Calibration-free Gaze Tracking for Automatic Measurement of Visual Acuity in Human Infants

Chunshui Xiong[1], Lei Huang[1] and Changping Liu[1]

***Abstract*— Most existing vision-based methods for gaze tracking need a tedious calibration process. In this process, subjects are required to fixate on a specific point or several specific points in space. However, it is hard to cooperate, especially for children and human infants. In this paper, a new calibration-free gaze tracking system and method is presented for automatic measurement of visual acuity in human infants. As far as I know, it is the first time to apply the vision-based gaze tracking in the measurement of visual acuity. Firstly, a polynomial of pupil center-cornea reflections (PCCR) vector is presented to be used as the gaze feature. Then, Gaussian mixture models (GMM) is employed for gaze behavior classification, which is trained offline using labeled data from subjects with healthy eyes. Experimental results on several subjects show that the proposed method is accurate, robust and sufficient for the application of measurement of visual acuity in human infants.**

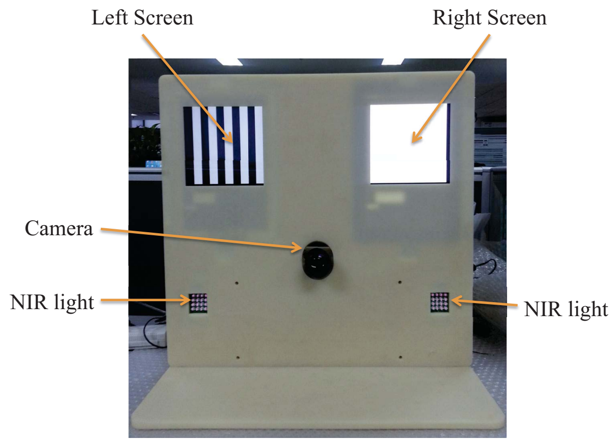


Fig. 1. Gaze tracking system for measurement of visual acuity.

I. INTRODUCTION

Measurement of visual acuity in human infants has been studied for many years. A variety of techniques have been proposed for infants acuity testing [1]. The preferential looking techniques were proposed for measuring of grating acuity, depending on the fact that infants spontaneously fixate striped filed in preference to a uniform filed [2]. Some studies based on the preferential looking techniques have been carried out, such as forced preferential looking [3], operant preferential looking [4] and acuity card procedure [5]. In the preferential looking technique, pairs of stimuli are presented to a subject. Additionally, an experienced observer locating behind a peephole between the two stimuli is needed to judge the fixation behavior of the subject.

With the development of the computer vision techniques, eye gaze can be tracked accurately and noninvasively. It can be used for replacing the experienced observer's observation of subjects' gaze behavior in the measurement of visual acuity, which is impersonal and simplifies the measurement process greatly. In this paper, we mainly focus on the vision-based gaze tracking for measurement of visual acuity.

Gaze tracking systems started to appear in the early 1970s [6], which were mainly used for psychological studies. It has been an active research topic for many years due to its potential usages in many applications, such as market and advertising analysis, studies of cognitive processes, graphic display and many other human-computer interactions. Many gaze tracking methods appeared, a recent survey can be seen in [7]. Vision-based techniques have become the most popular technique for gaze estimation due to their low intrusiveness and good performance. Usually, these techniques can be classified into two groups: 3D model-based methods and 2D feature-based methods.

The 3D model-based gaze tracking methods directly reconstruct eye's 3D visual axis, and the point of gaze (PoG) can be obtained by intersecting the visual axis with the screen. However, 3D model based methods need fully calibrated setups, which require exact knowledge of the relative positions of the cameras, the light sources and the monitor.

The feature-based gaze tracking methods extract local features such as pupil center, iris center, cornea reflections or eye corners. Then, a mapping function from features to PoGs can be founded by a calibration process. The most popular method is pupil center-cornea reflections (PCCR) based method. Sesma et al. [8] gave a thorough review of the PCCR-based gaze estimation using a framework of one camera and two near-infrared (NIR) light sources. They evaluated multiple models with different normalized PCCR vectors and different polynomial functions. However, the feature-based methods need a calibration process, which is difficult for infants to cooperate.

In this paper, we present a calibration-free gaze tracking method for measurement of infant visual acuity. A polynomial of PCCR vector is proposed to be used as the gaze feature to represent human's gaze information. Gaussian mixture models (GMM) is employed for gaze behavior classification. The method has the following advantages: 1) no calibration is needed, 2) noninvasive for the subjects, 3) compared with the existing commercial gaze tracking setups [9], the proposed system is much cheaper. What's more, the measurement of visual acuity in human infants can be greatly simplified by applying the proposed gaze tracking method.

[1]Institute of Automation, Chinese Academy of Sciences, 95 Zhongguancun East Road, 100190, Beijing, China.

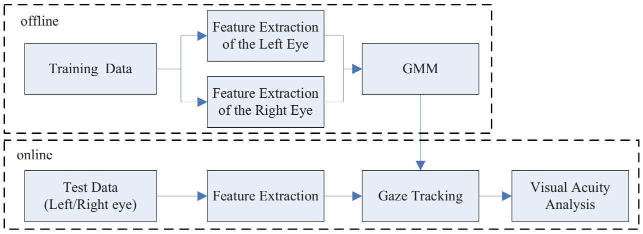


Fig. 2. Overview of the proposed method.

II. System Overview

The system is based on a camera, two NIR light sources and two screens (Fig. 1). The size of the screen is $12.4cm \times 12.4cm$ and the distance between the left screen and the right screen is $18.8cm$. The video camera is a CMOS camera with a resolution of 1280×1024, which is used for capturing the eye image for gaze tracking. The central wavelength of the NIR light source is 850nm. The distance from subjects to the screen is $50 \pm 5cm$. The two screens alternatively display a striped pattern and a gray pattern. As shown in Fig. 1, a striped pattern is displayed in the left screen and the gray pattern is displayed in the right screen. There are several striped patterns with different frequencies, which are designed specially for visual acuity measurement.

In traditional method of visual acuity measurement, an additional experienced observer is needed to observe the infant's gaze behavior, judging whether the subject is fixating on the screen that is displaying the striped pattern. In our system, the infant's gaze behavior is tracked automatically through the proposed gaze tracking method.

III. Method

The framework of the proposed gaze tracking method is shown in Fig. 2. The gaze tracking model is trained offline in advance. Then, the gaze of a subject can be predicted online for visual acuity analysis. The proposed gaze feature and the employed gaze tracking model are given in detail below.

A. Feature Extraction

The PCCR-based methods have been widely used for gaze estimation due to their simplicity. Usually, after an active calibration process, they could get high accuracy when multiple cornea-reflections are used. In this paper, a polynomial of the PCCR vector is proposed to be used as the gaze feature. Firstly, the pupil center and glints (cornea reflections) centers are precisely located. Then, a polynomial of the PCCR vector is calculated, which is used as the gaze feature.

The pupil and glints location procedures are shown in Fig. 3. Firstly, eye region detection is employed through sliding window based method, in which edge orientation histograms feature [10] is extracted and SVM [11] classifier is used to train the eye-region detector. After eye-region detection, an efficient fast radial symmetry [12] operator is adopted to precisely locate the center of glints, which is a fast gradient-based interest operator for detecting points of high radial symmetry. Based on the two glints, the pupil region can be cropped roughly (Fig. 3(d)). At last, since the pupil in eye image is approximate an ellipse, we employ Starburst [13] algorithm to precisely locate the center of pupil, which is a fast and accurate pupil location method.

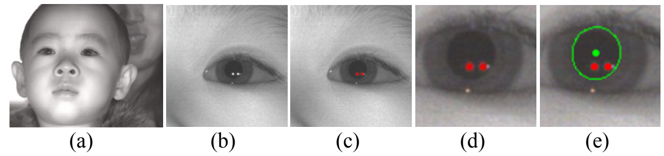


Fig. 3. Pupil and glints location. (a) Original image. (b) The detected eye region. (c) Glints location. (d) The cropped pupil region. (e) Pupil location.

Sesma-Sanchez et al. [8] have done many experiments to evaluate PCCR-based methods with different system configurations, different mapping functions and different normalization factors. Inspired by their experiments, a PCCR-based model with high accuracy and being robust against head movements can be designed as follows: the PCCR vector V can be calculated as

$$V = \frac{(P - R_1) + (P - R_2)}{2} \tag{1}$$

where P is the coordinate of pupil center, R_1 and R_2 are coordinates of two glints' centers (Fig. 3(d)). Then the normalized PCCR vector V_N is

$$V_N = \frac{V}{\|R_1 - R_2\|_2} \tag{2}$$

where $\|R_1 - R_2\|_2$ is the Euclidean distance between the two glints. Then the polynomial mapping function from vector V_N to PoG (S) can be established as follows:

$$\begin{pmatrix} S_x \\ S_y \end{pmatrix} = \begin{pmatrix} a_0 & a_1 & a_2 & a_3 & a_4 & a_5 \\ b_0 & b_1 & b_2 & b_3 & b_4 & b_5 \end{pmatrix} \times \begin{pmatrix} 1 \\ V_{Nx} \\ V_{Ny} \\ V_{Nx}V_{Ny} \\ V_{Nx}^2 \\ V_{Ny}^2 \end{pmatrix} \tag{3}$$

where $V_N = (V_{Nx}, V_{Ny})^T$, $S = (S_x, S_y)^T$, $(a_0,a_1,a_2,a_3,a_4,a_5, b_0,b_1,b_2,b_3,b_4,b_5)$ are coefficients of the polynomial to be estimated. As (3) is a quadratic regression model with feature of $(V_{Nx}, V_{Nx})^T$, it also can be viewed as a linear regression model with feature of $(1, V_{Nx}, V_{Ny}, V_{Nx}V_{Ny}, V_{Nx}^2, V_{Ny}^2)^T$, so we can define the polynomial of PCCR feature f as

$$f = \begin{pmatrix} V_{Nx} & V_{Ny} & V_{Nx}V_{Ny} & V_{Nx}^2 & V_{Ny}^2 \end{pmatrix}^T \tag{4}$$

When we use the feature f, only linear regression model is needed to achieve similar accuracy with the PCCR-based method. Inspired by the success of the PCCR-based method, the feature f is presented to be used as the gaze feature in our calibration-free gaze tracking method.

B. Gaze Tracking

In our method, unified gaze tracking model is built for both eyes. Gaussian mixture models (GMM) has been widely used for data analysis, therefore, it is used for modeling the gaze features in our method. Firstly, we briefly describe the GMM for gaze tracking. Then, Minimum Classification Error (MCE) training for updating the model parameters is introduced.

A GMM with K components in a D-dimensional ($D = 5$ in this paper) embedding space is defined as:

$$p(f;\lambda) = \sum_{k=1}^{K} \pi_k p(f|k) \tag{5}$$

$$p(f|k) = \mathscr{N}(f;\lambda_k) = \frac{1}{\sqrt{(2\pi)^D|\Sigma_k|}} \exp\left(-\frac{1}{2}(f-\mu_k)^{\mathrm{T}}\Sigma_k^{-1}(f-\mu_k)\right) \tag{6}$$

where f is a sample point in the D-dimensional space. $\lambda(\lambda = \{\lambda_k\}_{k=1}^K = \{\pi_k, \mu_k, \Sigma_k\}_{k=1}^K)$ is the set of model parameters, π_k is the prior probability of the k-th Gaussian component, μ_k, Σ_k are the mean and variance of the k-th component, $p(f|k)$ is the conditional probability density function.

For each gaze pattern, a GMM model is trained. There are $M(M = 2)$ gaze patterns, indicating the gaze on the left screen and the right screen (Fig. 1) in our system. Then, M class conditional likelihood functions $\{p_i(f;\lambda^{(i)})\}_{i=1}^M$ are trained through Expectation Maximization [14] independently, where $\lambda^{(i)}$ is the parameter set of the i-th GMM. However, the classifier relying on distribution estimation are suboptimal when the assumed distribution form is not the true one [15]. To optimize the classification result directly, the MCE criterion is employed for fine-tuning the parameters.

The MCE method directly formulates the classifier design problem as a classification error rate minimization problem. To achieve this goal, an optimization criterion need to be defined, which is usually a function of the class conditional likelihood functions $\{p_i(f;\lambda^{(i)})\}_{i=1}^M$. And the classifier makes its decision for each input sample by choosing the largest of the class conditional likelihood function evaluated on the input sample. In this respect, there exist many possibilities, one of which is a class misclassification measure taking the following form:

$$d_i(f) = -\log\left(p_i(f;\lambda^{(i)})\right) + \max_{j\neq i}\left(\log\left(p_j(f;\lambda^{(j)})\right)\right) \tag{7}$$

For a sample of the i-th class, $d_i(f) > 0$ implies misclassification and $d_i(f) < 0$ means correct decision. A general form of the loss function can then be defined as

$$\ell_i(f;\lambda^{(i)}) = \frac{1}{1+\exp\left(-\alpha\cdot d_i(f)+\beta\right)} \tag{8}$$

where Λ ($\Lambda = (\lambda^{(1)}, \lambda^{(2)}, \cdots, \lambda^{(M)})$) are the parameters, α is normally set to be not less than one and β is set to be zero. For a given training data set consisting of N samples $\{f_i\}_{i=1}^N$, the empirical probability measure defined on the training data set is a probability measure with equal weight at each sample. The empirical loss is thus expressed as

$$L(\Lambda) = \frac{1}{N}\sum_{j=1}^{N}\sum_{i=1}^{M}\ell_i(f_j;\lambda^{(i)})I(f_j \in C_i) \tag{9}$$

where $I(\cdot)$ is the indicator function. The model parameters can be updated through gradient descent:

$$\Lambda_{t+1} = \Lambda_t - \varepsilon_t \nabla L(\Lambda)\mid_{\Lambda=\Lambda_t} \tag{10}$$

where t is the number of iterations and ε_t is the learning rate. After getting the model parameters, the classifier is operating under the following decision rule:

$$\mathscr{C}(f) = C_i \quad if \quad p_i(f;\lambda^{(i)}) = \max_j\left(p_j(f;\lambda^{(j)})\right) \tag{11}$$

In practical applications, to further identify the classification's confidence, some results with low confidence should be dropped. For a sample f, it is considered to belong to C_i only if it also meets the following additional rule:

$$\log(p_i(f;\lambda^{(i)})) \geq \log(p_i(ft;\lambda^{(i)})) \tag{12}$$

where $ft = \sum_{k=1}^K \pi_k^{(i)} * \left(\mu_k^{(i)} + \sqrt{3 * diag(\Sigma_k^{(i)})}\right)$ in our experiments.

IV. Experiments

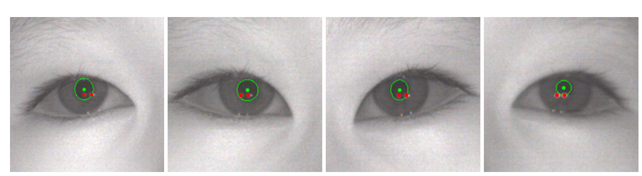
A. Data Acquisition

There is no public dataset for gaze tracking as far as I know. In this paper, the proposed gaze tracking method is designed for replacing the observer's observation of human infants' gaze behavior in the measurement of visual acuity. Based on the system, gaze data of 17 subjects (10 males and 7 females with age between 3 to 30 years old) with healthy eyes are collected to evaluate the gaze tracking method.

Each subject is asked to sit in front of the gaze tracking system (Fig. 1) and fixate on the screen that is showing a stripe pattern. The stripe pattern is shown in one screen and the white pattern is shown in the other screen, two patterns are shown in the two screens alternatively. The left eye and right eye's data are collected respectively. As to each eye, 20 samples are collected, 10 for the left screen and 10 for the right screen. Totally 680 samples are collected. After eliminating some inaccurate samples, 661 samples are selected. Ten subjects' data (385 samples) are used for training the GMMs and the rest seven subjects' data (276 samples) are used as the test dataset. Some data examples are shown in Fig. 4.

B. GMM Training

In order to validate the effectiveness of the training method for GMM, one training result is shown in this section. As to each GMM, three Gaussian components are used in our experiments. The initial mean values are set through k-means algorithms, the initial variances are set to be identity matrixes



(a) (b) (c) (d)

Fig. 4. Samples from one subject, results of pupil and glints location are also shown in the eye images. (a) Left eye's image when the subject fixates on the left screen. (b) Left eye's image when the subject fixates on the right screen. (c) Right eye's image when the subject fixates on the left screen. (d) Right eye's image when the subject fixates on the right screen.

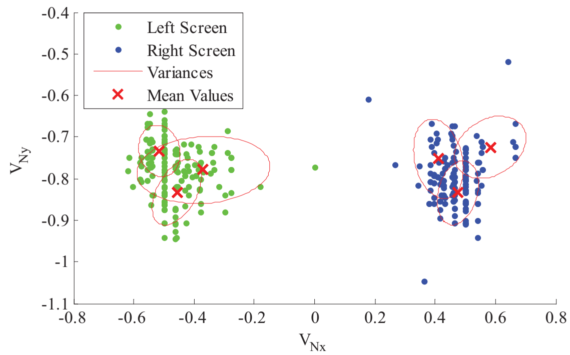


Fig. 5. A training result of the GMMs.

and the initial weights of the three Gaussian components are set to be $1/3$ equally.

As the dimension of the polynomial of PCCR feature is five. For visualization, we show the training results of the first two dimensions (V_{Nx}, V_{Ny}) of the gaze feature. As shown in Fig. 5, the training data are well modeled by the GMMs. And also, the two gaze patterns (fixating on the left screen and the right screen) are well differentiated by the first two dimensions of the feature, which shows the effectiveness of the polynomial of PCCR feature to represent human's gaze.

C. Gaze Tracking Results

To validate that the proposed gaze tracking method is available for the measurement of visual acuity, both offline test and online test are evaluated based on our system.

Offline test. As to classification operating under the decision rule in (11), the accuracy is 100% in the test dataset. In practical applications, some subjects may not fixates on the two screens, dropping some low confidence samples is essential. After dropping some samples with low confidence using rules in (12), the results are shown in Table I. The dropped rate is low for both eyes of the seven subjects in the test dataset. Note that the dropped rate of subject 12 is a little higher than other subjects. This is because he wears glasses, the reflection on the glasses affects the precision of the pupil and glints location. On the whole, the results show that the method achieves high accuracy, which is the premise for reliable medical application.

Online test. This test is carried out in C++ implementation on a laptop with an Intel Core 2 Duo 2.10GHz CPU, 2GB RAM. About thirty subjects are asked to sit in front of the system for an online test, in which they fixate on the left screen, the right screen or some other locations. The system give the predicting results left, right or not on the two screens accordingly. According to feedbacks of the subjects and artificial observations for some young children, the system predicts the gaze patterns accurately for all subjects even some subjects wear glasses or move his head naturally. The average computing time of the method is about 10 frames per second, which is sufficient for visual acuity analysis.

Both the offline test and online test show that the proposed gaze tracking method achieves high accuracy. And the process speed is also acceptable. What's more, the whole gaze tracking procedure is noninvasive and does not need a calibration process for a new subject, so even infants can be measured easily in our system.

TABLE I
GAZE TRACKING RESULTS

Subjects	Left Screen Dropped/Total number	Right Screen Dropped/Total number
Subject 11	0/20	0/20
Subject 12	1/20	2/20
Subject 13	1/20	0/20
Subject 14	0/19	0/20
Subject 15	0/19	0/20
Subject 16	0/20	0/20
Subject 17	1/18	0/20

V. CONCLUSIONS

We present a gaze tracking system and method for measurement of visual acuity. The system is noninvasive and calibration-free, so even infants can be measured easily. It only needs a single camera and two NIR light sources, which are cheap and simple. Experimental results show that the gaze tracking method is accurate with a speed of about 10 frames per second, which is sufficient for further visual acuity analysis. Possible tracks for further improvements include subpixel detection of corneal reflections and pupil center, testing different NIR lights configurations and prediction models.

REFERENCES

[1] V. Dobson and D. Y. Teller, "Visual acuity in human infants: a review and comparison of behavioral and electrophysiological studies," *Vision Research*, vol. 18, no. 11, pp. 1469–1483, 1978.
[2] R. L. Fantz, "Visual perception from birth as shown by pattern selectivity*," *Annals of the New York Academy of Sciences*, vol. 118, no. 21, pp. 793–814, 1965.
[3] D. Y. Teller, R. Morse, R. Borton, and D. Regal, "Visual acuity for vertical and diagonal gratings in human infants," *Vision Research*, vol. 14, no. 12, pp. 1433–1439, 1974.
[4] D. Mayer and V. Dobson, "Assessment of vision in young children: a new operant approach yields estimates of acuity.," *Investigative ophthalmology & visual science*, vol. 19, no. 5, pp. 566–570, 1980.
[5] M. McDonald, V. Dobson, S. Sebris, L. Baitch, D. Varner, and D. Teller, "The acuity card procedure: a rapid test of infant acuity.," *Investigative ophthalmology & visual science*, vol. 26, no. 8, pp. 1158–1162, 1985.
[6] L. R. Young and D. Sheena, "Eye-movement measurement techniques.," *American Psychologist*, vol. 30, no. 3, p. 315, 1975.
[7] D. W. Hansen and Q. Ji, "In the eye of the beholder: A survey of models for eyes and gaze," *IEEE T-PAMI*, vol. 32, no. 3, pp. 478–500, 2010.
[8] L. Sesma-Sanchez, A. Villanueva, and R. Cabeza, "Gaze estimation interpolation methods based on binocular data," *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 8, pp. 2235–2243, 2012.
[9] "Tobii technology." http://www.tobii.com/, 2014. [Online; accessed 24-February-2014].
[10] W. T. Freeman and M. Roth, "Orientation histograms for hand gesture recognition," in *International Workshop on Automatic Face and Gesture Recognition*, vol. 12, pp. 296–301, 1995.
[11] V. Vapnik, *The nature of statistical learning theory*. springer, 1999.
[12] G. Loy and A. Zelinsky, "Fast radial symmetry for detecting points of interest," *IEEE T-PAMI*, vol. 25, no. 8, pp. 959–973, 2003.
[13] D. H. Li, D. Winfield, and D. J. Parkhurst, "Starburst: A hybrid algorithm for video-based eye tracking combining feature-based and model-based approaches," in *Proc. CVPRW*, pp. 79–79, 2005.
[14] A. P. Dempster, N. M. Laird, D. B. Rubin, *et al.*, "Maximum likelihood from incomplete data via the em algorithm," *Journal of the Royal statistical Society*, vol. 39, no. 1, pp. 1–38, 1977.
[15] B.-H. Juang, W. Hou, and C.-H. Lee, "Minimum classification error rate methods for speech recognition," *Speech and Audio Processing, IEEE Transactions on*, vol. 5, no. 3, pp. 257–265, 1997.