IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received December 26, 2018, accepted January 10, 2019, date of publication January 30, 2019, date of current version March 5, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2896303

An Accurate and Robust Gaze Estimation Method Based on Maximum Correntropy Criterion

BEN YANG, XUETAO ZHANG, ZHONGCHANG LI, SHAOYI DU, AND FEI WANG
National Engineering Laboratory for Visual Information Processing and Applications, Xi'an Jiaotong University, Xi'an 710049, China
School of Electronic and Information Engineering, Xi'an Jiaotong University, Xi'an 710049, China

Corresponding author: Xuetao Zhang (xuetaozh@xjtu.edu.cn)

This work was supported in part by the National Key Research and Development Program of China under Grant 2017YFC0803905, and in part by the Fundamental Research Funds for the Central Universities.

ABSTRACT Accurately estimating the user's gaze is important in many applications, such as human–computer interaction. Due to great convenience, appearance-based methods for gaze estimation have been a popular subject of research for many years. However, the greatest challenges in the appearance-based gaze estimation in a desktop environment are how to simplify the calibration process and deal with other issues such as image noise and low resolution. To address the problems, we adopt a mapping relationship between the high-dimensional eye image features space and the low-dimensional gaze positions and propose a robust and accurate method for gaze estimation with a webcam. First, we utilize Kullback–Leibler divergence to reduce feature dimension and keep similarity between the feature space and the gaze space. Then, we construct the objective function using the maximum correntropy criterion instead of mean squared error, which can enhance the anti-noise ability, especially for outliers or pixel corruption. A regularization term is adopted to adaptively select the sparse training samples for gaze estimation. We conducted extensive experiments in a desktop environment, which verified that the proposed method was robust and efficient in dealing with sparse training samples, pixel corruption, and low-resolution problems in gaze estimation.

INDEX TERMS Appearance-based method, human computer interaction, gaze estimation, maximum correntropy criterion.

I. INTRODUCTION

Gaze estimation is the process of detecting the location where a person is fixating on or estimating the direction of the 3D visual axis of the eye. Much work can be accomplished by knowing the gaze of the eye. For instance, [1] by utilizing gaze to select a target, people with disabilities are able to type by looking at an on-screen keyboard. Paravati and Gatteschi [2] summarized the potential applications in developing smart environments.

To avoid the inconvenience of intrusive gaze trackers, non-intrusive gaze estimation methods have been developed based on computer vision technology. As surveyed in [3], these kinds of methods can be mainly divided into two categories: model-based methods and appearance-based methods.

Model-based methods utilize the geometric relationship of the eye model and the environment to estimate the 3D visual axis of the subjects. However, to achieve a high accuracy, stereo camera pairs, infrared light sources, and other hardware [4]–[7] are needed for the accurate extraction of the 3D eye features to determine the eye model and the gaze direction. The most common 3D eye features include the pupil center and corneal infrared [5]–[8]. After obtaining the visual axis from the 3D eye features, the gaze point on the screen can be obtained by intersecting the visual axis with the screen [9]. However, the process is complicated and difficult to calibrate.

Unlike model-based methods, appearance-based methods are simple (using only the eye image as a high dimensional input feature to map onto the gaze coordinates), and a single web camera is sufficient to set up the experiment. The latest methods, such as the saliency [10], human gaze patterns [11], and Convolutional Neural Network [12], [13], are proposed. However, these methods need many training samples to learn the model and they also cannot handle outliers or noise in the data. Our method assumes a fixed head pose like most appearance-based methods, which is based on the appearance manifold interpolation for gaze estimation. Tan *et al.* [14] used local linearity based on the appearance-manifold to

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

interpolate the unknown gaze points using 252 training samples. Although this method could obtain an accurate estimation, a large amount of training samples were needed, and this method did not take into consideration image noise. Williams *et al.* [15] introduced sparsity, and learned the mapping from a few semi-supervised training sets through a semi-supervised Gaussian Process. This method achieved a better result by using a few training samples, but this method did not consider noise or low-resolution problems. Lu *et al.* [16], [17] used adaptive-linear regression for gaze estimation by using fewer training samples. This method was based on Mean Squared Error (MSE), and they only needed about 33 samples to obtain excellent accuracy results considering that the noise was light and it was a Gaussian distribution. When the noise was heavy and a non-Gaussian distribution was used, this method could ensure high accuracy.

Appearance manifold interpolation mainly assumes that the appearance on the manifold can be approximated by linear combinations of neighbor samples. This idea is often used to estimate the gaze positions. However, these methods have many challenges, such as the assumption of similarity between human eye image manifold and the gaze feature space, and the need of many training samples and some disturbance problems. To deal with above challenges, in this paper we propose an accurate and robust gaze estimation method based on correntropy. It utilizes eye appearance features to predict the gaze position. After extracting and aligning eye images, a Kullback-Leibler divergence (KLD) -based feature dimension reduction is used to effectively improve the similarity of the human eye image manifold and the gaze feature space. For the reconstruction of the test samples, this paper proposes a gaze estimation method based on the correntropy algorithm due to the large number of training samples and image corrosion problems. This method can select the sparse and local training samples to represent the new eye feature and enhance the anti-noise performance. In the experiment, we tested the effectiveness of the proposed method by using sparse training samples. The main contributions of the paper can be summarized as follows:

1) In this paper, we firstly introduce the MCC into appearance based gaze estimation field. Comparing with the existing methods, which utilize MSE as optimization objective, the proposed gaze estimation method can efficiently cope with noisy and low resolution images.
2) For the purpose of improving the accuracy of gaze estimation, we also introduce a local linear constrain to make the selected samples close to the input image. Meanwhile, a KLD-based feature dimension reduction is used to improve the similarity between human eye image feature manifold and gaze space.

II. THE OVERVIEW OF PROPOSED ALGORITHM

A. EYE APPEARANCE MANIFOLD INTERPOLATION FOR GAZE ESTIMATION

Consider the features of N samples in eye images, and $E = \{\mathbf{e}_1, \mathbf{e}_2 \ldots, \mathbf{e}_N\} \in R^{m\times N}$ constitutes a manifold in the m-D space. Let $G = \{\mathbf{g}_1, \mathbf{g}_2, \ldots, \mathbf{g}_N\} \in R^{2\times N}$ denote the set of gaze positions on the screen, corresponding with eye images in E. The goal is to learn the mapping between the eye appearance features and the gaze positions.

The eye appearance feature manifold is assumed to be continuous and smooth, and one of the features can be interpolated by using some of its neighbors in the manifold. In [18], a spline interpolation-based construction approach was proposed for reconstructing the new sample by utilizing its neighbors. In this paper, we assumed a linear relationship. Therefore, given a new eye image, we estimate its feature $\hat{\mathbf{e}}$ by combining the training features in E as follows:

$$\hat{\mathbf{e}} = \sum_i w_i \mathbf{e}_i \tag{1}$$

where w_i is the weight corresponding with the ith sample.

In the literature, several works have assumed a similar structure between eye appearance feature manifold and gaze position space [14], [17], [19]. Thus, we can obtain the new corresponding gaze position by training gaze positions using the same weights in the feature manifold. That is, a new gaze position $\hat{\mathbf{g}}$ can be constructed as:

$$\hat{\mathbf{g}} = \sum_i w_i \mathbf{g}_i \tag{2}$$

As mentioned in [16], the above assumption is valid by assuming locality. In this paper, we also handle this with the idea of sparse training samples.

B. THE FRAMEWORK OF PROPOSED ALGORITHM

The proposed algorithm is based on the principle introduced in Section II-A. However, we improve the accuracy and the robustness of it in two different aspects. Firstly, we introduce a Kullback-Leibler divergence based optimization algorithm to learn a transformation. This can reduce feature dimension while enhancing the similarity between the feature projection space and gaze space. Secondly, we adopt a maximum correntropy criterion induced cost function, which can greatly improve the robustness and accuracy.

The entire process of the proposed algorithm is shown in Figure 1. Firstly, We use the AdaBoost algorithm along with Haar-Like features to detect the subject's face. This algorithm is very efficient and has high detection accuracy. Of course, it is also possible to use the other popular face detection methods. Secondly, the facial landmarks are located utilizing a constrained local model framework [20]. After locating the eye corner position, we segment the eye region with a fixed aspect ratio. Thirdly, in the feature extraction and reduction step, the eye image is divided into several small cells. We obtain a 1-D histogram of gradient orientations for each cell. By stacking the vectors of all cells into a column vector, we can obtain a vector as the eye appearance feature. To enhance the similarity and reduce computation cost, we introduce a KLD-based method which can improve the similarity efficiently between human eye feature space and gaze space via mapping human eye feature space to a

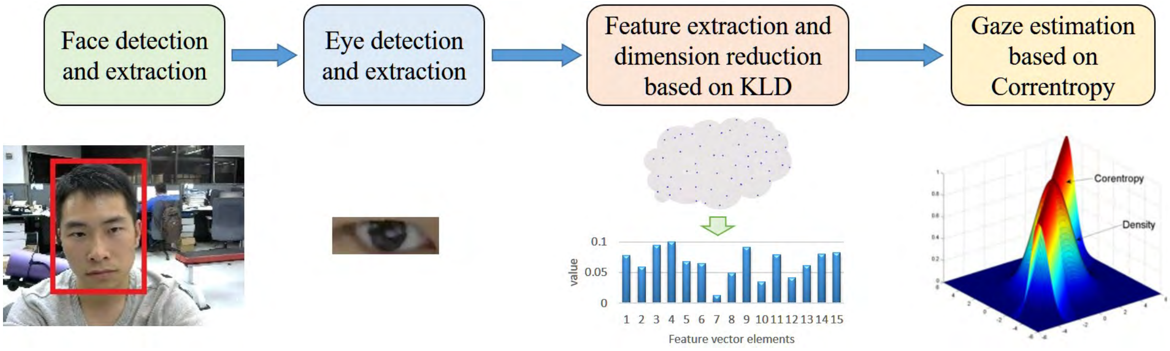

Face detection and extraction
Eye detection and extraction
Feature extraction and dimension reduction based on KLD
Gaze estimation based on Correntropy
0.1
0.05
0
value
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
Feature vector elements
Corentropy
Density

FIGURE 1. **The framework of proposed algorithm.**

feature projection space. The experiment results indicate that the KLD-based method can not only increase the accuracy of the result of experiment, but also can improve the estimation efficiency greatly. And then, we design a novel objective function based on MCC to obtain an accurate gaze estimation. Compared with the traditional MSE based method, MCC greatly increases the robustness of the overall algorithm due to its reliability in processing abnormal data, so that it can accurately and stably estimate the fixation position. Finally, in the stage of gaze coordinate estimation, we adopted the weight sharing framework in II-A, and finally reconstructed the gaze position through the training samples to complete the whole process of gaze estimation.

III. HUMAN EYE FEATURE DIMENSION REDUCTION BASED ON KL DIVERGENCE

The estimation accuracy and the similarity between the eye feature space and the gaze space are related because we use the method of weight sharing to estimate the gaze position. In order to improve the accuracy of estimation and reduce the estimation time, the proposed algorithm introduces feature space dimension reduction, which improves the similarity between feature space and gaze point space, which greatly reduces the computation time and improves the accuracy of the estimation algorithm. Compared with the traditional dimensionality reduction algorithm Principal Component Analysis (PCA) [21], we introduce a KLD-based dimensionality reduction method, which reduces the feature space dimension and ensures the similarity with the gaze space to improve the estimation accuracy. As shown in the Figure 2, the KDL-based feature space reduction results are significantly better than the PCA-based dimensionality reduction results. As we know, KLD as a distance measure is usually used to measure the distance between two probability distributions and some optimize process. In a certain range, KLD minimum value represents corresponding parameters are optimal. Compared with KLD, other distance measures, such as the Total variation distance and Bhattacharyya distance, have some problems. Such as the Total variation distance is mainly used to remove noise, which will change small gradient regions to constant. Bhattacharyya distance is mainly used to measure the separability between clusters, while the high dimension calculation of the cost would be larger. Therefore, this paper adopts KLD-based feature space dimension reduction, which can improve the accuracy of estimation while ensuring the efficiency of calculation. It is assumed that variable $X = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ has the projection transformation function $f(\mathbf{x})$ in the feature space, the distance of X between any two variables $\mathbf{x}_1, \mathbf{x}_2 \in X$ in feature space is $d[\mathbf{x}_1, \mathbf{x}_2]$, and the distance between any two variables is $d[f(\mathbf{x}_1), f(\mathbf{x}_2)]$ in projection feature space. Assuming the projection transformation function $f(\mathbf{x})$ of $\mathbf{x} \in X$ is linear, this means $f(\mathbf{x}) = C\mathbf{x}$, where C is the projection conversion matrix. Therefore, the Euclidean distance between the features $f(\mathbf{x}_1)$ and $f(\mathbf{x}_2)$ in projection feature space is $||f(\mathbf{x}_1) - f(\mathbf{x}_2)||_2 = (\mathbf{x}_1 - \mathbf{x}_2)^T A(\mathbf{x}_1 - \mathbf{x}_2)$, where $A = C^T CA = C^T C$ is positive semidefinite matrices, also known as distance matrices.

After extracting the HoG features from the eye images, let Ω, Ψ, Γ represent the eye feature space, the gaze point space, and the eye feature projection space, respectively. The function $D(\cdot, \cdot)$ represents the distance between two points in each spaces, while the measurement criterion between eye feature projection space and gaze space are assumed to be similar. That is: $D(C\mathbf{e}_i, C\mathbf{e}_j) \approx D(\mathbf{x}_i, \mathbf{x}_j)$, where C is the projection transformation matrix of the eye images. Then, one can establish the objective function between the dimensionality reduced space and the gaze space based on the distance measure criterion. The formula is as follows:

$$\min_{C} \mathcal{F}(\Omega, \Psi, \Gamma; C) \tag{3}$$

The main purpose of (3) is to obtain projection space Γ by performing the projection transformation matrix C on the human eye feature space Ω, so that the feature projection space Γ and the gaze space Ψ have a more similar structure.

In this paper, the Euclidean distance is used to represent the distance measurement of space, that is: $D(i, j) = ||i - j||_2$. The distance of the three spaces of the human eye feature space, the feature projection space, and the gaze space are respectively represented as follows:

$$D_\Omega(i,j) = (\mathbf{e}_i - \mathbf{e}_j)^T(\mathbf{e}_i - \mathbf{e}_j) \tag{4}$$

$$D_\Gamma(i,j) = (C\mathbf{e}_i - C\mathbf{e}_j)^T(C\mathbf{e}_i - C\mathbf{e}_j) \tag{5}$$

$$D_\Psi(i,j) = (\mathbf{x}_i - \mathbf{x}_j)^T(\mathbf{x}_i - \mathbf{x}_j) \tag{6}$$

Moreover, the distance in feature projection space can be converted into the following expression:

$$\begin{aligned} D_\Gamma(i,j) &= (C\mathbf{e}_i - C\mathbf{e}_j)^T(C\mathbf{e}_i - C\mathbf{e}_j) \\ &= (\mathbf{e}_i - \mathbf{e}_j)^T C^T C(\mathbf{e}_i - \mathbf{e}_j) \\ &= (\mathbf{e}_i - \mathbf{e}_j)^T A(\mathbf{e}_i - \mathbf{e}_j) \end{aligned} \tag{7}$$

where $A = C^T C$ is positive semidefinite matrix (PSD).

Common distance measurement learning algorithms, such as Principal Component Analysis (PCA) [21] and Linear Discriminant Analysis (LDA) [22], make use of the global covariance structure and cannot guarantee the local similarity of the structure between features projection space Γ and gaze space Ψ. Therefore, to guarantee the accuracy of the distance measurement, this paper uses the KLD based learning method. To guarantee the similarity of spatial structure between feature projection space and gaze space, this method learns the metric matrix based on the structure of the target space to obtain the distance matrix A.

A. KULLBACK-LEIBLER DIVERGENCE

In information theory and probability theory, the KLD [23] is an asymmetric metric formula for two probability distributions P and Q. It is mainly used to describe the difference between two distributions. Therefore, the KLD expression from probability distribution Q to probability distribution P is $D_{KL}(P||Q)$, and the KL divergence expression is as follows:

$$D_{KL}(P||Q) = \Sigma P(i) log \frac{P(i)}{Q(i)} \tag{8}$$

$D_{KL}(P||Q)$ means P on the KL divergence of Q, where P is real probability distribution and Q is the approximate probability distribution of P. When the value of $D_{KL}(P||Q)$ is smaller, the function shows that the probability distribution P is more similar to the approximate probability distribution Q. On the contrary, when the value is larger, there is a greater difference between them.

B. OBJECTIVE FUNCTION AND OPTIMIZATION

For the eye image feature space and the gaze space structure, the purpose of dimension reduction of the human eye feature space is projected so that the projected space Γ and the gaze space Ψ are more similar. To learn the measurement better, the conditional distribution between the training samples is defined as follows:

$$P(j|i) = \begin{cases} \dfrac{e^{-D(i,j)}}{\sum_{k \neq i} e^{-D(i,k)}} & j \neq i \\ 0 & j = i \end{cases} \tag{9}$$

Therefore, the conditional distributions of the human eye projection space and the gaze space are:

$$P_\Gamma(j|i) = \begin{cases} \dfrac{e^{-D_\Gamma(i,j)}}{\sum_{k \neq i} \mathbf{e}^{-D_\Gamma(i,j)}} & j \neq i \\ 0 & j = i \end{cases} \tag{10}$$

$$P_\Psi(j|i) = \begin{cases} \dfrac{e^{-D_\Psi(i,j)}}{\sum_{k \neq i} e^{-D_\Psi(i,j)}} & j \neq i \\ 0 & j = i \end{cases} \tag{11}$$

The purpose of human eye dimension reduction is to make $P_\Gamma(j|i)$ as close as possible to $P_\Psi(j|i)$ by finding a positive semi-definite matrix A. Moreover, by using KLD, the objective function can be written as:

$$\begin{aligned} &\min_A f(A) = D_{KL}(P_\Psi(j|i)||P_\Gamma(j|i)) \\ &s.t.\ \ A \in PSD \end{aligned} \tag{12}$$

The function has a minimum value since the objective function (12) is a convex function. It can be solved using the mutual iteration method, by applying gradient descent and projection to the PSD cone method of the main function. The gradient expression of the objective function is as follows:

$$\nabla f(A) = \sum_{i,j} (P_\Psi(j|i) - P_\Gamma(j|i))(\mathbf{e}_i - \mathbf{e}_j)^T(\mathbf{e}_i - \mathbf{e}_j) \tag{13}$$

For the tth iteration, adjust A by step ζ, and the operation is as follows:

$$A_{t+1} = A_t - \zeta \nabla f(A) \tag{14}$$

To ensure that Matrix A is a positive semi-definite matrix, the solution process is performed by projecting Matrix A onto the PSD cone. The operation process is as follows:

First, the eigenvalue solution operation of matrix A is as follows:

$$A_{t+1} = \sum_k \lambda_k u_k u_k^T \tag{15}$$

where λ_k is eigenvalue of A, u_k is the corresponding eigenvector of λ_k.

Through the elimination of negative eigenvalue operation, we obtain

$$A_{t+1} = \sum_k max(\lambda_k, 0) u_k u_k^T \tag{16}$$

The mutual iteration method is used to perform the gradient operation and the projection operation on the above process until A converges.

Figure 2 provides a visualization of the eye image feature space Ω, gaze space Ψ, and the projected feature space Γ. Here, we use 36 points for calibration. Figure 2a shows the structure of the eye image feature space. We can see that its structure is chaotic and distributed in a disorderly manner. Figure 2b is the gaze space structure, and the image shows a good regularity. Figure 2c shows the structure of the projected feature space based on KLD, and Figure 2d shows the structure of the projected feature space based on PCA. It can be

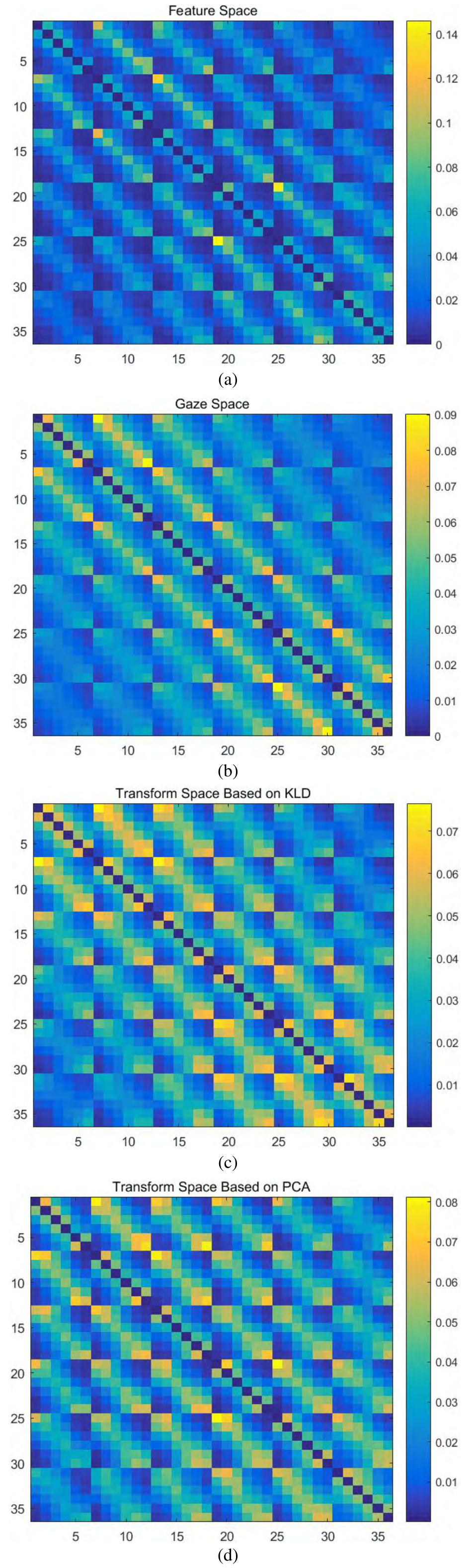


FIGURE 2. Three space structure diagram. In the figure, the abscissa and ordinate represent the number of points, and the right ordinate represents the distance between points. (a) Human eye feature space structure. (b) Gaze space structure. (c) Projection space based on KLD. (d) Projection space based on PCA.

clearly seen that after the linear projection of the eye image feature space Ω into the space Γ, the structure of the space Γ shows obvious periodicity and is noticeably similar to the structure of gaze space Ψ. And we can see obviously that the results of two dimension reduction methods from Figure 2c and Figure 2d. It sames that the method based on KLD is better than PCA.

IV. HUMAN EYES GAZE ESTIMATION BASED ON CORRENTROPY

A. CORRENTROPY

Correntropy was proposed in Information Theoretic Learning (ITL) [24] as a generalized similarity measurement [25]. It is always used to measure the similarity of the feature vectors. It can effectively deal with non-Gaussian noise [26]–[29]. For two arbitrary random variables $X = \{x_1, x_2, \ldots, x_N\}$ and $Y = \{y_1, y_2, \ldots, y_N\}$, their correlation is defined as follows:

$$\begin{aligned} V(X,Y) &= E\left[k_\sigma(X,Y)\right] \\ &= \iint k(x,y)p_{XY}(x,y)dxdy \end{aligned} \tag{17}$$

where $p_{XY}(x,y)$ is the joint PDF and $k(x,y)$ is any continuous, non-negative definite kernel which satisfies the Mercer theory [30]. If the kernel is a translation invariant kernel, such as Gaussian, then (17) becomes the correntropy.

In practice, if the joint PDF of $p_{XY}(x,y)$ is unknown and there are only a finite number of data points $\{(x_i,y_i)\}_{i=1}^{n}$, we can obtain the following simple estimate of correntropy:

$$V(X,Y) = \frac{1}{N}\sum_{i=1}^{N} k_\sigma(x_i,y_i) \tag{18}$$

where $k_\sigma(x_i,y_i) = \exp\left(-\frac{(x_i-y_i)^2}{2\sigma^2}\right)$

B. GAZE ESTIMATION BASED ON CORRENTROPY

We propose a gaze estimation method based on correntropy to further improve the accuracy and enhance the robustness of the gaze estimation. Compared to previous works, such as [14], [16], and [19], which are mainly based on MSE, our proposed method can enhance anti-noise performance and achieve more accurate estimation. Moreover, some recent papers have utilized l^1-Regularization to reconstruct the test sample by sparse representation. However, such methods cannot guarantee the locality of the selected training samples and this can introduce additional errors.

$$\begin{aligned} J = \max_{\mathbf{w}} \frac{1}{m}\sum_{j=1}^{m} \exp\left(-\frac{\left\|\sum_{i=1}^{N} w_i\mathbf{e}_i - \hat{\mathbf{e}}\right\|^2}{2\sigma^2}\right) \\ -\lambda\left\|\mathbf{d}^T\mathbf{w}\right\|^2 \\ s.t. \sum_{i=1}^{N} w_i = 1 \end{aligned} \tag{19}$$

Here, $E \in R^{m\times N}$ is the training sample set of human eye features, $\hat{e}$ is the test sample, δ is the bandwidth of Gaussian kernel, $\mathbf{w}$ is the weight of training samples, and $\mathbf{d}$ denotes the distance between the test sample $\hat{e}$ and the ith training samples in E, which can be calculated as $d_i = \exp(\frac{\|\hat{e}-e_i\|^2}{\rho})$. The second term in (19) is a constraint, which can reduce the weights of further samples. That means the constraint can stay away from selecting samples far from the testing sample, which breaks the local smooth assumption. Form (19) we can get the algorithm complexity of the algorithm we proposed is O(mN).

In this paper, we use a half-quadratic technique to solve the regularized entropy maximization problem, such as [31] and [32]. According to the property of convex conjugate function [33], we have:

For $G(x,\sigma) = \exp(-\frac{\|x\|^2}{2\sigma^2})$, there exists a convex conjugated function called ψ, so we can obtain:

$$G(x,\sigma)\max_{p}\left(p\frac{\|x\|^2}{\sigma^2} - \psi(p)\right) \tag{20}$$

where $p = \{p_1, p_2, \ldots, p_m\}$ are the auxiliary variables in half-quadratic optimization.

For a fixed x, the maximum is reached at $p = -G(x,\sigma)$ [33]. By substituting (20) into the (19), we can obtain the augmented function:

$$\begin{aligned}\hat{J}(w,p) = \max_{\mathbf{w},p}&\left[\frac{1}{m}\sum_{j=1}^{m}\left(p_j\left(-\frac{\left\|\sum_{i=1}^{N} w_i\mathbf{e}_i - \hat{\mathbf{e}}\right\|^2}{2\delta^2}\right)\right.\right.\\ &\left.\left. - \psi(p_j)\right) - \lambda\left\|\mathbf{d}^T\mathbf{w}\right\|^2\right]\\ s.t.&\sum_{i=1}^{N} w_i = 1\end{aligned} \tag{21}$$

For a fixed $\mathbf{w}$, we can obtain the following equation:

$$J(\mathbf{w})\max_{p}\hat{J}(\mathbf{w},p) \tag{22}$$

which is the same to this equation:

$$\max J(w) = \max_{w,p}\hat{J}(w,p) \tag{23}$$

To solve (21), alternative maximization can be utilized as follows:

$$p^{\tau+1} = -G\left(\sum_{i=1}^{N} w_i\mathbf{e}_i - \hat{\mathbf{e}}\right) \tag{24}$$

$$\begin{aligned}\mathbf{w}^{\tau+1} = \arg\max_{\mathbf{w}}&\left[\left(\hat{\mathbf{e}} - E\mathbf{w}\right)^T diag(p)\left(\hat{\mathbf{e}} - E\mathbf{w}\right)\right.\\ &\left. - \lambda\left\|\mathbf{d}^T\mathbf{w}\right\|^2\right]\\ s.t.&\sum_{i=1}^{N} w_i = 1\end{aligned} \tag{25}$$

From the definition, we can learn that $p \leq 0$. By replacing p with $-p$, we can obtain the equivalent minimal problem in (25), as follows:

$$\begin{aligned}w^{\tau+1} = \arg\min_{w}&\left[\frac{1}{2}\mathbf{w}^T\tilde{E}^T\tilde{E}\mathbf{w} - \left(\tilde{E}^T\tilde{\mathbf{e}}\right)^T\mathbf{w}\right.\\ &\left. - \frac{1}{2}\lambda\left\|\mathbf{d}^T\mathbf{w}\right\|^2\right]\\ s.t.&\sum_{i=1}^{N} w_i = 1\end{aligned} \tag{26}$$

where, $\tilde{E} = Ediag\left(\sqrt{-p^{\tau+1}}\right)$ and $\tilde{\mathbf{e}} = Ediag\left(-p^{\tau+1}\right)\hat{\mathbf{e}}$. Equation (26) is a l^2-norm regularization problem, the analytic solution can be obtained as below:

$$\mathbf{w} = \left(\tilde{E}^T - \tilde{E} - \lambda diag(\mathbf{d})^2\right)^{-1}\tilde{E}^T\tilde{\mathbf{e}} \tag{27}$$

The entire procedure is summarized in Algorithm 1.

Algorithm 1 The Procedure of Gaze Estimation Based on Correntropy

```
Require:
    The training eye features, $E$;
    The testing eye feature, $\hat{\mathbf{e}}$;
    The training gaze positions, $G$;
Ensure:
    Gaze position, $\hat{\mathbf{g}}$;
    Initialize the weight vector $\mathbf{w} = E^T\hat{\mathbf{e}}$;
    Update the weight vector $\mathbf{w}$ and the auxiliary variables $p$;

    for $\tau = 1$ to $n$ do
        Update the auxiliary variables $p$ by ((24));
        Update the weight vector by ((27));
        if $\left|J^{\tau+1} - J^{\tau}\right| \leq \epsilon$ then
            break;
        end if
    end for
    Calculate the gaze positions by $\hat{\mathbf{g}} = \sum_i w_i\mathbf{g}_i$.
```

V. EXPERIMENT RESULTS

In this section, we verify the effectiveness of our algorithm compared with other classical algorithms. We set different calibration modes and the number of calibration points were 9, 16, 25, and 36. We will begin by introducing the way we collected the experimental data. Then, the experimental process can be divided into three separate parts: a comparison of algorithms under different calibration patterns, a comparison of different algorithms under pixel corrosion, and a comparison of different algorithms under low-resolution conditions.

A. DATASET COLLECTION

As shown in the Figure 3 , during the data collection process, the distance between the tester and the screen was nearly 70 cm and the screen size was 52cm(H)×29cm(V).

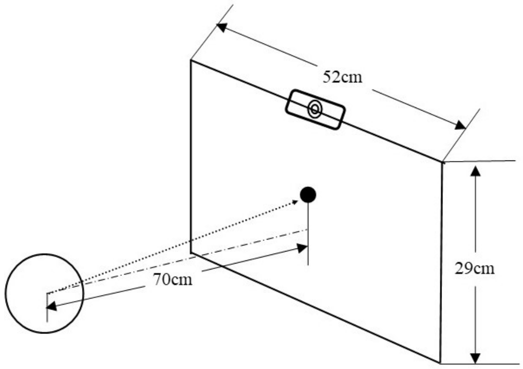


FIGURE 3. The experiment setup.

To analyse the effect of the number of calibration points, four calibration data sets were collected, respectively. The number of calibration points were 9, 16, 25, and 36, and the calibration patterns were as shown in Figure 4. The head images and corresponding gaze coordinate information were saved in the collection process.

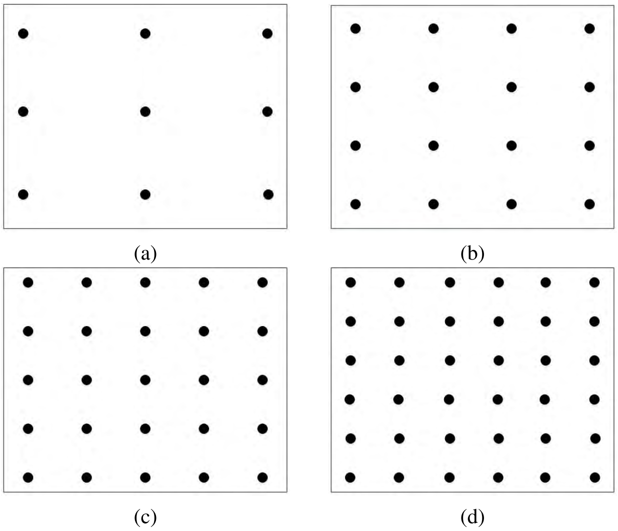


FIGURE 4. Human eye image of random corrosion.

For testing the dataset under the condition of a fixed head posture, the data was collected by a standard camera. During the collection process, a single bright spot was randomly displayed on the screen once, and the corresponding face image was captured. A total of 200 samples were collected.

For the experimental data of the pixel corrupted human eye image, a pixel corrosion operation was performed on each test image in the process of designing the dataset. The pixel value of each test human eye image was determined by a value belonging to a uniform distribution and the value was between [0 - 255], where the value and position of the replaced pixel were both random. The percentage of corrosion was set from between10% to 50% to verify the effectiveness of the proposed method during the experiment, as shown in Figure 5. Figure 5a shows the original image; Figure 5b shows the image at 10% corrosion; Figure 5c shows the image at 30% corrosion; and Figure 5d shows the image at 50% corrosion.

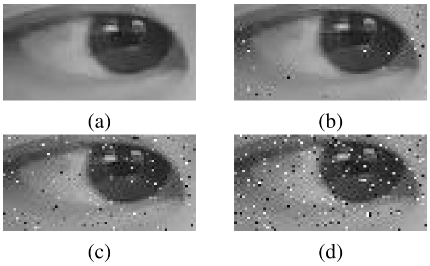


FIGURE 5. Human eye image of random corrosion.

During the experiment, the pixel scaling operation was performed on each test image using the same method to collect data in the case of low resolution. For each eye image being tested, the magnification was reduced to $\frac{1}{k}$, and then magnified to the original size. All the samples underwent the same treatment, so that could obtain low-resolution images. The scale value $\frac{1}{k}$ was in the range of [1 - 5], and the results are shown in Figure 6. Figure 6a shows the original image; Figure 6b shows the human eye image with 10% zoom; Figure 6c shows the human eye image with 30% zoom; and Figure 6d shows the human eye image with 50% zoom.

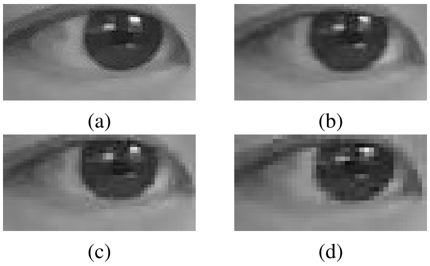


FIGURE 6. Human eye images at different scales.

During the experiment, we compared our method with several classic algorithms. These were Adaptive Linear Regression (ALR) proposed by Lu *et al.* [17], the local region method (Local Region) proposed by Tan *et al.* [14], Support Vector Regression (SVR) based on multi-scale HoG feature extraction (SVR+HoG) [34], the Local Linear Constraint method (LLC) proposed by Wang *et al.* [35], and Synthesis-based Low-cost gaze analyse (SLC) algorithm proposed by Chang *et al.* [36].

B. EXPERIMENTAL RESULTS AND ANALYSIS UNDER DIFFERENT CALIBRATION PATTERNS

We conducted the following comparative experiments on the feature dimension reduction problem. As shown in Table 1. In different calibration modes, adding KLD-based feature dimensionality reduction algorithm is significantly better than that without KLD. Therefore, in the process of gaze

TABLE 1. Comparison of the results of algorithms with KLD or not under different calibration modes.

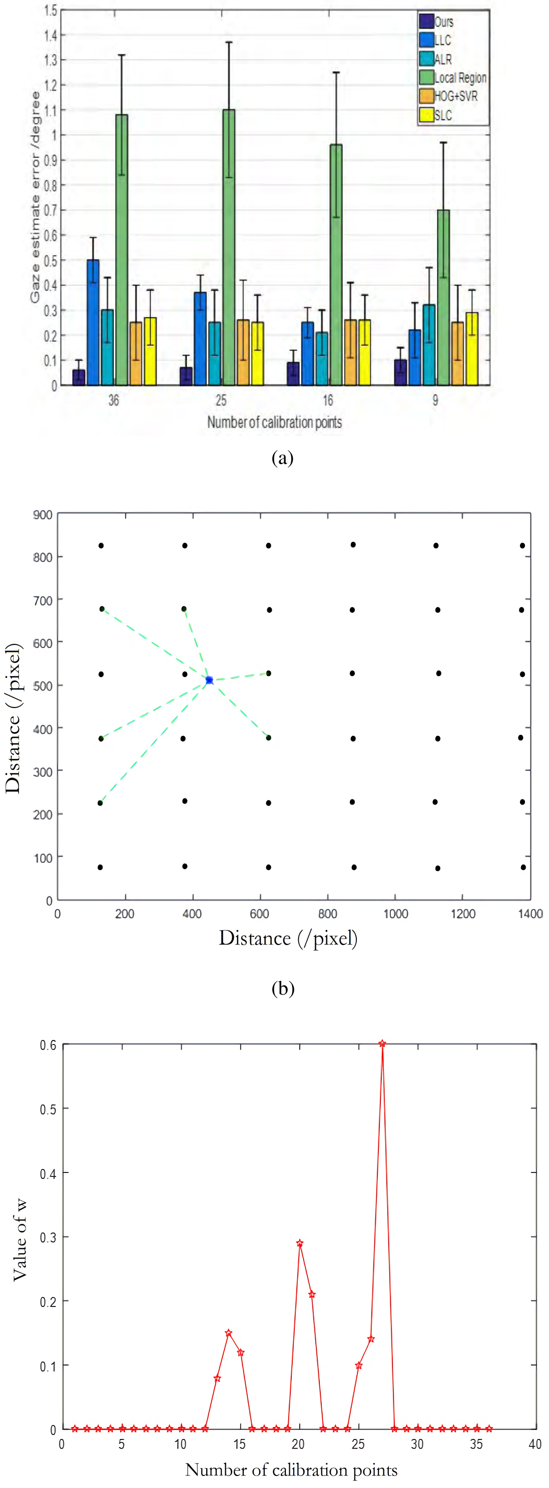
With KLD or not	36	25	16	9
with KLD	0.05°	0.07°	0.08°	0.10°
without KLD	0.16°	0.18°	0.16°	0.15°

point estimation, it is necessary to introduce the dimension reduction based on KLD features.

We tested the accuracy of gaze estimation with different calibration patterns. The four patterns used are shown in Figure 4. We compared the experimental results between the proposed method and five other methods (LLC, ALR, Local Region, SVR and SLC) using the same dataset. Table 2 gives the estimation error of different people and the overall average error.

In Figure 7a, the histogram shows more vividly a comparison of the results of the different methods. In this paper, the proposed method achieved the best estimation accuracy in all comparison methods. The estimation error was less than 0.1 degrees. The estimation errors for the other methods were all greater than 0.2 degrees. For the proposed method, as the data collected contained noise pollution, the samples were selected to be similar to the training samples in the process of reconstruction by using the correntropy advantage and the locality of the enhancing data, as shown in Figure 7b. When the distance between the test image and the training sample was large, the weight became correspondingly smaller. In practice, the weight was reset to zero if the weight was less than a certain value, as shown in Figure 7c, and the noise in the image under the correntropy framework could be effectively solved. Although the LLC method also uses the locality of the data, the Gaussian kernel function was used in this method, which can give a high-dimensional projection of the data. Under noisy conditions, the proposed method in this paper had better recognition ability. Although ALR uses the coefficient regularization term and could achieve good estimation accuracy under the condition of fewer training samples, the experimental effect was obviously lower than the algorithm in this paper due to the data collection being contaminated by noise. The Local Region method used all the training samples, which resulted in long consumption time and it ignored the differences in the data. In summary, the proposed method could select different training samples to reconstruct the test images. It could achieve good experimental accuracy by using the locality between the enhancing data.

We also compared the computational time of all the algorithms. The codes were written in MatLab on a 3.60GHz Core i7 CPU with Windows 10 system machine. Table 2 shows the computational times of different algorithms in different calibration modes. We can see that although the computational time of the algorithm we proposed is slightly larger than some comparison algorithms in some calibration modes, it can be regarded as a higher operation rate also. If we combine Table 2 and Table 2, we can see that the algorithm we proposed have the best performance, and it can greatly improve the estimation accuracy under the condition of ensuring the operation speed.



(a)

(b)

(c)

FIGURE 7. Graphical illustration of result of different calibration pattern experiment. (a) Comparison between different algorithm. (b) The selected samples by the proposed algorithm. (c) Sample weights obtained by the proposed algorithm.

TABLE 2. Comparison of the results of different algorithms under different calibration modes.

Sample number	Methods	Subject 1(/°)	Subject 2(/°)	Subject 3(/°)	Subject 4(/°)	Subject 5(/°)	Average error(/°)
36	*LLC*	0.46 ± 0.15	0.41 ± 0.15	0.55 ± 0.16	0.56 ± 0.16	0.53 ± 0.16	0.50 ± 0.09
	ALR	0.42 ± 0.13	0.26 ± 0.18	0.43 ± 0.15	0.22 ± 0.10	0.19 ± 0.10	0.30 ± 0.13
	LocalRegion	1.00 ± 0.32	0.85 ± 0.23	1.28 ± 0.22	1.17 ± 0.15	1.10 ± 0.28	1.08 ± 0.24
	HoG + SVR	0.25 ± 0.16	0.25 ± 0.14	0.25 ± 0.16	0.26 ± 0.13	0.24 ± 0.15	0.25 ± 0.15
	SLC	0.26 ± 0.12	0.25 ± 0.12	0.29 ± 0.09	0.31 ± 0.13	0.22 ± 0.10	0.27 ± 0.11
	Ours	**0.05 ± 0.03**	**0.10 ± 0.07**	**0.05 ± 0.05**	**0.07 ± 0.04**	**0.05 ± 0.05**	**0.06 ± 0.04**
25	*LLC*	0.35 ± 0.12	0.31 ± 0.11	0.39 ± 0.16	0.41 ± 0.18	0.40 ± 0.12	0.37 ± 0.07
	ALR	0.26 ± 0.11	0.26 ± 0.17	0.26 ± 0.12	0.21 ± 0.11	0.24 ± 0.15	0.25 ± 0.13
	LocalRegion	1.01 ± 0.28	0.96 ± 0.25	1.16 ± 0.28	1.16 ± 0.28	1.17 ± 0.28	1.10 ± 0.27
	HoG + SVR	0.26 ± 0.16	0.23 ± 0.16	0.27 ± 0.17	0.29 ± 0.12	0.27 ± 0.17	0.26 ± 0.16
	SLC	0.22 ± 0.12	0.24 ± 0.12	0.25 ± 0.09	0.29 ± 0.11	0.23 ± 0.09	0.25 ± 0.11
	Ours	**0.05 ± 0.04**	**0.11 ± 0.07**	**0.07 ± 0.05**	**0.07 ± 0.04**	**0.04 ± 0.04**	**0.07 ± 0.05**
16	*LLC*	0.24 ± 0.08	0.23 ± 0.09	0.26 ± 0.12	0.28 ± 0.10	0.25 ± 0.08	0.25 ± 0.06
	ALR	0.21 ± 0.09	0.23 ± 0.09	0.23 ± 0.10	0.18 ± 0.09	0.19 ± 0.10	0.21 ± 0.09
	LocalRegion	0.99 ± 0.24	0.61 ± 0.31	1.05 ± 0.30	1.04 ± 0.30	1.11 ± 0.31	0.96 ± 0.29
	HoG + SVR	0.26 ± 0.18	0.26 ± 0.14	0.28 ± 0.18	0.29 ± 0.10	0.23 ± 0.15	0.26 ± 0.15
	SLC	0.26 ± 0.10	0.24 ± 0.12	0.27 ± 0.08	0.29 ± 0.11	0.23 ± 0.09	0.26 ± 0.10
	Ours	**0.04 ± 0.03**	**0.16 ± 0.07**	**0.08 ± 0.04**	**0.09 ± 0.05**	**0.06 ± 0.05**	**0.09 ± 0.05**
9	*LLC*	0.21 ± 0.11	0.24 ± 0.11	0.22 ± 0.10	0.23 ± 0.15	0.22 ± 0.10	0.22 ± 0.11
	ALR	0.52 ± 0.22	0.28 ± 0.13	0.28 ± 0.13	0.26 ± 0.13	0.27 ± 0.13	0.32 ± 0.15
	LocalRegion	0.31 ± 0.12	0.51 ± 0.28	0.88 ± 0.33	0.88 ± 0.34	0.94 ± 0.28	0.70 ± 0.27
	HoG + SVR	0.28 ± 0.16	0.24 ± 0.13	0.27 ± 0.15	0.24 ± 0.15	0.22 ± 0.14	0.25 ± 0.15
	SLC	0.29 ± 0.09	0.26 ± 0.11	0.32 ± 0.08	0.33 ± 0.10	0.26 ± 0.08	0.29 ± 0.09
	Ours	**0.07 ± 0.04**	**0.15 ± 0.09**	**0.10 ± 0.24**	**0.07 ± 0.03**	**0.09 ± 0.05**	**0.10 ± 0.05**

TABLE 3. Comparison of the time of different algorithms under different calibration modes.

Sample Numbers	LLC	ALR	Local Region	HoG+SVR	SLC	**Ours**
36	69.67s	7.86s	12.11s	26.78s	13.51s	11.01s
25	6.38s	6.87s	11.34s	13.79s	12.39s	8.29s
16	6.13s	6.31s	10.91s	6.63s	11.57s	6.50s
9	5.92s	6.09s	10.61s	3.04s	11.05s	5.80s

TABLE 4. Comparison of the results of different algorithms under different corrosion rates.

Corrosion rates	LLC	ALR	Local Region	HoG+SVR	SLC	**Ours**
10%	1.03°	0.39°	1.25°	0.36°	0.19°	**0.19°**
20%	1.05°	0.43°	1.31°	0.35°	0.18°	**0.17°**
30%	1.01°	0.40°	1.42°	0.32°	0.18°	**0.18°**
40%	1.04°	0.41°	1.51°	0.38°	0.17°	**0.19°**
50%	1.08°	0.46°	1.56°	0.36°	0.18°	**0.17°**

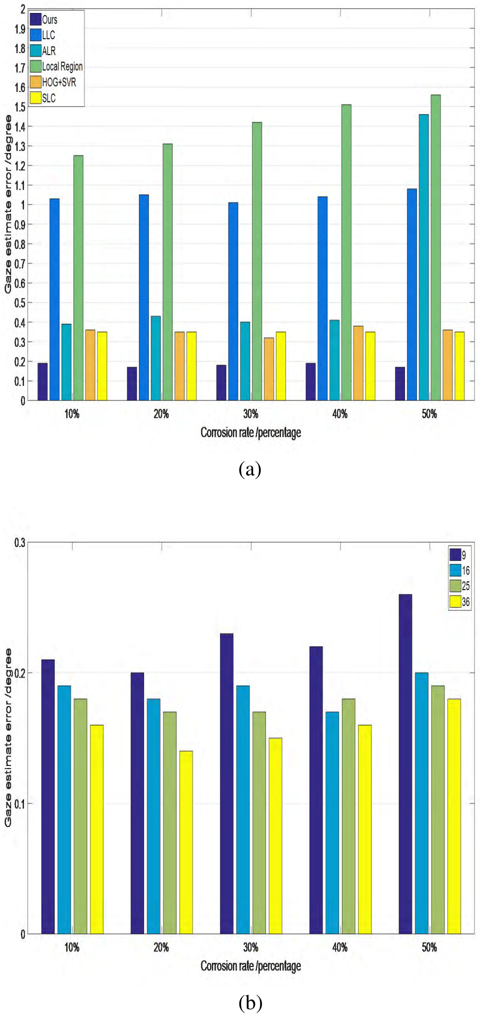
C. EXPERIMENTAL RESULTS AND ANALYSIS UNDER NOISY ENVIRONMENT

These experiments of human eye gaze estimation focuses on the pixel corrosion of the human eye image and compared with the results obtained by other methods. Two experiments were conducted. The first was a comparison of the results of different methods under different noise corrosion rates. The second was a comparison of the effectiveness of the proposed method under different training samples. The HoG feature was no longer used for the noise image feature. In this paper, all the images in the noise-containing human eye test sample were adjusted to a 15 × 30 pixel size, and the pixel value was directly extracted as the image feature and normalized, simultaneously. Finally, we obtained the characteristics of each image, which were the 450-dimensional eigenvectors.

The results of different methods under 36 training samples and under different noise erosion rates are shown in Table 3. Figure 8a is a comparison chart of the corresponding results in Table 3. The experimental results of the proposed algorithm with LLC, ALR, Local Region, HoG + SVR and SLC at the different corrosion rates are also shown in Figure 8a. From the results, we can see that the proposed algorithm and SLC have higher noise immunity than the other algorithms. Since the random forest itself has high noise immunity, the SLC has higher estimation accuracy under noise conditions. Under the framework of correntropy, we can see that the noise could be effectively processed as well as SLC. Compared with

TABLE 5. Comparison of the results of different calibration patterns under different corrosion rates.

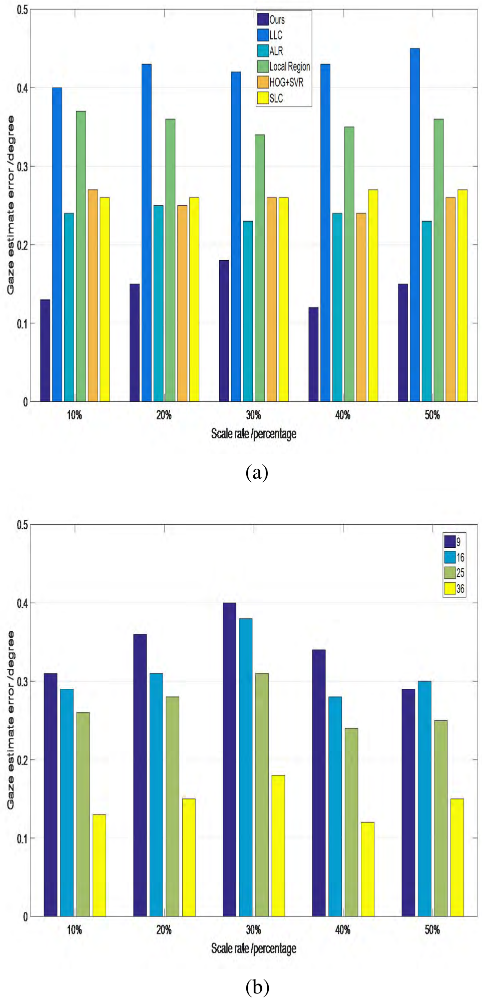
Corrosion rates	9	16	25	36
10%	0.21°	0.19°	0.18°	0.16°
20%	0.20°	0.18°	0.17°	0.14°
30%	0.23°	0.19°	0.17°	0.15°
40%	0.22°	0.17°	0.18°	0.16°
50%	0.26°	0.20°	0.19°	0.18°



(a)

(b)

FIGURE 8. Graphical illustration of results of noise experiment. (a) Comparison between different algorithm. (b) Comparison between different calibration patterns using the proposed algorithm.



(a)

(b)

FIGURE 9. Graphical illustration of results of resolution experiment. (a) Comparison between different algorithm. (b) Comparison between different calibration patterns using the proposed algorithm.

LLC, ALR, and Local Region algorithms, these algorithms were based on the MSE criterion to reconstruct the test samples, and the proposed algorithm had strong anti-noise performance.

Table 4 shows a comparison of the results with different pixel corrosion rates under different calibration mode conditions. Figure 8b is a comparison chart of the corresponding results in Table 4. As shown, the proposed method performed well at estimating the gaze position with a small number of training samples.

D. EXPERIMENTAL RESULTS AND ANALYSIS AT LOW RESOLUTION

The results of the different methods for one of the testers under 36 normal training samples at different scales are shown in Table 5. It can be seen from the table that each algorithm could effectively evaluate the gaze point at a low resolution, and that the low-resolution image belongs to the fuzzy image relative to the standard data. The proposed method had the best precision and showed the stability of the gaze estimation algorithm, as shown in Figure 9a.

TABLE 6. Comparison of the results of different algorithms at different scales.

Corrosion rates	LLC	ALR	Local Region	HoG+SVR	SLC	**Ours**
10%	1.03°	0.39°	1.25°	0.36°	0.19°	**0.19°**
20%	1.05°	0.43°	1.31°	0.35°	0.18°	**0.17°**
30%	1.01°	0.40°	1.42°	0.32°	0.18°	**0.18°**
40%	1.04°	0.41°	1.51°	0.38°	0.17°	**0.19°**
50%	1.08°	0.46°	1.56°	0.36°	0.18°	**0.17°**

TABLE 7. Comparison of the results of different calibration modes at different scales.

Corrosion rates	9	16	25	36
10%	0.21°	0.19°	0.18°	0.16°
20%	0.20°	0.18°	0.17°	0.14°
30%	0.23°	0.19°	0.17°	0.15°
40%	0.22°	0.17°	0.18°	0.16°
50%	0.26°	0.20°	0.19°	0.18°

Table 6 shows a comparison of the results of different scales under different calibration mode conditions. Under the condition of a small number of training samples, the proposed method could also estimate the fixation point. In addition, as the amount of training increased, the experimental precision was higher, as shown in Figure 9b.

VI. CONCLUSION

This paper introduces gaze estimation under the correntropy algorithm. The implementation process optimizes the problems that existed in the process of gaze estimation such as the sub-pixel extraction and alignment of the human eye image, the dimension reduction of the human eye feature space, excessive training samples, and noise interference. Improvement of the local neighborhood constraint of the objective function was based on the relevant entropy algorithm. Our improved method can effectively ensure that the calibration function can obtain good accuracy, even with a small number of calibration points. Finally, the performance of this algorithm was verified through multiple experiments.

REFERENCES

[1] R. A. Naqvi, M. Arsalan, and K. R. Park, "Fuzzy system-based target selection for a NIR camera-based gaze tracker," *Sensors*, vol. 17, no. 4, p. 864, 2017. [Online]. Available: http://www.mdpi.com/1424-8220/17/4/862

[2] G. Paravati and V. Gatteschi, "Human-computer interaction in smart environments," *Sensors*, vol. 15, no. 8, pp. 19487–19494, 2015. [Online]. Available: http://www.mdpi.com/1424-8220/15/8/19487

[3] D. W. Hansen and Q. Ji, "In the eye of the beholder: A survey of models for eyes and gaze," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 3, pp. 478–500, Mar. 2010.

[4] Z. Zhu, Q. Ji, and K. P. Bennett, "Nonlinear eye gaze mapping function estimation via support vector regression," in *Proc. 18th Int. Conf. Pattern Recognit. (ICPR)*, vol. 1, Aug. 2006, pp. 1132–1135.

[5] A. Villanueva and R. Cabeza, "A novel gaze estimation system with one calibration point," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 38, no. 4, pp. 1123–1138, Aug. 2008.

[6] E. D. Guestrin and M. Eizenman, "General theory of remote gaze estimation using the pupil center and corneal reflections," *IEEE Trans. Biomed. Eng.*, vol. 53, no. 6, pp. 1124–1133, Jun. 2006.

[7] E. D. Guestrin and M. Eizenman, "Remote point-of-gaze estimation requiring a single-point calibration for applications with infants," in *Proc. Eye Tracking Res. Appl. Symp. (ETRA)*, 2008, pp. 267–274.

[8] S. Du, J. Liu, Y. Liu, X. Zhang, and J. Xue, "Precise glasses detection algorithm for face with in-plane rotation," *Multimedia Syst.*, vol. 23, no. 3, pp. 293–302, Jun. 2017.

[9] D. Yoo and M. Chung, "A novel non-intrusive eye gaze estimation using cross-ratio under large head motion," *Comput. Vis. Image Understand.*, vol. 98, no. 1, pp. 25–51, Apr. 2005.

[10] Y. Sugano, Y. Matsushita, and Y. Sato, "Appearance-based gaze estimation using visual saliency," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 2, pp. 329–341, Feb. 2013.

[11] F. Alnajar, T. Gevers, R. Valenti, and S. Ghebreab, "Calibration-free gaze estimation using human gaze patterns," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2013, pp. 137–144.

[12] N. Liu, J. Han, D. Zhang, S. Wen, and T. Liu, "Predicting eye fixations using convolutional neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Boston, MA, USA, Jun. 2015, pp. 362–370.

[13] X. Zhang, Y. Sugano, M. Fritz, and A. Bulling, "Appearance-based gaze estimation in the wild," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 4511–4520.

[14] K.-H. Tan, D. J. Kriegman, and N. Ahuja, "Appearance-based eye gaze estimation," in *Proc. 6th IEEE Workshop Appl. Comput. Vis. (WACV)*, Dec. 2002, pp. 191–195.

[15] O. Williams, A. Blake, and R. Cipolla, "Sparse and semi-supervised visual mapping with the S^3GP," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 1, Jun. 2006, pp. 230–237.

[16] F. Lu, Y. Sugano, T. Okabe, and Y. Sato, "Inferring human gaze from appearance via adaptive linear regression," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Nov. 2011, pp. 153–160.

[17] F. Lu, Y. Sugano, T. Okabe, and Y. Sato, "Adaptive linear regression for appearance-based gaze estimation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 10, pp. 2033–2046, Oct. 2014.

[18] S. A. Nene and S. K. Nayar, "A simple algorithm for nearest neighbor search in high dimensions," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 9, pp. 989–1003, Sep. 1997.

[19] S. Roweis and L. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, Dec. 2000.

[20] T. Baltrušaitis, P. Robinson, and L.-P. Morency, "Continuous conditional neural fields for structured regression," in *Computer Vision—ECCV*, D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, Eds. Cham, Switzerland: Springer, 2014, pp. 593–608.

[21] S. Wold, K. Esbensen, and P. Geladi, "Principal component analysis," *Chemometrics Intell. Lab. Syst.*, vol. 2, nos. 1–3, pp. 37–52, 1987.

[22] R. A. Fisher, "The use of multiple measurements in taxonomic problems," *Ann. Eugenics*, vol. 7, no. 2, pp. 179–188, 1936.

[23] W. H. Highleyman, "Linear decision functions, with application to pattern recognition," *Proc. IRE*, vol. 50, no. 6, pp. 1501–1514, Jun. 1962.

[24] J. C. Principe, "Information theoretic learning," in *Proc. Int. Workshop Pattern Recognit. Inf. Syst., Conjunct (ICEIS)*, Miami, FL, USA, May 2009, pp. 1385–1392.

[25] I. Santamaria, P. P. Pokharel, and J. C. Principe, "Generalized correlation function: Definition, properties, and application to blind equalization," *IEEE Trans. Signal Process.*, vol. 54, no. 6, pp. 2187–2197, Jun. 2006.

[26] W. Liu, P. P. Pokharel, and J. C. Principe, "Correntropy: Properties and applications in non-Gaussian signal processing," *IEEE Trans. Signal Process.*, vol. 55, no. 11, pp. 5286–5298, Nov. 2007.

[27] B. Chen, J. Wang, H. Zhao, N. Zheng, and J. C. Príncipe, "Convergence of a fixed-point algorithm under maximum correntropy criterion," *IEEE Signal Process. Lett.*, vol. 22, no. 10, pp. 1723–1727, Oct. 2015.

[28] B. Chen, L. Xing, H. Zhao, N. Zheng, and J. C. Príncipe, "Generalized correntropy for robust adaptive filtering," *IEEE Trans. Signal Process.*, vol. 64, no. 13, pp. 3376–3387, Jul. 2016.

[29] B. Chen, X. Liu, H. Zhao, and J. C. Principe, "Maximum correntropy Kalman filter," *Automatica*, vol. 76, pp. 70–77, Feb. 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S000510981630396X

[30] V. Cherkassky, "The nature of statistical learning theory∼," *IEEE Trans. Neural Netw.*, vol. 8, no. 6, p. 1564, Nov. 1997.

[31] R. He, W.-S. Zheng, and B.-G. Hu, "Maximum correntropy criterion for robust face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 8, pp. 1561–1576, Aug. 2011.

[32] R. He, W.-S. Zheng, B.-G. Hu, and X.-W. Kong, "A regularized correntropy framework for robust pattern recognition," *Neural Comput.*, vol. 23, no. 8, pp. 2074–2100, 2011.

[33] V. Boyd and Faybusovich, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004, pp. 90–95.

[34] F. Martinez, A. Carbone, and E. Pissaloux, "Gaze estimation using local features and non-linear regression," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2013, pp. 1961–1964.

[35] J. Wang, J. Yang, K. Yu, F. Lv, T. Huang, and Y. Gong, "Locality-constrained linear coding for image classification," in *Proc. Comput. Vis. Pattern Recognit.*, 2010, pp. 3360–3367.

[36] Z. Chang, Q. Qiu, and G. Sapiro, "Synthesis-based low-cost gaze analysis," in *HCI International Posters' Extended Abstracts*, C. Stephanidis, Ed. Cham, Switzerland: Springer, 2016, pp. 95–100.

BEN YANG is currently pursuing the master's degree with the School of Electronic and Control Science and Engineering, Xi'an Jiaotong University, Xi'an, China. His current research interests include computer vision and machine learning.

XUETAO ZHANG received the B.S. degree in information engineering and the M.S. and Ph.D. degrees in pattern recognition and intelligence system from Xi'an Jiaotong University, China, in 2003, 2006, and 2012, respectively, where he is currently an Assistant Professor with the Institute of Artificial Intelligence and Robotics. He visited the Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, from 2009 to 2010. His research interests include computer vision, human vision, and machine learning.

ZHONGCHANG LI received the master's degree from the School of Electronic and Information Engineering, Xi'an Jiaotong University, in 2017. His research interests include computer vision and machine learning.

SHAOYI DU received the B.S. degrees in computational mathematics and computer science and the M.S. degree in applied mathematics and the Ph.D. degree in pattern recognition and intelligence system from Xi'an Jiaotong University, China, in 2002, 2005, and 2009, respectively, where he is currently a Professor with the Institute of Artificial Intelligence and Robotics. He was a Postdoctoral Fellow with Xi'an Jiaotong University, from 2009 to 2011, and visited the University of North Carolina at Chapel Hill, from 2013 to 2014. His research interests include computer vision, machine learning, and pattern recognition.

FEI WANG received the B.S. degree in electronics from Northwest University, in 1998, the M.S. degree in communication and information system from the Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, in 2002, and the Ph.D. degree in pattern recognition and intelligence system from Xi'an Jiaotong University, China, in 2009, where he is currently a Professor with the Institute of Artificial Intelligence and Robotics. He visited the North Carolina State University, USA, from 2012 to 2013. His research interests include computer vision and intelligent systems.