Manuscript received 12 February 2023; revised 31 March 2023; accepted 7 April 2023. Date of publication 4 May 2023; date of current version 31 May 2023. This work was supported in part by the National Key Research and Development Program of China (Grant No. 2019YFA0706200), in part by the Natural Science Foundation of Gansu Province, China (Grant No. 22JR5RA488), in part by the National Natural Science Foundation of China (Grant No. 61632014, No. 61627808). Supported by Supercomputing Center of Lanzhou University. (*Corresponding author: Rong Ma*)

Minqiang Yang, Xiang Feng, and Rong Ma are with the School of Information Science and Engineering, Lanzhou University, Lanzhou 730000, China (e-mail: yangmq@lzu.edu.cn; fengx2020@lzu.edu.cn; marong@lzu.edu.cn).

Xiongying Li is with the Institute of Higher Education, Lanzhou University, Lanzhou 730000, China (e-mail: lixying@lzu.edu.cn).

Chengsheng Mao is with the Department of Preventive Medicine, Feinberg School of Medicine, Northwestern University, Chicago, IL 60611 USA (e-mail: chengsheng.mao@northwestern.edu).

Digital Object Identifier 10.1109/TCSS.2023.3268505

and cognitive states from pupil dilation signals and eye movements. Despite this, the field of affective computing using pupil dilation signals and eye movements still has room for improvement. Conventional time- and frequency-domain features may be affected by scaling, translation, and noise, reducing the effectiveness of the model. Moreover, the information redundancy between conventional features is relatively high, which causes high computational complexity and extended training time. In addition, there has been a lack of research on using the attraction level of emotional movie clips as a measure in physiological-signal-based affective computing studies.

Physiological signals are widely used in affective computing research, such as electroencephalogram (EEG) [7], [8], electromyogram (EMG) [9], electrocardiogram (ECG) [10], and pupil dilation signal [11], [12]. Pupil dilation signals, which can be well measured automatically [13], advantages over other physiological signals include: 1) ease of acquisition; 2) high signal quality (signal-to-noise ratio); and 3) immunity to the influence of ambient electromagnetic fields.

Feature extraction is a critical research area to detect emotion and cognitive information within the field of affective computing. Many previous studies of physiological signals had impressive classification accuracy using extracted time-domain, frequency-domain, and time-frequency domain features of specific frequency bands [14], [15]. We referred to these features as conventional features in this article. However, given the complexity of physiological signals, it is clear that there is still much room for improvement in feature extraction. In this article, we will use novel features in combination with conventional features to get better classification performance.

The moments are a probabilistic and statistical concept often used as an image feature extraction operator, representing a random variable's numerical characteristic [16]. According to the theory of moments, signals can be projected onto polynomials of different orders, producing moments of different orders. If the polynomials are orthogonal to each other, the moments of different orders have minimal information redundancy [17], [18]. The orthogonal moments can be divided into continuous and discrete moments according to whether the definition domain belongs to the real field [19]. The primary continuous orthogonal moments include Zernike, pseudo-Zernike, Legendre, Bessel-Fourier [20], etc. Since the basis function of the Legendre moments is defined in the Cartesian system, the calculation is relatively simple [21]. Moreover, it performs better than Zernike and pseudo-Zernike at resisting white noise [22]. Furthermore, Legendre

moments have been successfully applied in pattern matching [17] and biometric [23], [24] and ECG signal denoising [25]. Similarly, the primary discrete moments include Krawtchouk, discrete Hahn, and Tchebichef [20]. However, the discrete Hahnc [26] and Tchebichef [27] are unstable when the order of the moments becomes larger. Besides, Krawtchouk moments have also been proven effective for classification purposes over 2-D images and 3-D objects [28], [29], [30]. In addition, feature extraction methods based on these two orthogonal moments are rarely used to analyze pupil dilation signals and eye movements for affective computing studies. Considering the above, we selected the Legendre and Krawtchouk moment from the continuous moments and discrete moments for experiments, which may mitigate the flaws of conventional features.

Emotional movie clips are widely used as affective computing stimulus [31], [32]. In this article, we explored attraction assessment by ocular hints in emotional movie clip stimulation circumstances. We proposed a feature extraction method for pupil dilation signals based on orthogonal moments. The main contributions of this article can be summarized in two points.

- 1) We designed an emotional evocation experimental paradigm using emotional movie clips as stimulus material. Pupil dilation signals and eye movements were collected from 50 subjects. The data can be used to study attraction levels.
- 2) We proposed an orthogonal-moment-based feature extraction method for pupil dilation signals. Subsequently, we compared it with the conventional feature set in terms of independence and informativeness. Furthermore, we conducted ablation experiments and quantitative comparison tests, which verified that the orthogonal-moment-based feature extraction method of pupil dilation signals could extract effective features for emotion classification.

The rest of this article is organized in the following manner. Section II presents the design and procedure of the data acquisition experiment. Section III describes our methods, including data preprocessing, feature extraction based on orthogonal moments, feature space exploration, and classification experiments. Section IV introduces the ablation experiments and comparative analysis. Section V summarizes the experimental results and discusses possible future research.

II. DATA ACQUISITION

In this section, pupil dilation signals and eye movements were collected from 50 participants while watching emotional movie clips. We also collected their attraction level and emotion report for each clip.

A. Experiment Preparation

1) *Participant:* A total of 50 volunteers (37 males and 13 females) from Lanzhou University participated in the experiment. Subjects were thoroughly informed about the research and signed an informed consent form. The average age of the subjects was 19 years ($SD = 0.78$, range = 17–21 years). Before starting the experiment, each subject would complete the Patient Health Questionnaire-9 (PHQ-9) and Generalized

TABLE I
MOVIES USED TO EVOKE EMOTIONS AND THEIR TYPES
AND SCORES (AS OF MARCH 2023)

Title	Type	Score
Let The Bullets Fly	Drama,Comedy	9.0
Mr. Bean's Holiday	Comedy	8.2
Goodbye Mr. Loser	Comedy	7.8
To Live	Tragedy	9.3
My Best Summer	Drama,Romance,Tragedy	5.1
Shanghai Fortress	Science Fiction, Action	2.9
John Wick: Chapter3	Action	7.9
Pure Hearts: Into Chinese Showbiz	Drama	2.2

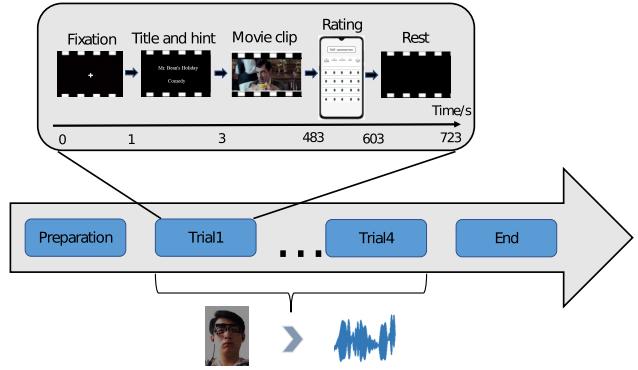


Fig. 1. Experimental paradigm.

Anxiety Disorder-7 (GAD-7) to ensure emotional health to participate in the experiment.

2) *Experimental Equipment:* The experiment was conducted in a soundproof, moderately bright room with no outside interference. A 21-in monitor for displaying the stimulus material was placed 60 cm away from the participant. The data acquisition tool used for the experiment was a Pupil Core eye tracker [33] with a sampling frequency of 250 Hz. The eye-tracking software, Pupil Capture, was used to view and record the data and to complete the confidence assessment.

3) *Emotional Inducer:* Since the subjects were all from China, we selected six movie clips with Chinese language and two movie clips without language content as the stimulus material, as shown in Table I. The movies were selected from the high to the low rating to avoid the influence of movie quality on the results, the ratings of which were collected from Douban. We selected 8-min clip of each movie with apparent emotion and could be understood without explanation as the stimulus material.

B. Experimental Procedure

The experiment consisted of four trials for each participant. Each trial presented a movie clip and allowed the subject to complete a self-report. The whole experiment took about 48 min, of which each trial lasted about 12 min (including a 2-min break). The experimental paradigm is shown in Fig. 1. To avoid the influence of the former trial on the latter one, we ensured that the emotions of the clips used in the two adjacent trials were not the same. In the preparation phase, we brought the subject into the room. We placed and adjusted

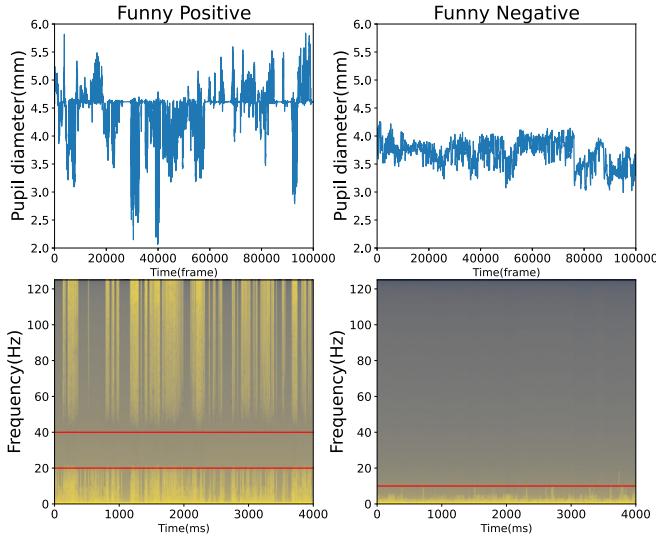


Fig. 2. Waveform and time–frequency diagrams.

the eye tracker to ensure the subjects are comfortable. Before the experiment, each subject randomly chooses a set of movie clips.

The specific steps of each trial are as follows.

- 1) Before the formal experiment, 3 s was used to show the title and genre of the movie.
- 2) 8 min of play of the stimulus material.
- 3) After viewing, there is a 2-min period for self-report, which include labels of attraction level, fun, sadness, and excitement.
- 4) Finally, there is a 2-min break.

III. METHOD

In this section, we preprocessed the data and extracted five feature sets. Then, we compared our proposed feature sets with conventional feature sets regarding independence and informativeness. After that, we combined the feature sets and tested them on three classifiers.

A. Data Preprocessing

The labels we collected during the experimental phase were Likert scale scores with a total score of 5 [34]. To simplify the model, we defined the label's score greater than or equal to 3 as positive and those less than 3 as negative. Fig. 2 shows an example of waveform and time–frequency diagram for two groups of pupil dilation signals with different labels after viewing the same emotional movie clip. In the time–frequency diagram, the bright part represents the concentration of the energy. For the negative ones, most of the energy is concentrated below 10 Hz. In contrast, in the positive ones, a significant portion of the energy is concentrated above 40 Hz. This difference makes it possible to predict emotions using pupil dilation signals.

For data cleaning, we first removed the data with a confidence level below 0.95 based on the confidence assessment of each piece of data by eye tracker. After that, we used moving average interpolation to restore the data to 250 Hz. Finally,

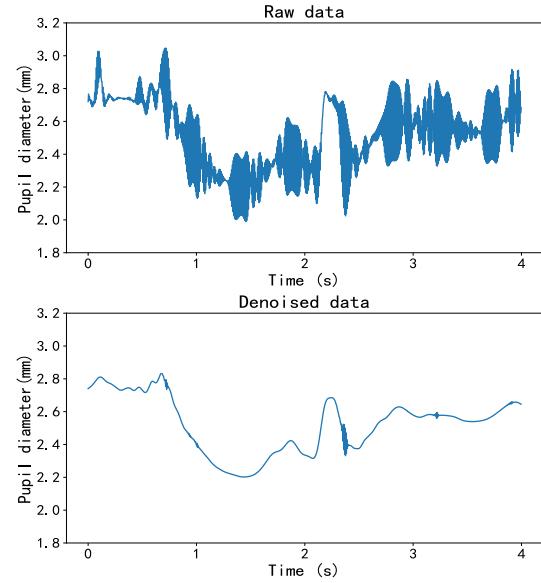


Fig. 3. Waveform before and after denoising.

we use the bandpass filtering technique to remove the data below 12.5 Hz or above 112.5 Hz as noise [35]

$$T = \frac{\text{median}(|cD1|) * \frac{\sqrt{2 \ln N}}{\log_2(j+1)}}{0.6745}. \quad (1)$$

After that, we used a soft-threshold wavelet denoising technique for further data cleaning, and the threshold function is shown in (1) [36]. We selected the Daubechies family with orthogonality, high vanishing moments, tight branching, and symmetry as wavelet bases. After observing the denoising effect of three to eight layers, we finally chose to use a decomposition reconstruction of six layers. The waveforms of 4-s signals before and after denoising are shown in Fig. 3.

B. Feature Extraction

The features extracted from pupil dilation signals and eye movements can be divided into five sets. The first is the time-domain features, including eye movements, denoted as feature set T [37]. Since we only used blink frequency in eye movements. For simplicity, we merged it into the time-domain feature set. The second is the frequency-domain features calculated from pupil dilation signals, denoted as feature set F [14]. The third feature set is the percentage of different eye movement patterns obtained based on sliding windows, denoted as feature set S [38]. The fourth is the first 15 Legendre moments calculated from pupil dilation signals, denoted as feature set L. The fifth is the Krawtchouk moments of first five orders for the pupil dilation signals as feature set K. The features contained in the first two feature sets are shown in Table II.

When dealing with conventional images with intensity function $f(x, y)$, the Legendre moment is defined as [39]

$$L_{pq} = \frac{(2p+1)(2q+1)}{4} \int_{-1}^1 \int_{-1}^1 P_p(x) P_q(y) f(x, y) dx dy. \quad (2)$$

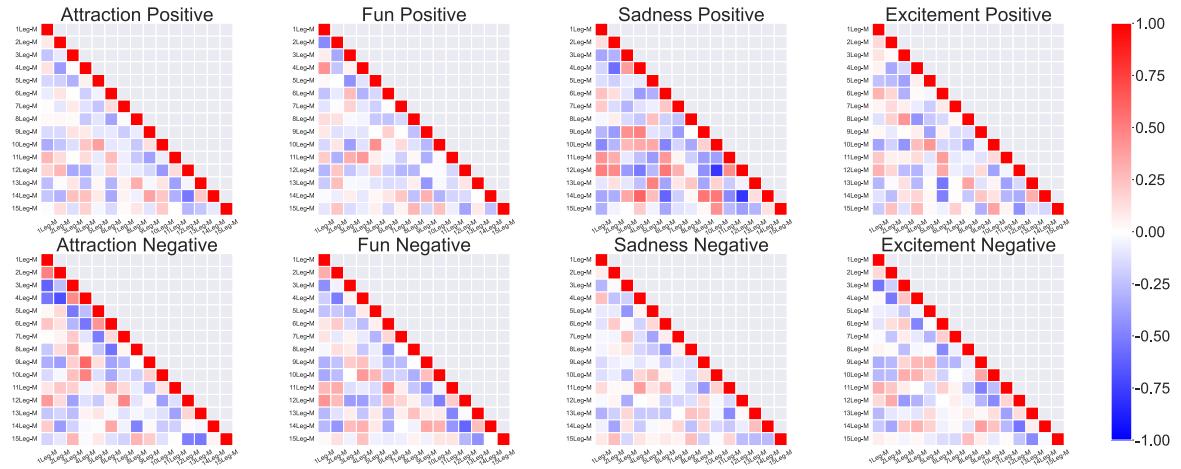


Fig. 4. Heat map of Pearson's correlation coefficient between moments.

TABLE II
TIME-DOMAIN FEATURES AND FREQUENCY-DOMAIN FEATURES

Feature Set	Features
T	max, min, pp distance, mean, SRM, std, var, c3, auto correlation, binned entropy complexity, RMS, skewness, slope, kurtosis form factor, margin factor, CREST factor pulse factor, MAC, blink frequency
F	spectrum centroid, spectrum variance, spectrum skew, spectrum kurtosis, spectrum coefficient, wavelet abs mean, wavelet entropy, wavelet var, wavelet std, wavelet energy

By projecting the 1-D signal to the Legendre polynomials, the formula for the 1-D Legendre moments was introduced [40]

$$L_p = \frac{2p+1}{N-1} \sum_{i=1}^N P_p(x_i) f(x_i) \quad (3)$$

where $x_i = ((2i - N - 1)/(N - 1))$, $P_p(x)$ is the p th-order Legendre polynomial.

The p th-order Legendre polynomial is given by

$$P_p(x) = \frac{1}{2^p} \sum_{k=0}^{p/2} (-1)^k \frac{(2p-2k)!}{k!(p-k)!(p-2k)!} x^{p-2k}. \quad (4)$$

The following recursive form of the Legendre polynomial was used to reduce the computational complexity [41]:

$$P_{p+1}(x) = \frac{2p+1}{p+1} x P_p(x) - \frac{p}{p+1} P_{p-1}(x) \quad (5)$$

where $P_0(x) = 1$ and $P_1(x) = x$.

We compute the first 15 orders of Legendre moments for the pupil dilation signals. Only the first 15 orders are extracted as set L since using higher orders increases complexity and does not provide more information.

Algorithm 1 Flow for Calculating Legendre Moments and Krawtchouk Moments

Input: Pupil dilation signal matrix M with size of $s \times N$
Output: Legendre moments matrix and Krawtchouk moments matrix with size of $s \times 15$, $s \times 5$

Initialization: $Lmoments \leftarrow \{\}, Kmements \leftarrow \{\}$

```

1: for  $p = 0$  to  $s$  do
2:   Initialize  $Larray \leftarrow [\dots]_{15}$  and  $Karray \leftarrow [\dots]_5$ 
3:   /* calculate Legendre moments */
4:   for  $order = 0$  to  $14$  do
5:      $f_1 \leftarrow \frac{2order+1}{N-1}$ 
6:      $f_2 \leftarrow 0$ 
7:     for  $index = 1$  to  $N$  do
8:        $f_2 \leftarrow f_2 + P_{order}(M[p][x_{index}])$ 
9:     end for
10:    Update  $Larray$ , appending the value of  $f_1 \times f_2$ 
11:  end for
12:  Update  $Loments$ , appending the  $Larray$ 
13:  /* calculate Krawtchouk moments */
14:  for  $order = 0$  to  $4$  do
15:     $f \leftarrow 0$ 
16:    for  $index = 1$  to  $N$  do
17:       $f \leftarrow f + \bar{K}_{order}(index-1) \times M[p][x_{index}]$ 
18:    end for
19:    Update  $Karray$ , appending the value of  $f$ 
20:  end for
21:  Update  $Koments$ , appending the  $Karray$ 
22: end for
23: return  $Loments$  and  $Koments$ 

```

The formula for calculating the 1-D Krawtchouk moments is given by [28], [42]

$$\bar{Q}_n = \sum_{i=1}^N \bar{K}_n(i-1; p, N-1) f(x_i) \quad (6)$$

where \bar{K}_n is weighted Krawtchouk polynomial and can be defined as

$$\bar{K}_n(x; p, N) = K_n(x; p, N) \sqrt{\frac{w(x; p, N)}{r(n; p, N)}} \quad (7)$$

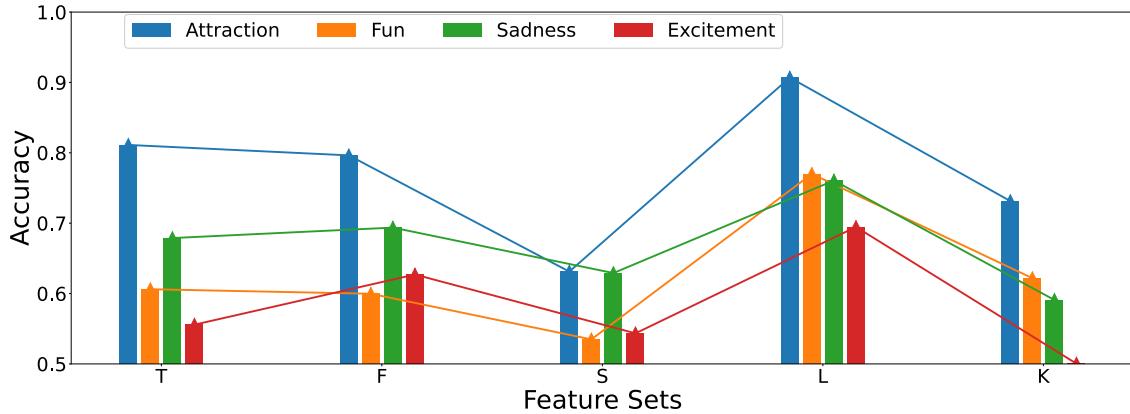


Fig. 5. Accuracy when using five feature sets, respectively.

and $r(n; p, N)$ and $w(n; p, N)$ are given by

$$w(n; p, N) = \left(\frac{N}{x}\right)^n p^x (1-p)^{N-x} \quad (8)$$

$$r(n; p, N) = (-1)^n \left(\frac{1-p}{p}\right)^n \frac{n!}{(-N)_n}. \quad (9)$$

To simplify the calculation, recursive form of K_n and $w(x; p, N)$ was introduced [43], [44]

$$K_{n+1} = \left(1 + \frac{n - np - x}{pN - pn}\right) K_n - \frac{n - np}{pN - pn} K_{n-1} \quad (10)$$

$$w(x+1; p, N) = \frac{w(x; p, N)p(N-x)}{x+1-p-xp} \quad (11)$$

where $K_0(x; p, N) = 1$ and $w(0; p, N) = (1-p)^N$.

Similarly, we compute only the first five orders of Krawtchouk moments as set K.

In the case of feature set S, we calculated the proportion of different eye movement patterns using sliding windows and unsupervised clustering. Considering that the pupil dilates once in 2–7 s, we set a window size of 4 and a window coverage of 50%. Within this window, we calculated a series of features, including the longest continuous dilation time, the proportion of dilation time to the total time, and the dilation maximum [45]. Afterward, we applied unsupervised clustering algorithm to these features. All the windows were considered as three different eye movement patterns [38]. The weights of three patterns are taken as the set S.

C. Feature Space Exploratory

A superior feature set should satisfy two conditions, i.e., independent and informative. The former means that there is less redundant information between features. The latter means that the features can provide sufficient information for classification. We used a Pearson correlation coefficient heat map and the accuracy using five feature sets, respectively, to verify these two points.

1) *Independent*: Since self-reports of four emotions, each with two attributes, were collected, we have eight kinds of emotion labels. We calculated the Pearson correlation coefficients for the first 15 orders of Legendre moment under the

eight kinds of emotion labels and plotted them as heat maps, as shown in Fig. 4. Most regions of the heat map are white or light-colored, representing little redundant information among the features. Different features can extract different information from the data. In particular, the 4th and 15th of Legendre moments correlate weakly with other orders in most cases. Low correlation represents that the Legendre moments are independent and suitable for classification.

2) *Informative*: We calculated the classification accuracy of different labels using five feature sets separately and plotted them in a histogram, as shown in Fig. 5. We used the support vector machine (SVM) classifier for classification and did not perform feature selection. It is observed from the figure that the feature set based on the Legendre moments has the best classification results for the five feature sets. The accuracy of the feature set based on the Krawtchouk moments is similar to the conventional time-domain, frequency-domain feature set. These two points imply that these two moments are informative and can be used to improve emotion classification performance.

D. Classification

We designed ablation experiments with control and baseline for testing the validity of the above two orthogonal moments as features of the pupil dilation signals. We used T + F and L + K for the classification baseline, respectively. We used the feature set S as a control for comparing L and K. The feature sets S, L + K, and S + L + K were combined on the feature set with T + F. The feature combinations used in the experiments are L + K, T + F, T + F + S, T + F + L + K, and T + F + S + L + K. Moreover, the fusion method chosen for the classification is traditional feature level fusion using the above combinations.

The selected classifiers were the most common SVM [46], k-nearest neighbor (kNN) [47], and random forest (RF) [48]. Then, we upsample the data using the adaptive synthetic sampling approach to reduce label imbalance [49]. Furthermore, we normalize the data using the z-score method [50]. The feature selection algorithm used for the experiments is an SVM-based feature recursive elimination algorithm. For the classification, we randomly selected 80% of the data as the

TABLE III

ATTRACTION CLASSIFICATION RESULTS OF DIFFERENT FEATURE SETS(**BOLD**: BEST PERFORMANCE FOR EACH CLASSIFIER)

Feature Set	Classifier	Classification Metrics			
		Accuracy[SD]	Precision[SD]	Recall[SD]	F1 score[SD]
L+K	SVM	0.849 [0.06]	0.909 [0.07]	0.787 [0.09]	0.840 [0.06]
	kNN	0.802 [0.05]	0.923 [0.08]	0.671 [0.09]	0.772 [0.07]
	RF	0.874 [0.05]	0.904 [0.07]	0.847 [0.07]	0.872 [0.05]
T+F (Baseline)	SVM	0.828 [0.06]	0.886 [0.10]	0.759 [0.10]	0.811 [0.07]
	kNN	0.785 [0.06]	0.927 [0.09]	0.639 [0.10]	0.751 [0.08]
	RF	0.866 [0.05]	0.930 [0.07]	0.802 [0.09]	0.857 [0.06]
+S	SVM	0.881 [0.04]	0.964 [0.06]	0.796 [0.09]	0.867 [0.06]
	kNN	0.782 [0.06]	0.987 [0.04]	0.574 [0.10]	0.721 [0.08]
	RF	0.892 [0.05]	0.956 [0.06]	0.831 [0.09]	0.885 [0.05]
+L+K	SVM	0.902 [0.05]	0.963 [0.06]	0.839 [0.08]	0.894 [0.05]
	kNN	0.807 [0.06]	0.957 [0.06]	0.641 [0.10]	0.763 [0.08]
	RF	0.910 [0.05]	0.950 [0.06]	0.869 [0.08]	0.905 [0.05]
+S+L+K	SVM	0.914 [0.04]	0.963 [0.05]	0.868 [0.08]	0.910 [0.05]
	kNN	0.773 [0.07]	0.947 [0.08]	0.582 [0.11]	0.714 [0.09]
	RF	0.909 [0.04]	0.953 [0.06]	0.861 [0.08]	0.902 [0.05]

TABLE IV

FUN CLASSIFICATION RESULTS OF DIFFERENT FEATURE SETS(**BOLD**: BEST PERFORMANCE FOR EACH CLASSIFIER)

Feature Set	Classifier	Classification Metrics			
		Accuracy[SD]	Precision[SD]	Recall[SD]	F1 score[SD]
L+K	SVM	0.738 [0.06]	0.751 [0.11]	0.741 [0.11]	0.736 [0.07]
	kNN	0.667 [0.07]	0.630 [0.10]	0.829 [0.10]	0.711 [0.08]
	RF	0.776 [0.07]	0.778 [0.11]	0.768 [0.12]	0.766 [0.09]
T+F (Baseline)	SVM	0.595 [0.07]	0.590 [0.10]	0.671 [0.16]	0.610 [0.09]
	kNN	0.600 [0.08]	0.564 [0.10]	0.775 [0.10]	0.645 [0.08]
	RF	0.644 [0.08]	0.635 [0.11]	0.665 [0.13]	0.639 [0.10]
+S	SVM	0.665 [0.07]	0.659 [0.11]	0.740 [0.10]	0.689 [0.08]
	kNN	0.606 [0.08]	0.587 [0.10]	0.813 [0.10]	0.676 [0.08]
	RF	0.673 [0.07]	0.669 [0.12]	0.731 [0.11]	0.689 [0.08]
+L+K	SVM	0.752 [0.07]	0.750 [0.12]	0.730 [0.12]	0.732 [0.08]
	kNN	0.664 [0.08]	0.623 [0.10]	0.815 [0.09]	0.700 [0.07]
	RF	0.753 [0.07]	0.764 [0.11]	0.723 [0.13]	0.731 [0.08]
+S+L+K	SVM	0.773 [0.08]	0.779 [0.13]	0.756 [0.12]	0.758 [0.09]
	kNN	0.672 [0.08]	0.620 [0.11]	0.863 [0.09]	0.715 [0.08]
	RF	0.749 [0.08]	0.761 [0.11]	0.711 [0.13]	0.726 [0.09]

TABLE V

SADNESS CLASSIFICATION RESULTS OF DIFFERENT FEATURE SETS(**BOLD**: BEST PERFORMANCE FOR EACH CLASSIFIER)

Feature Set	Classifier	Classification Metrics			
		Accuracy[SD]	Precision[SD]	Recall[SD]	F1 score[SD]
L+K	SVM	0.723 [0.07]	0.738 [0.11]	0.668 [0.12]	0.692 [0.09]
	kNN	0.678 [0.09]	0.629 [0.11]	0.839 [0.09]	0.712 [0.08]
	RF	0.771 [0.08]	0.753 [0.11]	0.782 [0.11]	0.761 [0.09]
T+F (Baseline)	SVM	0.744 [0.07]	0.717 [0.09]	0.828 [0.10]	0.762 [0.07]
	kNN	0.685 [0.08]	0.664 [0.10]	0.768 [0.11]	0.705 [0.08]
	RF	0.729 [0.07]	0.723 [0.11]	0.764 [0.10]	0.736 [0.08]
+S	SVM	0.728 [0.06]	0.687 [0.11]	0.824 [0.09]	0.742 [0.07]
	kNN	0.699 [0.08]	0.651 [0.10]	0.834 [0.11]	0.724 [0.08]
	RF	0.730 [0.07]	0.718 [0.10]	0.754 [0.11]	0.727 [0.07]
+L+K	SVM	0.784 [0.07]	0.767 [0.11]	0.808 [0.09]	0.781 [0.08]
	kNN	0.705 [0.06]	0.647 [0.09]	0.861 [0.08]	0.734 [0.07]
	RF	0.792 [0.07]	0.798 [0.10]	0.790 [0.12]	0.785 [0.08]
+S+L+K	SVM	0.802 [0.06]	0.770 [0.10]	0.840 [0.10]	0.797 [0.07]
	kNN	0.704 [0.07]	0.634 [0.10]	0.890 [0.08]	0.734 [0.07]
	RF	0.776 [0.07]	0.773 [0.11]	0.768 [0.12]	0.762 [0.09]

training set and 20% as the test set. For each classification, we repeated the experiment 200 times. Tables III–VI show the

TABLE VI

EXCITEMENT CLASSIFICATION RESULTS OF DIFFERENT FEATURE SETS(**BOLD**: BEST PERFORMANCE FOR EACH CLASSIFIER)

Feature Set	Classifier	Classification Metrics			
		Accuracy[SD]	Precision[SD]	Recall[SD]	F1 score[SD]
L+K	SVM	0.640 [0.08]	0.692 [0.19]	0.445 [0.16]	0.514 [0.13]
	kNN	0.588 [0.09]	0.564 [0.14]	0.614 [0.13]	0.575 [0.10]
	RF	0.659 [0.09]	0.690 [0.15]	0.551 [0.17]	0.591 [0.12]
T+F (Baseline)	SVM	0.608 [0.08]	0.648 [0.18]	0.402 [0.15]	0.472 [0.12]
	kNN	0.613 [0.08]	0.558 [0.13]	0.582 [0.16]	0.552 [0.11]
	RF	0.642 [0.10]	0.618 [0.18]	0.560 [0.15]	0.567 [0.12]
+S	SVM	0.608 [0.08]	0.622 [0.19]	0.364 [0.14]	0.433 [0.12]
	kNN	0.543 [0.10]	0.490 [0.17]	0.500 [0.16]	0.479 [0.14]
	RF	0.616 [0.09]	0.597 [0.16]	0.509 [0.15]	0.532 [0.12]
+L+K	SVM	0.576 [0.07]	0.612 [0.17]	0.387 [0.15]	0.450 [0.12]
	kNN	0.525 [0.09]	0.496 [0.11]	0.613 [0.15]	0.534 [0.09]
	RF	0.614 [0.09]	0.597 [0.16]	0.538 [0.14]	0.550 [0.12]
+S+L+K	SVM	0.597 [0.08]	0.596 [0.18]	0.431 [0.15]	0.480 [0.12]
	kNN	0.506 [0.09]	0.481 [0.13]	0.552 [0.13]	0.504 [0.11]
	RF	0.608 [0.09]	0.608 [0.15]	0.530 [0.16]	0.551 [0.12]

TABLE VII

ATTRACTION T-STATISTICS AND p-VALUES FOR PAIRED T-TESTS WITH DIFFERENT FEATURE SETS(*: p < 0.05, **: p < 0.01, ***: p < 0.001)

Null Hypothesis	Classifier	Classification Indicators			
		Accuracy	Precision	Recall	F1 score
≥ +S	SVM	-7.758***	-6.850***	-3.055**	-6.747***
	kNN	0.257	-5.944***	4.193	2.498
	RF	-3.645***	-2.913**	-2.322*	-3.550***
Baseline ≥ +L+K	SVM	-10.006***	-5.998***	-6.813***	-9.267***
	kNN	-2.404**	-2.636**	-0.136	-1.066
	RF	-6.589***	-2.442**	-6.098***	-6.274***
≥ +S+L+K	SVM	-11.612***	-6.569***	-8.156***	-11.413***
	kNN	1.363	-1.624	3.958	3.303
	RF	-6.226***	-2.493**	-5.339***	-5.757***
T+F+S ≥ T+F+L+K	SVM	-3.006**	0.138	-3.658***	-3.178***
	kNN	-2.842**	4.105	-4.782***	-3.726***
	RF	-2.515**	0.766	-3.115**	-2.500**

TABLE VIII

FUN T-STATISTICS AND p-VALUES FOR PAIRED t-TESTS WITH DIFFERENT FEATURE SETS(*: p < 0.05, **: p < 0.01, ***: p < 0.001)

Null Hypothesis	Classifier	Classification Indicators			
		Accuracy	Precision	Recall	F1 score
≥ +S	SVM	-6.932***	-4.956***	-3.433***	-6.638***
	kNN	-0.599	-1.666*	-2.794**	-2.837**
	RF	-2.477**	-2.114*	-3.585***	-3.612***
Baseline ≥ +L+K	SVM	-16.460***	-10.603***	-3.250***	-10.669***
	kNN	-5.815***	-4.018***	-2.794**	-5.159***
	RF	-11.059***	-8.337***	-3.340***	-8.143***
≥ +S+L+K	SVM	-16.739***	-11.992***	-3.939***	-11.925***
	kNN	-6.609***	-3.723***	-6.328***	-6.041***
	RF	-9.565***	-7.580***	-2.386**	-6.323***
T+F+S ≥ T+F+L+K	SVM	-8.512***	-5.985***	0.573	-3.568***
	kNN	-4.907***	-2.666**	-0.150	-2.188*
	RF	-8.253***	-6.076***	0.501	-3.779***

mean values of the four metrics predicting the results of the four emotion labels.

IV. RESULT AND ANALYSIS

A. Experimental Results' Analysis

To verify the feasibility of the proposed orthogonal-moment-based method for feature extraction of pupil dilation

TABLE IX
SADNESS T-STATISTICS AND p-VALUES FOR PAIRED t-TESTS WITH
DIFFERENT FEATURE SETS(*: p < 0.05, **: p < 0.01, ***: p < 0.001)

Null Hypothesis	Classifier	Classification Indicators			
		Accuracy	Precision	Recall	F1 score
$\geq +S$	SVM	1.625	1.973	0.296	1.977
	kNN	-1.319	0.889	-4.069***	-1.717*
	RF	-0.142	0.283	0.682	0.792
Baseline $\geq +L+K$	SVM	-3.726***	-3.319***	1.343	-1.708*
	kNN	-2.031*	1.261	-6.408***	-2.705**
	RF	-6.408***	-5.127***	-1.684*	-4.563***
$\geq +S+L+K$	SVM	-6.223***	-3.950***	-0.879	-3.701***
	kNN	-1.944*	2.091	-8.704***	-2.727**
	RF	-4.794***	-3.443***	-0.250	-2.295*
$T+F+S \geq T+F+L+K$	SVM	-5.837***	-5.146***	1.175	-3.637***
	kNN	-0.540	0.283	-2.018*	-0.867
	RF	-6.028***	-5.510***	-2.212*	-5.169***

TABLE X
EXCITEMENT T-STATISTICS AND p-VALUES FOR PAIRED t-TESTS WITH
DIFFERENT FEATURE SETS(*: p < 0.05, **: p < 0.01, ***: p < 0.001)

Null Hypothesis	Classifier	Classification Indicators			
		Accuracy	Precision	Recall	F1 score
$\geq +S$	SVM	0.063	1.029	1.919	2.254
	kNN	5.758	3.338	3.929	4.747
	RF	2.050	1.081	2.321	2.060
Baseline $\geq +L+K$	SVM	2.827	1.489	0.744	1.354
	kNN	7.758	3.700	-1.348	1.376
	RF	1.925	0.924	0.949	0.919
$\geq +S+L+K$	SVM	0.992	2.040	-1.391	-0.476
	kNN	8.871	3.760	1.505	3.048
	RF	2.539	0.453	1.467	0.885
$T+F+S \geq T+F+L+K$	SVM	2.925	0.385	-1.100	-0.982
	kNN	1.417	-0.266	-5.215***	-3.256***
	RF	0.154	-0.002	-1.405	-1.078

signals, we conducted ablation experiments with baseline and control. The classification metrics of five feature sets combinations are shown in Tables III–VI.

Table III shows the average of the metrics for each feature set in predicting the attraction level. It can be observed that all the highest accuracy, recall, and f1 appear in the experiments where orthogonal moments are used. When using the SVM classifier, adding the feature set L + K to the baseline improves the accuracy by about 8%, while adding the feature set S improves it by only about 6%. Adding both L + K and S will result in an accuracy improvement of about 9% to an acceptable %. The orthogonal moment features significantly improve recall and f1 compared with feature set S. In terms of precision, the orthogonal moment features only perform worse on the kNN classifier. It has the same improvement effect as feature set S on the other two classifiers.

Table IV shows the performance of the model for predicting the fun labels. The improvement of the orthogonal moment features for the baseline is very significant compared with the feature set S. The orthogonal moment features will improve the accuracy by about 16% on SVM. In contrast, the feature set S is only about 7%. The noticeable point is that for the RF classifier, the classification is better using only orthogonal moment features. The remaining two classifiers work best by

combining orthogonal moment features, feature set S, and the baseline.

Table V shows the model performance for predicting sadness labels. When using SVM and RF classifiers, the improvement of orthogonal moment features is more pronounced than the feature set S. For the kNN classifier, the highest accuracy, recall, and F1 values occur with the use of moment features. However, the improvement is insignificant compared with the feature set S.

Table VI shows the performance of the model for predicting excitement labels. Although most of the highest values appear in the experiments using orthogonal moment features alone, the improvement is not significant compared with the baseline's performance. Moreover, the overall effect of predicting excitement is unsatisfactory.

B. Comparative Statistical Test Analysis

To test whether our proposed features statistically significantly improved the model's prediction performance, we performed paired-sample *t*-tests between the feature groups for the above experiments. Tables VII–X present the results of the statistical tests when predicting the four sets of labels. Each number in the table represents the *t* statistic of the paired samples *t*-test. The trailing “*” represents the *p*-value.

As shown in Table VII, both the feature set S and the orthogonal moment features and their combinations improve the baseline classification when predicting the attraction level. The effect of orthogonal moment features is more significant than the feature set S. Except for prediction, the metrics of the orthogonal moment feature are significantly higher than that of feature set S.

From Table VIII, the orthogonal moment features can improve the three classifiers when predicting the fun labels. However, feature set S can only improve SVM and RF. Furthermore, except for recall which is the same, orthogonal moment features also significantly outperform feature set S in other metrics. This difference indicates that the generalization of orthogonal moment features is better for predicting fun labels.

We can see more evidently the improvement of the orthogonal moment features on the baseline when predicting sadness from Table IX. Meanwhile, the feature set S has no significant improvement effect. Table X shows that neither the orthogonal moment features nor the feature set S has evident improvement on the baseline when predicting excitement.

V. DISCUSSION AND CONCLUSION

A. Discussion

Due to previous studies neglecting the importance of the attraction level, few affective computing research investigated the relationship between attraction level and ocular hints (i.e., pupil dilation signals and eye movements). Moreover, affective computing experiments have never adopted orthogonal moments as feature extractors of the pupil dilation signals. With the ablation experiments, our work makes the first step to exploring the above two problems.

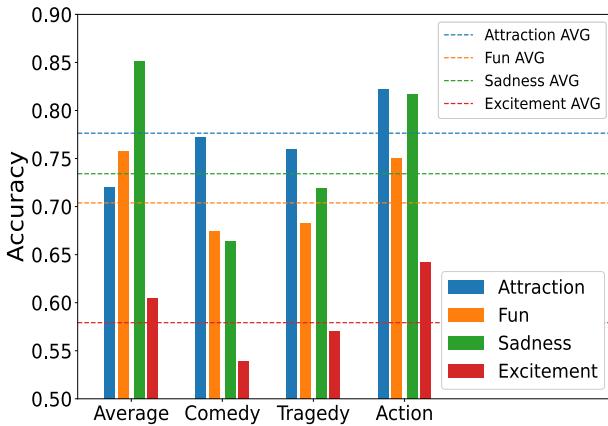


Fig. 6. Accuracy of classification of different movie genres.

1) Feature Analysis: We perform correlation analysis on the Legendre moments to determine whether the features are informative and independent. The Pearson correlation coefficient heat map shows very little information redundancy between the Legendre moments. Most of the images are light or colorless (low correlation). Furthermore, when using the five feature sets separately, the feature set based on the Legendre moments achieves better classification accuracy than the conventional time- and frequency-domain features. The feature set based on the Krawtchouk moments has a similar classification accuracy as the conventional feature sets. As a result, the features based on these two orthogonal moments can be at least as informative as the conventional feature set.

2) Model Defect: By ablation experiments, we found that using Legendre moments and Krawtchouk moments can significantly improve the model's prediction ability at the attraction level, fun, and sadness. In particular, the prediction accuracy can reach more than 91% at the attraction level. However, when predicting excitement, the model has weak classification capability. Since using only one kind of physiological signals, it is to be expected that the model cannot have a valid classification for all emotions [51]. Furthermore, except for excitement which has higher arousal in arousal-valance circumplex model, the arousal of the remaining emotions is neutral [52]. When a subject's excitement is negative, it can be considered in the symmetric emotion in the valence-arousal model coordinate system about the origin, i.e., bored [52]. Arousal difference between these two emotions is significant. However, it has been previously pointed out that a linear relationship exists between pupil diameter and arousal [53]. We extracted too many features, which may have led to dimensional disaster and an overfitting phenomenon.

3) Further Discussion: To further explore the model's validity in different situations, we calculated the average accuracy of the three classifiers using five feature sets simultaneously when different kinds of movies were used as emotional inducers. Figs. 6 and 7 show the models' accuracy for different genres and different score bands of movies, where the dashed line represents the average classification accuracy of one kind of emotion. Among them, the difference in the model's prediction accuracy for sadness is more significant under different kinds of movie clip stimuli. Similarly, the

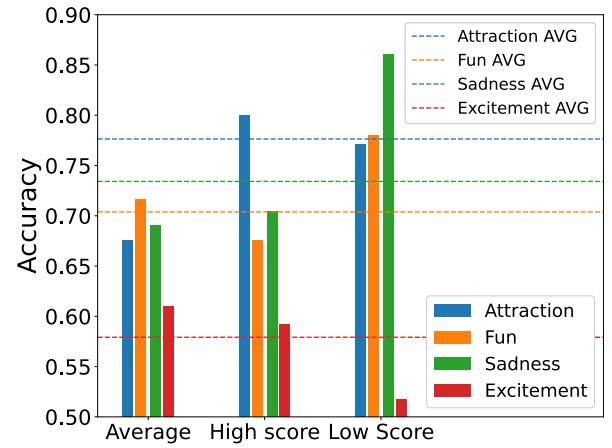


Fig. 7. Accuracy for different scores of movie.

accuracy of the models for predicting attraction also differed significantly for movies in different score bands. We will continue investigating the reasons for this evident difference in future work.

B. Conclusion

This article focuses on the relationship between ocular hints, i.e., pupil dilation signals and eye movements, and the attraction level under emotional movie clip stimulation. The study shows that the pupil dilation signals and eye movements can objectively reflect the attraction level of subjects under video stimulus. We proposed an orthogonal-moment-based feature extraction method for pupil dilation signals and conducted comparative test and ablation experiments. The results show that the proposed method achieves higher emotional classification accuracy and lower information redundancy, which means that the proposed method could provide quantitatively reference for attraction level assessment. Furthermore, we built a new database of eye movements and pupil dilation signals that can be used to study the attraction level under emotional movie clip stimulation.

However, there are some limitations in our experiment. First, we collected so little data that a deep-learning-based approach is unsuitable. Second, the experiment time is short for each subject, typically around 40 min, which may result in the emotions inspired by the previous movie clip affecting the results of the subsequent trial. Finally, in data acquisition, we used a 5-point Likert scale. Nevertheless, considering that there is a difference in how each person feels about emotional intensity, we should use a self-assessment method that is more accurate and intuitive for the subjects.

Besides the pupil dilation, the eye movement behavior patterns, e.g. fixations and saccades, which might be able to reveal the attraction level, deserve more investigation in the future. Further on, the emotional and cognitive process behind the ocular change patterns is an interesting exploratory topic.

REFERENCES

- [1] S. D. Goldinger and M. H. Papesh, "Pupil dilation reflects the creation and retrieval of memories," *Current Directions Psychol. Sci.*, vol. 21, no. 2, pp. 90–95, Apr. 2012.

- [2] M. A. Just, P. A. Carpenter, and A. Miyake, "Neuroindices of cognitive workload: Neuroimaging, pupillometric and event-related potential studies of brain work," *Theor. Issues Ergonom. Sci.*, vol. 4, nos. 1–2, pp. 56–88, Jan. 2003.
- [3] M. M. Bradley, L. Miccoli, M. A. Escrig, and P. J. Lang, "The pupil as a measure of emotional arousal and autonomic activation," *Psychophysiology*, vol. 45, no. 4, pp. 602–607, Jul. 2008.
- [4] T. Partala and V. Surakka, "Pupil size variation as an indication of affective processing," *Int. J. Hum.-Comput. Stud.*, vol. 59, nos. 1–2, pp. 185–198, Jul. 2003.
- [5] E. Geangu, P. Hauf, R. Bhardwaj, and W. Bentz, "Infant pupil diameter changes in response to others' positive and negative emotions," *PLoS ONE*, vol. 6, no. 11, Nov. 2011, Art. no. e27132.
- [6] M. Agarwal and J. Mostow, "Semi-supervised learning to perceive children's affective states in a tablet tutor," in *Proc. AAAI Conf. Artif. Intell.*, 2020, vol. 34, no. 9, pp. 13350–13357.
- [7] W. Zheng, "Multichannel EEG-based emotion recognition via group sparse canonical correlation analysis," *IEEE Trans. Cogn. Devel. Syst.*, vol. 9, no. 3, pp. 281–290, Sep. 2016.
- [8] J.-J. Guo, R. Zhou, L.-M. Zhao, and B.-L. Lu, "Multimodal emotion recognition from eye image, eye movement and EEG using deep neural networks," in *Proc. 41st Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2019, pp. 3071–3074.
- [9] B. Cheng and G. Liu, "Emotion recognition from surface EMG signal using wavelet transform and neural network," in *Proc. 2nd Int. Conf. Bioinf. Biomed. Eng.*, May 2008, pp. 1363–1366.
- [10] F. Agrafioti, D. Hatzinakos, and A. K. Anderson, "ECG pattern analysis for emotion detection," *IEEE Trans. Affective Comput.*, vol. 3, no. 1, pp. 102–115, Jan./Mar. 2012.
- [11] T. Zhang, A. E. Ali, C. Wang, X. Zhu, and P. Cesar, "CorrFeat: Correlation-based feature extraction algorithm using skin conductance and pupil diameter for emotion recognition," in *Proc. Int. Conf. Multimodal Interact.*, Oct. 2019, pp. 404–408.
- [12] W.-L. Zheng, W. Liu, Y. Lu, B.-L. Lu, and A. Cichocki, "EmotionMeter: A multimodal framework for recognizing human emotions," *IEEE Trans. Cybern.*, vol. 49, no. 3, pp. 1110–1122, Mar. 2018.
- [13] J. Zhong et al., "Filterable sample consensus based on angle variance for pupil segmentation," *Digit. Signal Process.*, vol. 130, Oct. 2022, Art. no. 103695.
- [14] B. Cheng, C. Fan, H. Fu, J. Huang, H. Chen, and X. Luo, "Measuring and computing cognitive statuses of construction workers based on electroencephalogram: A critical review," *IEEE Trans. Computat. Social Syst.*, vol. 9, no. 6, pp. 1644–1659, Dec. 2022.
- [15] S. Talmaki and V. R. Kamat, "Real-time hybrid virtuality for prevention of excavation related utility strikes," *J. Comput. Civil Eng.*, vol. 28, no. 3, May 2014, Art. no. 04014001.
- [16] K. R. Castleman, *Digital Image Processing*. Upper Saddle River, NJ, USA: Prentice-Hall, 1996.
- [17] C.-H. Teh and R. T. Chin, "On image analysis by the methods of moments," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-10, no. 4, pp. 496–513, Jul. 1988.
- [18] J. Yan et al., "Electronic nose feature extraction methods: A review," *Sensors*, vol. 15, no. 11, pp. 27804–27831, Nov. 2015.
- [19] S. Qi, Y. Zhang, C. Wang, J. Zhou, and X. Cao, "A survey of orthogonal moments for image representation: Theory, implementation, and evaluation," *ACM Comput. Surv.*, vol. 55, no. 1, pp. 1–35, Jan. 2023.
- [20] P. Kaur and H. S. Pannu, "Comprehensive review of continuous and discrete orthogonal moments in biometrics," *Int. J. Comput. Math., Comput. Syst. Theory*, vol. 3, no. 2, pp. 64–91, Apr. 2018.
- [21] P. Kaur, H. S. Pannu, and A. K. Malhi, "Comprehensive study of continuous orthogonal moments—A systematic review," *ACM Comput. Surv.*, vol. 52, no. 4, pp. 1–30, Jul. 2020.
- [22] S. J. Hosaini, S. Alirezaei, M. Ahmadi, and S. V.-A.-D. Makki, "Comparison of the Legendre, Zernike and pseudo-Zernike moments for feature extraction in iris recognition," in *Proc. 5th Int. Conf. Comput. Intell. Commun. Netw.*, Sep. 2013, pp. 225–228.
- [23] A. Sarmah and C. J. Kumar, "Iris verification using Legendre moments and KNN classifier," *Int. J. Eng. Sci. Invention*, vol. 2, no. 8, pp. 52–59, 2013.
- [24] C. L. Deepika, A. Kandaswamy, C. Vimal, and B. Satish, "Palmpoint authentication using modified Legendre moments," *Proc. Comput. Sci.*, vol. 2, pp. 164–172, Jan. 2010.
- [25] B.-H. Kwan, K.-M. Ong, and R. Paramesran, "Noise removal of ECG signals using Legendre moments," in *Proc. IEEE Eng. Med. Biol. 27th Annu. Conf.*, 2005, pp. 5627–5630.
- [26] B. M. Mahmood, S. H. Abdulhussain, T. Suk, and A. Hussain, "Fast computation of Hahn polynomials for high order moments," *IEEE Access*, vol. 10, pp. 48719–48732, 2022.
- [27] S. H. Abdulhussain, B. M. Mahmood, T. Baker, and D. Al-Jumeily, "Fast and accurate computation of high-order Tchebichef polynomials," *Concurrency Comput., Pract. Exper.*, vol. 34, no. 27, p. e7311, Dec. 2022.
- [28] P. T. Yap, R. Paramesran, and S.-H. Ong, "Image analysis by Krawtchouk moments," *IEEE Trans. Image Process.*, vol. 12, no. 11, pp. 1367–1377, Nov. 2003.
- [29] A. Mademlis, A. Axenopoulos, P. Daras, D. Tzovaras, and M. G. Strintzis, "3D content-based search based on 3D Krawtchouk moments," in *Proc. 3rd Int. Symp. 3D Data Process., Visualizat., Transmiss. (3DPVT)*, Jun. 2006, pp. 743–749.
- [30] D. Ioannidis, D. Tzovaras, I. G. Damousis, S. Argyropoulos, and K. Moustakas, "Gait recognition using compact feature extraction transforms and depth information," *IEEE Trans. Inf. Forensics Security*, vol. 2, no. 3, pp. 623–630, Sep. 2007.
- [31] A. Maffei and A. Angrilli, "Spontaneous blink rate as an index of attention and emotion during film clips viewing," *Physiol. Behav.*, vol. 204, pp. 256–263, May 2019.
- [32] M. Yang, Y. Ma, Z. Liu, H. Cai, X. Hu, and B. Hu, "Undisturbed mental state assessment in the 5G era: A case study of depression detection based on facial expressions," *IEEE Wireless Commun.*, vol. 28, no. 3, pp. 46–53, Jun. 2021.
- [33] M. Kassner, W. Patera, and A. Bulling, "Pupil: An open source platform for pervasive eye tracking and mobile gaze-based interaction," in *Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Comput.*, Sep. 2014, pp. 1151–1160.
- [34] K. Rasmussen, *Encyclopedia of Measurement and Statistics*, vol. 1. Sage, 2007.
- [35] J. M. Findlay, "Frequency analysis of human involuntary eye movement," *Kybernetik*, vol. 8, no. 6, pp. 207–214, Jun. 1971.
- [36] L. Jing-yi, L. Hong, Y. Dong, and Z. Yan-sheng, "A new wavelet threshold function and denoising application," *Math. Problems Eng.*, vol. 2016, pp. 1–8, 2016.
- [37] C. Mao, B. Hu, M. Wang, and P. Moore, "EEG-based biometric identification using local probability centers," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2015, pp. 1–8.
- [38] M. Yang, J. Tang, Y. Wu, Z. Liu, X. Hu, and B. Hu, "A behaviour patterns extraction method for recognizing generalized anxiety disorder," in *Proc. IEEE Int. Conf. E-Health Netw., Appl. Services (HEALTHCOM)*, Mar. 2021, pp. 1–4.
- [39] R. Mukundan and K. Ramakrishnan, *Moment Functions in Image Analysis: Theory and Applications*. Singapore: World Scientific, 1998.
- [40] G. Y. Yang, H. Z. Shu, C. Toumoulin, G. N. Han, and L. M. Luo, "Efficient Legendre moment computation for grey level images," *Pattern Recognit.*, vol. 39, no. 1, pp. 74–80, Jan. 2006.
- [41] R. Mukundan and K. R. Ramakrishnan, "Fast computation of Legendre and Zernike moments," *Pattern Recognit.*, vol. 28, no. 9, pp. 1433–1442, Sep. 1995.
- [42] P. A. Raj and A. Venkataramana, "Fast computation of inverse Krawtchouk moment transform using Clenshaw's recurrence formula," in *Proc. IEEE Int. Conf. Image Process.*, Apr. 2007, pp. 1–37.
- [43] R. Koekoek and R. F. Swarttouw, "The Askey-scheme of hypergeometric orthogonal polynomials and its q-analogue," 1996, *arXiv:math/9602214*.
- [44] D. Giakoumis, D. Tzovaras, K. Moustakas, and G. Hassapis, "Automatic recognition of boredom in video games using novel biosignal moment-based features," *IEEE Trans. Affect. Comput.*, vol. 2, no. 3, pp. 119–133, Jul. 2011.
- [45] O. Banos, J.-M. Galvez, M. Damas, H. Pomares, and I. Rojas, "Window size impact in human activity recognition," *Sensors*, vol. 14, no. 4, pp. 6474–6499, Apr. 2014.
- [46] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995.
- [47] T. Cover and P. Hart, "Nearest neighbor pattern classification," *IEEE Trans. Inf. Theory*, vol. IT-13, no. 1, pp. 21–27, Jan. 1967.
- [48] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.
- [49] H. He, Y. Bai, E. A. Garcia, and S. Li, "ADASYN: Adaptive synthetic sampling approach for imbalanced learning," in *Proc. IEEE Int. Joint Conf. Neural Netw.*, Jun. 2008, pp. 1322–1328.
- [50] H. Abdi, "Z-scores," *Encyclopedia Meas. Statist.*, vol. 3, pp. 1055–1058, Jan. 2007.

- [51] L. Shu et al., "A review of emotion recognition using physiological signals," *Sensors*, vol. 18, no. 7, p. 2074, Jun. 2018.
- [52] J. A. Russell, "A circumplex model of affect," *J. Personality Social Psychol.*, vol. 39, no. 6, p. 1161, Dec. 1980.
- [53] M. P. Janisse, "Pupil size, affect, and exposure frequency," *Social Behav. Personality, Int. J.*, vol. 2, no. 2, pp. 125–146, Jan. 1974.



Rong Ma received the B.S., M.S., and Ph.D. degrees from Lanzhou University, Lanzhou, China, in 2004, 2007, and 2017, respectively.

She is currently an Assistant Professor with Lanzhou University. Her current research interests include digital image processing, computer vision, and neural network.



Minqiang Yang (Member, IEEE) received the Ph.D. degree in computer science from Lanzhou University, Lanzhou, China, in 2022.

He is currently an Engineer with the Gansu Provincial Key Laboratory of Wearable Computing, School of Information Science and Engineering, Lanzhou University. His current research interests include affective computing, image processing, machine learning, and automatic depression detection.



Xiongying Li received the master's degree in professional development and educational psychology from Northwest Normal University, Lanzhou, in 2003, and the Ph.D. degree in higher education from Xiamen University, Xiamen, China, in 2013.

He is currently a Professor with the Institute of Higher Education, Lanzhou University, Lanzhou, a Registered Psychologist, and the Director of the Mental Health Education and Counseling Center, Lanzhou University. His research interests include college student learning and development, adolescent mental health education, family education, and psychological counseling.



Xiang Feng is currently pursuing the B.E. degree in computer science and technology with the School of Information Science and Engineering, Lanzhou University, Lanzhou, China.

His current major research interests include multimodal fusion, affective computing, and computer vision.



Chengsheng Mao (Member, IEEE) received the B.S. degree in computer science and technology from the Huazhong University of Science and Technology, Wuhan, China, in 2008, and the Ph.D. degree in computer application technology from Lanzhou University, Lanzhou, China, in 2016.

He is currently a Research Assistant Professor with the Department of Preventive Medicine, Feinberg School of Medicine, Northwestern University, Evanston, IL, USA. His research interests include data mining and machine learning in a range of domains including medical informatics, computer vision, and natural language processing.