Model Construction and Validation in Low-cost Interpolation-Based Gaze Tracking System

Suwitchaya Rattarom, Surapong Uttama, Nattapol Aunsri*, *Member, IAENG.*

***Abstract*—There are numbers of polynomial equations used in a low-cost gaze tracking system with interpolation-based gaze estimation; however, most of the polynomials are not statistically validated. The objective of this study is to propose a new model for low-cost gaze tracking system and perform a statistical analysis to validate the model. The process consists of two main steps. First, the candidate equations were constructed by finding the relationships between the position on the screen and pupil-glint vector. Then the equations matching the conditions were chosen. The equations contain 8-9 terms with an acceptable relation rate from the fitting process and are suitable with 9-25 calibration targets. Second, the statistical hypothesis testing to test the null hypothesis that the coefficient of the equation is not significant was performed for assuring that the obtained equations were validated. Afterward, the equations were merged into the candidate model. The results reveal that the best model is derived from the relationships between position X on the screen and a pupil-glint vector with polynomial degree 3 and 1, and the relationships between position Y on the screen and pupil-glint vector with a polynomial degree of 2 and 2. The accuracy of the model is 2.28°-2.51° for 12-25 calibration targets. The accuracy can be further improved if the gaze estimation splits is considered based on the screen areas: left, right and middle area of the screen. The left region is mapped by the pupil-glint vector of the right eye, the right region is mapped by the left eye and the middle region is mapped by the average from both eyes. By using a region based estimation at 5% of the sweep distance, we can improve the accuracy of the model 3122 for 25 points of calibration from 2.28° to 1.96°.**

We suggest that the cutting point at 5% of the sweep distance can improve the accuracy from 2.28° to 1.96°.

***Index Terms*—Gaze tracking, Interpolation-based gaze estimation, Mapping function, Model validation**

I. Introduction

HUMAN gazing indicates where we are looking at and shows that the rich information is useful for many applications, especially for those who are disables. For example, people with Lock-in syndrome, a condition preventing patients to move nearly all voluntary muscles in the body except the eyes, need to gaze in order to communicate with the others. Marketing research needs gazing data to know whether people are looking at the right spots of their advertisement. In user interface design, knowing where people are looking at is crucially important to create the proper interfaces. Researches in automobile use gazing data of drivers to detect their instantaneous situation such as paying attention to the road or feeling sleepy [1], [2].

Gaze tracking system, a system that tracks the eye position and measures the eye-gaze direction, is the important tool for those applications mentioned earlier. Recently, there are some commercial products available for the public. However, all of those systems use special hardware like 3D camera or high-resolution camera to track the eyes and are too expensive to afford. On the other hand, many ongoing researches have been trying to use ordinary and inexpensive cameras, like web camera, to reduce the system's cost in which larger groups of people with disability are reachable. However, due to the limitation of this type of hardware only some methodologies are applicable.

One of the famous methods used in low-cost gaze tracking systems is interpolation-based or regression-based gaze estimation. This method uses general-purpose polynomial equations to map the eye gaze. There are various proposed forms of polynomial equation suggested by many researchers, but most of them are not yet statistically validated, especially in the environment of low-cost gaze tracking systems.

Therefore, the aim of this work is to propose a new model for web camera-based gaze tracking system and perform statistical analysis to validate the constructed model. In this paper, we present how the model is formulated relying on the interpolation-based technique. The validation process is then applied to ensure that the model is well-fitted and accurate.

This paper is structured as follows. Section II presents the interpolation-based gaze estimation systems. The detailed methodology for the generation of the polynomial used in the system is provided in Section III. The experimental results are found in Section IV and conclusions are presented in Section V.

II. Interpolation-based gaze estimation systems

A. Gaze tracking system

Most gaze tracking systems can be divided into two parts: eye features extraction and gaze estimation [3]. The first part is the process that extracts features from eye images. The features consist of center of the pupil, iris spot, eyelid position, eye corner location, and position of corneal reflection (glint) – reflection of the infrared light on the corneal. Gaze estimation, which is another part of the system, consists of mapping functions that map some features from the first part to the coordinates on the screen to estimate the gaze direction.

B. Low-cost Gaze tracking system

Although there are many studies reported in literatures about the attempts to find the various polynomial models, most of the research studies in low-cost gaze tracking system still use simple models like quadratic function or linear function. The work from [4], using ITU Gaze Tracker software, demonstrated the use of this software with a low-cost web camera that's built with an infrared light. Moreover,

Manuscript received April 4, 2018; revised October 12, 2018.

S. Rattarom and S. Uttama are with the School of Information Technology, Mae Fah Luang University, Chiang Rai, 57100 Thailand e-mail: shya_en30@hotmail.com.

N.Aunsri is with the Brain Science and Engineering Innovation Research Unit and with the School of Information Technology, Mae Fah Luang University, Chiang Rai, 57100 Thailand, e-mail: nattapol.aun@mfu.ac.th.

* Author to whom correspondence should be addressed.

[5] constructed the low-cost gaze tracking system that works with ITU Gaze tracker. They tested the accuracy of their system and compared it with a commercial one. The result showed that low-cost system still had a lower accuracy than that of the commercial system. However, both devices offer an acceptable accuracy. In [6]–[8], they focused on building a 'Do-It-Yourself' head-mount gaze tracker. All of their systems used low-cost web camera and used quadratic function to estimate eye gaze. Researches from [9] and [10] used quadratic function in their systems. However, they concentrated on head pose estimation and the system worked in a natural light environment. Work from [11] is another study that paid attention to low-cost system in the natural light. Nevertheless, this work was based on linear function instead.

C. Interpolation-based (regression-based) gaze estimation

The regression method uses the general-purpose polynomial equations with unknown coefficients to estimate the gaze direction. The equation consists of two independent variables for each of the response variables. The independent variables are the coordinates x and y of pupil center of the eyes extracted from the eyes' image. In order to fix the error position from the acquired image, most of the researches used glint reflection from the infrared light source as a reference point for pupil position and use pupil-glint vector to represent x and y. The response variables are the coordinates X and Y of the eye-gaze direction on the screen. The model equations for left and right eyes do not need to be the same. The equation for mapping the position is represented by a form of simple linear regression as:

$$X = a_0 + a_1x + a_2y + a_3xy \tag{1a}$$

$$Y = b_0 + b_1x + b_2y + b_3xy \tag{1b}$$

Before performing gaze estimation, the method needs a special process called "calibration process." This process lets users to gaze at some pre-defined specific points on the screen. When the user is looking at a point, the positions of calibration target points on the screen (x and y) and pupil-glint vectors (X and Y) of the eyes image are collected. Subsequently the coefficients $a_0 \dots a_3$, $b_0 \dots b_3$ are calculated using a linear regression. Finally, the gaze estimation function is applied to the rest of the screen. However, the shape of the eyes is almost spherical, so the relationship between pupil-glint vector (point on the spherical object) and position on the screen (point on the plane object) should be constructed using a polynomial equation rather than a linear equation. As a consequence, in most studies, the polynomial equations are considered instead of the linear equations.

D. Polynomial model

There are many polynomial models used in the regression-based gaze tracking. The common and widely used model found in [12], [13], and also used in the open source software "The ITU Gaze Tracker [14]" is:

$$X = a_0 + a_1x + a_2y + a_3x^2 + a_4xy + a_5y^2 \tag{2a}$$

$$Y = b_0 + b_1x + b_2y + b_3x^2 + b_4xy + b_5y^2 \tag{2b}$$

Quantities x and y refer to the normalized x and y components of the pupil-glint vector of a specific eye at a specific point in time. X and Y refer to the X-coordinate and Y-coordinate of the 'Point of Regard' for the specific eye on the two dimensional plane of the screen. The coefficient a_i and b_i are determined through a calibration process. This model works with both left and right eyes.

Research studies done by [12] and [13] used this model with a video camera and two infrared light sources. It should be noted that the ITU Gaze tracker program supports one camera with both one and two infrared light sources.

Work in [15] used two cameras to make a 3D eye position and one infrared light source. Their proposed model is described by the two equations below.

$$X = a_0 + a_1x + a_2y + a_3xy \tag{3a}$$

$$Y = b_0 + b_1x + b_2y + b_3y^2 \tag{3b}$$

Some researchers tried to find another model that provides a good accuracy for the basic system with one video camera and one light source. [16] studied over 400,000 models and found that, increasing number of terms or the order of the polynomial may not necessarily affect the system accuracy. It is the same suggestion with work reported in [17]. From 400,000 models, they suggested two models that provide a good accuracy and have a small number of terms. The first one is

$$X = a_0 + a_1x + a_2y + a_3x^2 \tag{4a}$$

$$Y = b_0 + b_1y + b_2x^2 + b_3xy + b_4x^2y \tag{4b}$$

and the second one is

$$X = a_0 + a_1x + a_2x^3 + a_3y^2 \tag{5a}$$

$$Y = b_0 + b_1x + b_2y + b_3x^2y \tag{5b}$$

Another research by [18] built a system with one video camera and one light source and tested 625 polynomial models with multiple sets of calibration points. However, the best accuracy model was different from the model proposed by [16]. They claimed that this was from the different environments. They proposed a model that has the best accuracy if it is used with eight or more calibration points while having a set of small terms. The model is given as follows:

$$X = a_0 + a_1x + a_2x^3 + a_3y^2 + a_4xy \tag{6a}$$

$$Y = b_0 + b_1x + b_2x^2 + b_3y + b_4y^2 + b_5xy + b_6x^2y \tag{6b}$$

Instead of using brute force techniques to find the best fitting model, the work from [19], by using the same system of their previous research, proposed the method to find the causes of the polynomial equation. This method looked for the relationships between one independent variable (while controlling another one) with the response variable in the same axis, such as finding the relationships between x and X in the same Y position. After that, the coefficients of the first relation were used to find the relation of another variable again. The process is done like that in both response variables X and Y, and both left and right eyes. The best model is given as follows.

$$X = a_0 + a_1x + a_2x^2 + a_3x^3 + a_4y + a_5xy + a_6x^2y + a_7x^3y \tag{7a}$$

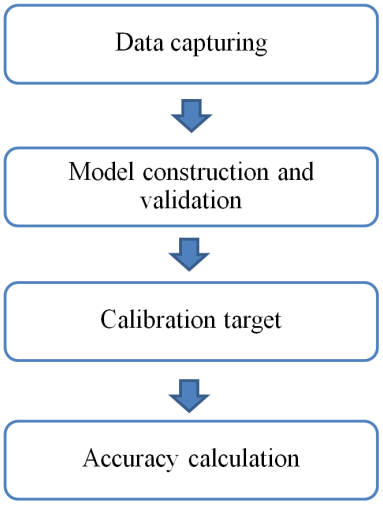

Fig. 1. Model construction process.

$$Y = b_0 + b_1x + b_2x^2 + b_3y + b_4y^2 + b_5xy + b_6x^2y \quad (7b)$$

Nevertheless, the extended version of this research increased participants from 9 to 25 persons and added up some steps in the data capturing process [20]. The paper provided a slightly differed polynomial model and showed slightly less accuracy than the model from the previous research.

The reason that the simple models were chosen is that they are common and provide the acceptable accuracy. However, it is not clearly seen that the simple models are the best choices without just the conclusion from the accuracy obtained from the experiments, especially in a low-cost gaze tracking system that was established by the web camera. Therefore, research from [21] compared the accuracy of simple and many other models that had been described in the previous topic under the same environmental settings for low-cost gaze tracking. The results showed that the model from [19] provided better accuracy than simple models.

III. Methodology

A. Model construction process

The model construction and validation framework contains four main processes, the framework is shown in Fig.1 and outlined as follows.

(i) *Data capturing and outliers removal* - This process is for capturing data for estimating the eye gaze and removing the outliers occurred from image processing. To remove the outliers, for every target point in each experiment, we selected the acquired images in which the pupil position or glint position lie within two standard deviation of the mean and remove those that lie outside two standard deviation of the mean. The selection process on the left and the right eyes are done separately.

(ii) *Model construction and validation* - This process used the relationships between gaze positions and pupil-glint vector to construct the candidate equations. Then, it is performed with the statistical hypothesis testing to diminish the inessential coefficients. Finally, candidate models were constructed by merging these equations.

(iii) *Calibration target* - This task is to setup of the calibration simulation. This work simulated the gaze estimation results in five calibration sets with different calibration points.

(iv) *Accuracy calculation* - With the gaze estimation results, the accuracy was calculated by using the remaining calibration points.

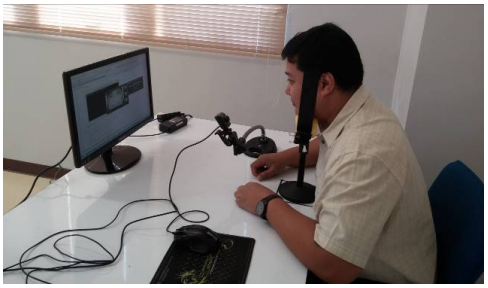

Fig. 2. Hardware and environment configuration.

B. Experimental setting

The gaze tracking hardware is composed of an ordinary web camera and 20-inch LED monitor. The camera model is WC-208 from SIGNO Technology Co., Ltd. It contains three infrared LEDs built-in under the lens. The monitor is the Samsung SyncMaster model LS20B300BS-ZA. The resolution of the screen was set to 1600x900 pixels. The video from the web camera was set to 30 frames per second with a resolution of 640x480 pixels, and processed as binary images. The monitor was placed in front of the participants with a distance of 600 mm while the camera was attached to the bottom side of the monitor with a distance of 220 mm from the eyes. The observers used a chinrest in order to fix their head position while running the experiments. Fig.2 illustrates the hardware setting.

Twelve experimenters including eight females and four males between the ages from 21 to 40 years old participated in the study; the age average was 30.6 years old. Among the participants, six of them have a normal vision whereas the others were near-sighted and have astigmatism in both eyes. One subject wore an eyeglass during the experiment while two subjects used contact lenses. All of them had never taken part with the eye-gaze experiment before.

C. Data capturing

TTo cover the significant area of the screen, the screen was divided into 16x9 grids with a dimension of 100x100 pixels each. 144 calibration points were placed in the center of grids, one by one, as shown in Fig.3. We used the ITU Gaze Tracker program to collect the raw data (pupil and glint position of both eyes in each target point). However, this program only allowed 16 calibration target points to be collected and does not offer an option to save the raw data. So, we modified the program in such a way that it can produce 16x9 calibration target points and can export raw data for our study.

The modification of the ITU Gaze Tracker program can display the calibration points and capture the pupil-center of the eyes and the glint reflection from the IR light source. For the capturing process, the participants were asked to look at all calibration targets respectively and each point was shown for 1 second. For each target, approximately 30 pupil-glint

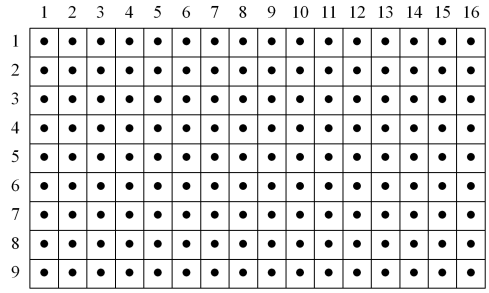

Fig. 3. 144 calibration points used in data capturing process.

images were captured. Three data sets were collected from each participant.

Due to the limitations of the low-resolution camera, some outliers had to be eliminated before moving to the next step. These data were later used in the calibration procedure for various combinations of calibration target arrangements and mapping models.

D. Model construction and validation

From the work done in [22], they fitted the model by using the relationships between gaze positions and pupil-glint vector in the same row/column of calibration point. Afterwards, the first relation was used to find the relation of the overall model. Then the validation process with statistical hypothesis testing was performed to check if all coefficients are significant. However, there were many candidate polynomial models that could fit the relation by just considering the coefficient of determination, R^2, which is the quantity to indicate the fitness of the model that was used in most model fitting problems, [23] is an example. Therefore, in this study we constructed the model in a different manner outlined.

The explanation of the numbering associated with the model is described as follows. Number of the model means the degree of polynomial in the relationship of full-size model. For instance, 3122; 31 is polynomial for X and 22 is polynomial for Y. 31 which means that the relationship between X and pupil-glint vector x is a polynomial degree 3, and the relationship between coefficient a, b, c, d and pupil-glint vector y is a polynomial degree 1. 22 which means that the relationship between Y and pupil-glint vector y is a polynomial degree 2, and the relationship between coefficient a, b, c and pupil-glint vector x is a polynomial degree 2. Below, we explain the two-step model construction and validation.

1) We first formulated the possible polynomial equations from the number of the relationships between gaze positions and pupil-glint vector [19], [20], [22]. Then we tested them with the normalized average data. From our experiment, the two best equations are 31 and 22 equations. The 31 equation came from the first relationship between x and X that is the polynomial degree 3, then the relationships between coefficients a,b,c,d and Y are the polynomial degree 1. The equation for Y was created by the same polynomial degree. This equation has eight terms and is in the form of

$$X = a_0x^3y + a_1x^3 + a_2x^2y + a_3x^2 + a_4xy + a_5x + a_6y + a_7 \quad (8a)$$

$$Y = b_0y^3x + b_1y^3 + b_2y^2x + b_3y^2 + b_4yx + b_5y + b_6x + b_7 \quad (8b)$$

The 22 equation came from the first relationship between x and X that is the polynomial degree 2, then the relationships between coefficients a, b, c and Y are the polynomial degree 2. The same approach was repeated to derive a model for Y. This equation has nine terms and is given below.

$$X = a_0x^2y^2 + a_1x^2y + a_2x^2 + a_3xy^2 + a_4xy + a_5x + a_6y^2 + a_7y + a_8 \quad (9a)$$

$$Y = b_0y^2x^2 + b_1y^2x + b_2y^2 + b_3yx^2 + b_4yx + b_5y + b_6x^2 + b_7x + b_8 \quad (9b)$$

2) From the full-size equation obtained from step 1), we tested the null hypothesis that the coefficient of the equation is equal to zero (no effect). The null hypothesis is rejected when p-value $< .05$. The stored data of 144 calibration points from all experiments was used in this process. The testing was separated into four conditions operating with four groups of data:
   a) Test for X in left eye: This condition operates with the positions X on the screen, the pupil-glint vectors x and the pupil-glint vectors y of the left eye.
   b) Test for Y in left eye: This condition operates with the positions Y on the screen, the pupil-glint vectors x and the pupil-glint vectors y of the left eye.
   c) Test for X in right eye: This condition operates with the positions X on the screen, the pupil-glint vectors x and the pupil-glint vectors y of the right eye.
   d) Test for Y in right eye: This condition operates with the positions Y on the screen, the pupil-glint vectors x and the pupil-glint vectors y of the right eye.

   From this process, some coefficients that are insignificant to the equation were removed. Then four candidate models were constructed by merging these two equations. The models are 3131, 3122, 2231 and 2222. The details are presented in Section IV.

The candidate models for X and Y are shown in Table I and Table II, respectively. Model from [19] gives the best accuracy as appeared in the literatures and models from other researches [21] are also used to compare with our proposed model.

E. Calibration targets

There is a fact that more calibration targets should provide better accuracy. On the other hand, more calibration targets take more time to calibrate and make the user's eyes fatigue. The acceptable solutions compromising the accuracy and time that have been used in many researches [4], [16], [18], [24] are found to use 9 to 25 calibration points that take the calibration time in less than a minute.

This work used the stored data of pupil-glint vector in 144 calibration points to simulate the calibration process. We set up five different calibration sets, which vary the number and arrangement of targets as displayed in Figs. 4-8, and

TABLE I
THE CANDIDATE POLYNOMIAL MODEL FOR X USED IN OUR EXPERIMENT

No.	Model	eye	Polynomial for X
1	3131	Left	$x^3y, x^3, x^2y, x^2, xy, x, y, 1$
		Right	$x^3y, x^3, x^2y, x^2, xy, x, y, 1$
2	3122	Left	$x^3y, x^3, x^2y, x^2, xy, x, y, 1$
		Right	$x^3y, x^3, x^2y, x^2, xy, x, y, 1$
3	2231	Left	$x^2y^2, x^2y, x^2, xy^2, xy, x, y^2, y, 1$
		Right	$x^2y^2, x^2, xy^2, xy, x, y^2, y, 1$
4	2222	Left	$x^2y^2, x^2y, x^2, xy^x, xy, x, y^2, y, 1$
		Right	$x^2y^2, x^2, xy^2, xy, x, y^2, y, 1$
5	[19]	Left	$x^3y, x^3, x^2y, x^2, x, xy, y, 1$
		Right	$x^3y, x^3, x^2y, x^2, x, xy, y, 1$

Note: For simplicity, polynomials are written without coefficients. The polynomial $b_1x^2y+b_2x+b_3y+b_4$ is, for example, written as $x^2y, x, y, 1$

TABLE II
THE CANDIDATE POLYNOMIAL MODEL FOR Y USED IN OUR EXPERIMENT

No.	Model	eye	Polynomial for Y
1	3131	Left	$y^3x, y^3, y^2x, y^2, yx, y, x, 1$
		Right	$y^3x, y^3, y^2x, y^2, yx, y, x, 1$
2	3122	Left	$y^2x^2, y^2x, y^2, yx^2, y, x^2, x, 1$
		Right	$y^2x, y^2, yx^2, yx, y, x^2, x, 1$
3	2231	Left	$y^3x, y^3, y^2x, y^2, yx, y, x, 1$
		Right	$y^3x, y^3, y^2x, y^2, yx, y, x, 1$
4	2222	Left	$y^2x^2, y^2x, y^2, yx^2, y, x^2, x, 1$
		Right	$y^2x, y^2, yx^2, yx, y, x^2, x, 1$
5	[19]	Left	$x^2y, xy, y, x^2, x, y^2, 1$
		Right	$x^2y, xy, y, x^2, x, y^2, 1$

Note: For simplicity, polynomials are written without coefficients. The polynomial $b_1x^2y+b_2x+b_3y+b_4$ is, for example, written as $x^2y, x, y, 1$

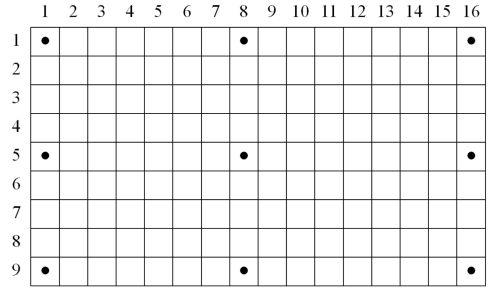

Fig. 4. Positions of calibration point in 9 calibration targets.

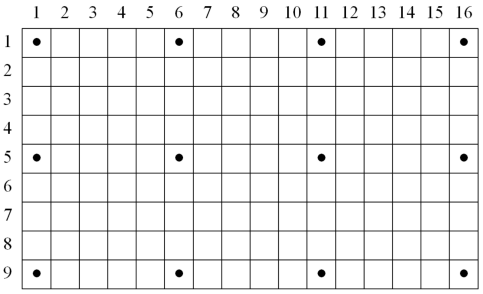

Fig. 5. Positions of calibration point in 12 calibration targets.

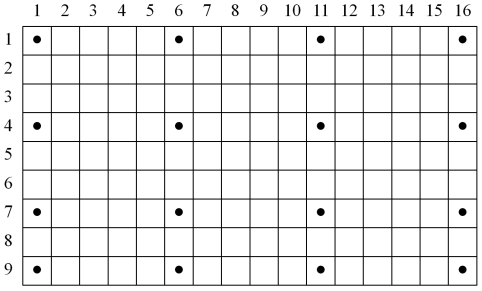

Fig. 6. Positions of calibration point in 16 calibration targets.

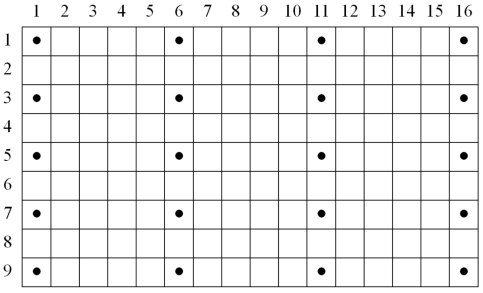

Fig. 7. Positions of calibration point in 20 calibration targets.

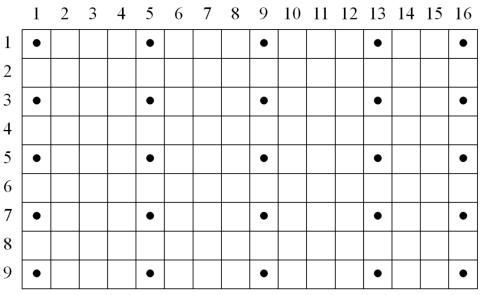

Fig. 8. Positions of calibration point in 25 calibration targets.

selected the data at a specific point of calibration. The pattern of the calibration sets displayed symmetry because of two reasons. First, the accuracy leans to be best in the middle of the screen and worst in the corner/border of the screen [25]. Therefore, the calibration points must cover all corners and borderlines in order to reduce the errors. Second, setting calibration points in the same lines of both X and Y axes provide more data at the same value X or Y to the equations.

The chosen points that are related to each set of the calibration were used to find the coefficients of the polynomial model from the previous section. The calibration process was done individually on all participants and experiments.

F. Accuracy calculation

The accuracy of mapping function means that the distance between the actual gaze direction and the measured gaze direction in degrees [25], [26]. Fig. 9 illustrates the concept of accuracy measurement for gaze mapping. Low value of accuracy is better than high value of accuracy. The actual gaze direction is the position of calibration point while the measured gaze direction is the result from polynomial model calculation. To compute the accuracy, this work used all data of 144 calibration points, except for the point used in calibration targets in each participant/experiment, to calculate the average accuracy of the model in all calibration sets.

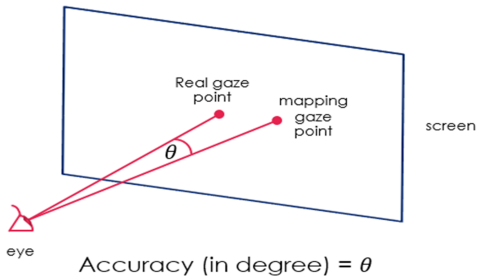

Fig. 9. Accuracy measurement in degree.

TABLE III
P-VALUES OF STATISTICAL HYPOTHESIS TESTING OF POLYNOMIAL DEGREE 3, 1

Coefficients	Test for X_{Left}	Y_{Left}	X_{Right}	Y_{Right}
a_0 / b_0	9.00×10^{-104}	2.40×10^{-198}	0	2.40×10^{-198}
a_1 / b_1	0	0	0	0
a_2 / b_2	2.58×10^{-197}	1.10×10^{-168}	3.81×10^{-318}	1.10×10^{-168}
a_3 / b_3	0	0	7.70×10^{-298}	0
a_4 / b_4	2.67×10^{-31}	2.79×10^{-123}	0	2.79×10^{-123}
a_5 / b_5	0	1.59×10^{-205}	0	1.59×10^{-205}
a_6 / b_6	3.97×10^{-307}	4.54×10^{-54}	0	4.54×10^{-54}
a_7 / b_7	0	0	0	0

Note: P-value in boldfaced mean accepted the null hypothesis.

IV. Experimental Results

A. Derivation of the equation with polynomial degree 3, 1

From the full-size equation explained in the previous session, we test the null hypothesis that the coefficient of the equation is equal to zero (no effect). The p-values of four conditions are given in Table III.

From the results, all coefficients are likely to be meaningful to the model equations. So, the validation equation for X and Y, both left and right eyes are the same as full-size equation. The equations for X and Y are given as:

$$X_{Left,Right} = a_0x^3y + a_1x^3 + a_2x^2y + a_3x^2 + a_4xy + a_5x + a_6y + a_7 \quad (10a)$$

$$Y_{Left,Right} = b_0y^3x + b_1y^3 + b_2y^2x + b_3y^2 + b_4yx + b_5y + b_6x + b_7 \quad (10b)$$

B. Derivation of the equation with polynomial degree 2, 2

In order to derive the equation with the polynomial degree 2, 2, the similar process with the derivation of the polynomial degree 3, 1 was performed. The p-values of four conditions are displayed in Table IV.

Table IV shows that some coefficients are insignificant to the equations. So the reduced equations for X and Y are summarized below.

$$X_{Left} = a_0x^2y^2 + a_1x^2y + a_2x^2 + a_3xy^2 + a_4xy + a_5x + a_6y^2 + a_7y + a_8 \quad (11a)$$

$$Y_{Left} = b_0y^2x^2 + b_1y^2x + b_2y^2 + b_3yx^2 + b_5y + b_6x^2 + b_7x + b_8 \quad (11b)$$

$$X_{Right} = a_0x^2y^2 + a_2x^2 + a_3xy^2 + a_4xy + a_5x + a_6y^2 + a_7y + a_8 \quad (11c)$$

$$Y_{Right} = b_1y^2x + b_2y^2 + b_3yx^2 + b_4yx + b_5y + b_6x^2 + b_7x + b_8 \quad (11d)$$

TABLE IV
P-VALUES OF STATISTICAL HYPOTHESIS TESTING OF POLYNOMIAL DEGREE 2, 2

Coefficients	Test for X_{Left}	Y_{Left}	X_{Right}	Y_{Right}
a_0 / b_0	4.70×10^{-19}	2.48×10^{-4}	1.45×10^{-27}	**0.227804**
a_1 / b_1	1.05×10^{-82}	1.39×10^{-5}	**0.207135**	5.52×10^{-9}
a_2 / b_2	2.64×10^{-193}	2.5×10^{-282}	0	0
a_3 / b_3	5.81×10^{-66}	1.13×10^{-11}	7.12×10^{-22}	0.00463
a_4 / b_4	4.53×10^{-76}	**0.059075**	3.26×10^{-5}	1.18×10^{-50}
a_5 / b_5	0	0	0	0
a_6 / b_6	9.55×10^{-3}	1.7×10^{-17}	4.15×10^{-85}	5.67×10^{-7}
a_7 / b_7	5.80×10^{-27}	1.6×10^{-4}	2.00×10^{-188}	2.5×10^{-105}
a_8 / b_8	0	0	0	0

Note: P-value in boldfaced mean accepted the null hypothesis.

C. Derivation of the candidate models

After validating the equations, we merge them into four sets of candidate models 3131, 3122, 2231 and 2222. The patterns of merging models on left and right eye are the same. However, the models may be different based on the validated equations. Each model can be described below.

- **Model 3131.** This model uses the polynomial degree 3, 1 in both X and Y equations. Because this equation is full-size, the model equations of the left and the right eye are the similar. The models equations can be written as follows.

$$X_{Left,Right} = a_0x^3y + a_1x^3 + a_2x^2y + a_3x^2 + a_4xy + a_5x + a_6y + a_7 \quad (12a)$$

$$Y_{Left,Right} = b_0y^3x + b_1y^3 + b_2y^2x + b_3y^2 + b_4yx + b_5y + b_6x + b_7 \quad (12b)$$

- **Model 3122.** This model uses the polynomial degree 3,1 in X equation and polynomial degree 2,2 in Y equation. For this model, Y equations of left and right eye are different based on the validated equation. The model equations are shown below.

$$X_{Left,Right} = a_0x^3y + a_1x^3 + a_2x^2y + a_3x^2 + a_4xy + a_5x + a_6y + a_7 \quad (13a)$$

$$Y_{Left} = b_0y^2x^2 + b_1y^2x + b_2y^2 + b_3yx^2 + b_5y + b_6x^2 + b_7x + b_8 \quad (13b)$$

$$Y_{Right} = b_1y^2x + b_2y^2 + b_3yx^2 + b_4yx + b_5y + b_6x^2 + b_7x + b_8 \quad (13c)$$

- **Model 2231.** This model uses the polynomial degree 2,2 in X equation and the polynomial degree 3,1 in Y equation. For X_{right} equation, the polynomial is reduced and the polynomial is the full-size for X_{left} equation. The models could be written as the following.

$$X_{Left} = a_0x^2y^2 + a_1x^2y + a_2x^2 + a_3xy^2 + a_4xy + a_5x + a_6y^2 + a_7y + a_8 \quad (14a)$$

$$X_{Right} = a_0x^2y^2 + a_2x^2 + a_3xy^2 + a_4xy + a_5x + a_6y^2 + a_7y + a_8 \quad (14b)$$

$$Y_{Left,Right} = b_0y^3x + b_1y^3 + b_2y^2x + b_3y^2 + b_4yx + b_5y + b_6x + b_7 \quad (14c)$$

TABLE V
AVERAGE ACCURACY (IN DEGREE) OVER 12 PARTICIPANTS WITH THREE REPETITIONS FOR EACH MODEL WITH 9 AND 12 CALIBRATION TARGETS.

Model	9 calibration targets			12 calibration targets		
	Average	Left	Right	Average	Left	Right
3131	3.90	3.77	4.04	2.66	2.66	2.65
3122	3.75	3.54	3.97	**2.51**	**2.50**	**2.52**
2231	3.30	3.33	3.26	2.79	2.75	2.83
2222	**3.11**	**3.09**	**3.13**	2.63	2.58	2.68
[19]	3.71	3.46	3.97	2.53	**2.50**	2.57

Note: The best accuracy for every calibration configuration is boldfaced.

TABLE VI
AVERAGE ACCURACY (IN DEGREE) OVER 12 PARTICIPANTS WITH THREE REPETITIONS FOR EACH MODEL WITH 16 AND 20 CALIBRATION TARGETS.

Model	16 calibration targets			20 calibration targets		
	Average	Left	Right	Average	Left	Right
3131	2.45	2.46	2.45	2.40	2.38	2.43
3122	**2.40**	**2.41**	**2.39**	**2.33**	**2.34**	**2.33**
2231	2.54	2.53	2.55	2.45	2.47	2.43
2222	2.47	2.48	2.47	2.38	2.42	2.33
[19]	2.44	2.42	2.45	2.36	**2.34**	2.39

Note: The best accuracy for every calibration configuration is boldfaced.

- **Model 2222.** This model uses the polynomial degree 2,2 in both X and Y equations. Like the first three models, the model equations are given below.

$$X_{Left} = a_0x^2y^2 + a_1x^2y + a_2x^2 + a_3xy^2 + a_4xy + a_5x + a_6y^2 + a_7y + a_8 \quad (15a)$$

$$Y_{Left} = b_0y^2x^2 + b_1y^2x + b_2y^2 + b_3yx^2 + b_5y + b_6x^2 + b_7x + b_8 \quad (15b)$$

$$X_{Right} = a_0x^2y^2 + a_2x^2 + a_3xy^2 + a_4xy + a_5x + a_6y^2 + a_7y + a_8 \quad (15c)$$

$$Y_{Right} = b_1y^2x + b_2y^2 + b_3yx^2 + b_4yx + b_5y + b_6x^2 + b_7x + b_8 \quad (15d)$$

D. Accuracy of candidate models

The average accuracy over 12 participants with three repetitions for each model and calibration targets is shown in Tables V-VII. For the case of 9 calibration targets, model 2222 gives the best result with the average accuracy of 3.11°. But the accuracy is very low. For the rest of the calibration sets, model number 3122 provides the best result for all left and right eyes, and definitely, the average accuracy of 2.50° to 2.28°, respectively.

To ensure the validation process, the best proposed model (model 3122) is compared to the same model before performing the validation. Table VIII presents the accuracy results for 9, 12, 16, 20 and 25 calibration targets. From the table, the full-size model provides less average accuracy than the validated ones for 9 and 25 points of calibration, in particular. Also, it gives the similar results for 12, 16 and 20 points of calibration cases. It confirms that the deleted terms were absolutely insignificant to the model.

Fig. 10-14 demonstrated the distribution of accuracy of our best proposed model against the X and Y positions of the target points in every configuration. Y-axis is reversed because the position of the screen starts at th top-left corner. Accuracy is showed by using the size of a bubble. Bubble with minimum radius indicates an average error at that target point below 2.5°. The positions that contain no bubbles are the points of calibration, no accuracy evaluation at these positions. Bubble represents an average error between 2.5° to 3.5° which is the medium size, while the biggest one represents an average error over 3.5°. The figures show that increasing the calibration points is helping to diminish the errors, and the best results come from 25 points of calibration. In this configuration, most target points have and error that is less than 2.5° and a few points that the errors had an excess of 2.5° locating at the corner areas of the screen.

According to the proposed method, we suggest that model 3122 with 25 calibration points is the best solution for low-cost gaze tracking system as it delivers the best accuracy both in average and distribution.

TABLE VII
AVERAGE ACCURACY (IN DEGREE) OVER 12 PARTICIPANTS WITH THREE REPETITIONS FOR EACH MODEL WITH 25 CALIBRATION TARGETS.

Model	25 calibration targets		
	Average	Left	Right
3131	2.32	2.31	2.33
3122	**2.28**	**2.29**	**2.27**
2231	2.38	2.40	2.36
2222	2.33	2.36	2.30
[19]	2.31	2.30	2.32

Note: The best accuracy for every calibration configuration is boldfaced.

TABLE VIII
AVERAGE ACCURACY (IN DEGREE) COMPARING BETWEEN MODEL 3122 AND THE SAME MODEL BEFORE VALIDATION.

Calibration targets	Our proposed model			Model before validation		
	Average	Left	Right	Average	Left	Right
9	3.75	3.54	3.97	3.84	3.64	4.04
12	2.51	2.50	2.52	2.51	2.49	2.54
16	2.40	2.41	2.39	2.40	2.41	2.40
20	2.33	2.34	2.33	2.33	2.34	2.32
25	2.28	2.29	2.27	2.30	2.30	2.29

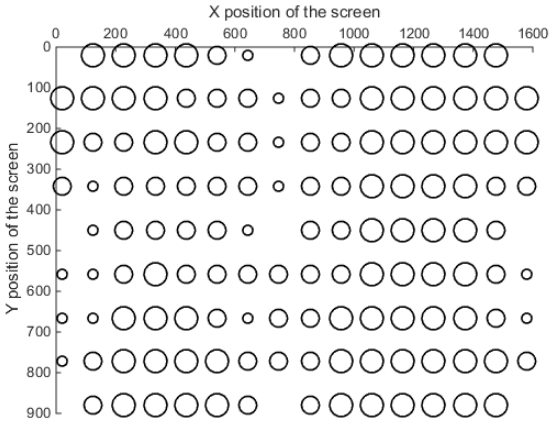

Fig. 10. Distribution of accuracy of model 3122 in 9 points of calibration.

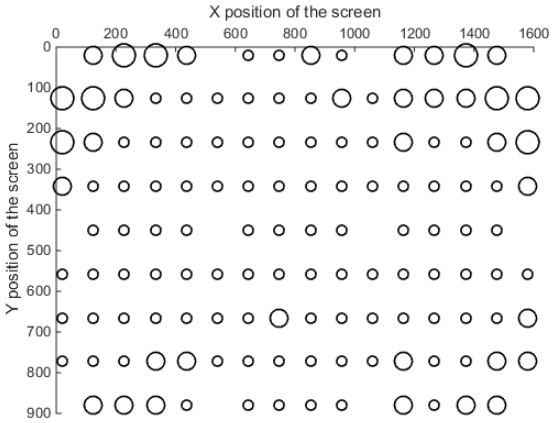

Fig. 11. Distribution of accuracy of model 3122 in 12 points of calibration.

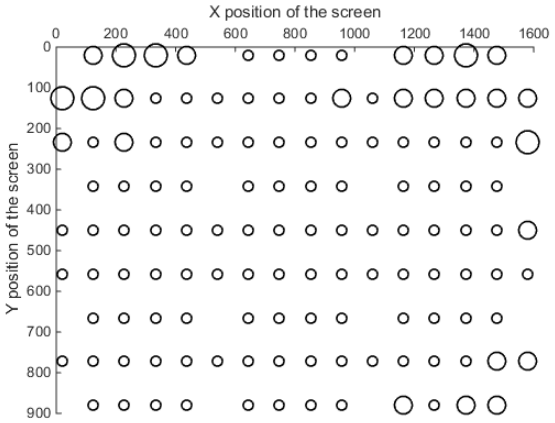

Fig. 12. Distribution of accuracy of model 3122 in 16 points of calibration.

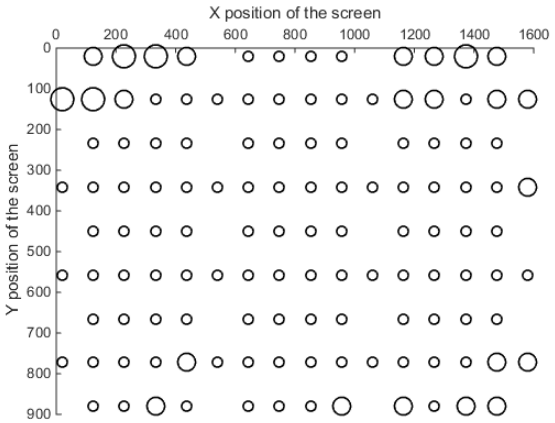

Fig. 13. Distribution of accuracy of model 3122 in 20 points of calibration.

E. Increasing the accuracy by using region estimation

This subsection describes the way to improve the accuracy of the system by considering the area of the screen that the subject is looking at based on the length of pupil-glint vector. This stems from the fact that when the subject looks at the left side of the screen the pupil of the left eye must move to the left side of the eye while the glint deviates to the right side of the eye. In addition, because the eye is spherical, this

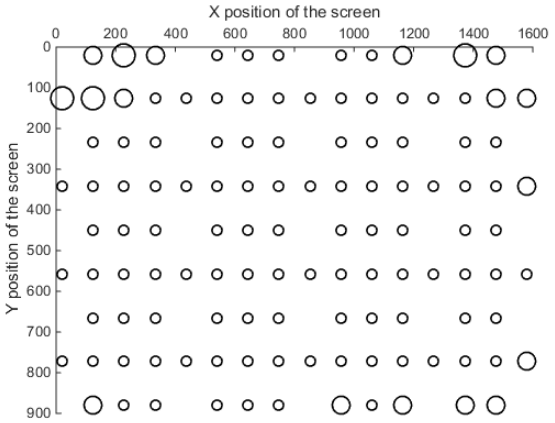

Fig. 14. Distribution of accuracy of model 3122 in 25 points of calibration.

leads to the distortion of the distance of pupil-glint vector of the left eye in the horizontal plan. The distance vector at the right eye in the X-axis, on the other hand, has less distortion. It should be noted that the pupil and glint are on the same side of the eye. In contrast, when the subject gazes at the right side of the screen the distance vector at the left eye is less distorted than from the right eye. These distortions are the cause of errors estimation in gaze mapping.

We increase the accuracy of eye gaze estimation by splitting the gaze mapping into 3 regions, as described below.

- Gaze mapping at the left region of the screen: The eye gaze is estimated by using only the pupil-glint vector from the right eye.
- Gaze mapping at the right region of the screen: The eye gaze is estimated by using only the pupil-glint vector from the left eye.
- Gaze mapping at the middle region of the screen: The eye gaze is estimated by using the average of gaze estimation from both left and right eyes.

In calibration step, we added that this is the process to calculate the average distance of pupil-glint vector in the X-axis, both left and right eyes, when the subject looks at the target points at the left edge and the right edge of the screen. This process was done as an individual experiment. The sweep distances of the left and the right eyes can be estimated by this process. Then, the left region of the screen is determined by the percentage of the sweep distance of the right eye. Similarly, the right region of the screen is defined by the percentage of the sweep distance of the left eye. The other distance value of pupil-glint vector in the X-axis means the eye is gazing at the middle region of the screen. Fig. 15 shows the selection areas of left and right regions. The distances d_1 and d_2 are the sweep distances of left and right eyes, respectively.

Our best proposed model, model 3122 at 25 points of calibration, was used to test with the various percentages of the sweep distances (both left and right eyes) at 5%, 10%, 15%, 20% and 25%, consequently. The accuracy results are presented in Table IX. From the results, cutting left and right areas at 5% of the sweep distance offer the best average accuracy of 1.96°. This is higher than the average accuracy obtained from the previous step.

Figure 16 displays the distribution of accuracy of 5%

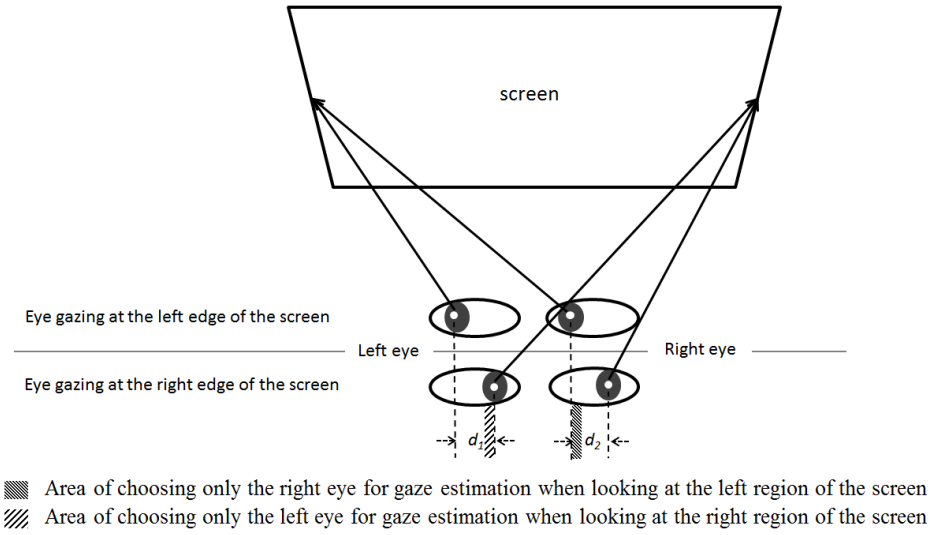

Fig. 15. The selections of gaze mapping at the left and right regions of the screen.

TABLE IX
AVERAGE ACCURACY (IN DEGREE) OF REGION ESTIMATION IN VARIOUS PERCENTAGE OF THE SWEEP DISTANCES.

Percentage (%)	Accuracy
5	1.96
10	1.97
15	1.97
20	1.99
25	2.01

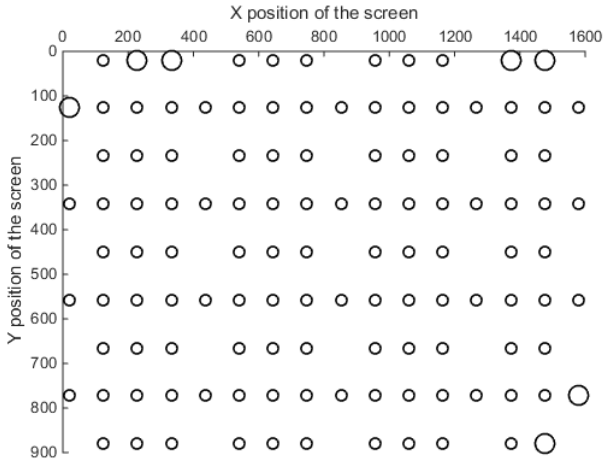

Fig. 16. Distribution of accuracy of region estimation at 5% of the sweep distance.

region estimation against the X and Y positions of the target points in every configurations. This figure shows that the distribution of accuracy had also improved. Most targets have an error that is less than 2.5° and none of them have an error that is greater than 3.5°.

V. CONCLUSIONS

In this paper, we presented a framework for a low-cost gaze tracking system. A model setting and calibration process with confidence illustrated that the system can operate under the best condition according to the operating environment and a user who is using it are clearly explained. We have proven that, for the low-cost gaze tracking system, the reliability of the derived polynomial model can be assured by doing a statistical hypothesis testing. Candidate models were constructed by choosing the equations that provide great relationships between position on the screen and pupil-glint vector. The suitable candidate models for the system contain 8-9 terms because the other models that were more complicated need more calibration targets and this resulted in a problem of eye fatigue. Then, the null hypothesis testing was conducted to ensure that the coefficients of the equation are not significant with the p-value < 0.05 that was performed. Four candidate models, 2222, 2231, 3122 and 3131, were built according to the proposed process and tested for the accuracy of the candidate models. We compared the accuracy with the model from [19] which had been reported that, in the low-cost gaze tracking system, they gave a better accuracy than those of the simple models [21].

Our proposed model (model 3122) provides the greatest performance than that of the champion [19] for most of the calibration targets. The resulting accuracy also suggests that, based on the proposed method, nine calibration targets may not be suitable for low-cost system. We also suggest that the combination of model 3122 with 25 calibration targets is the best for the low-cost gaze tracking system under the normal environment system setting. The time consuming was found to be less than a minute for the calibration process and the obtained model delivered an average accuracy of 2.28°. This accuracy is acceptable for many practical applications. Moreover, we presented a way to increase the accuracy by splitting the gaze estimation into three regions including the left, right and middle areas of the screen. We found that the cutting point at 5% of the sweep distance can improve the accuracy from 2.28° to 1.96°.

Finally, since the calibration time for each model is less than a minute it allows most real-time processing to be possible. This results in the usability of the proposed method, especially when being equipped with a very low price hardware.

REFERENCES

[1] A. T. Duchowski, "A breadth-first survey of eye-tracking applications," *Behavior Research Methods, Instruments, & Computers*, vol. 34, no. 4, pp. 455–470, 2002.

[2] D. W. Hansen and Q. Ji, "In the eye of the beholder: A survey of models for eyes and gaze," *IEEE transactions on pattern analysis and machine intelligence*, vol. 32, no. 3, pp. 478–500, 2010.

[3] S. Sheela and P. Vijaya, "Mapping functions in gaze tracking," *International Journal of Computer Applications.–2011.–26 (3).–36-42*, 2011.

[4] J. San Agustin, H. Skovsgaard, E. Mollenbach, M. Barret, M. Tall, D. W. Hansen, and J. P. Hansen, "Evaluation of a low-cost open-source gaze tracker," in *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications*. ACM, 2010, pp. 77–80.

[5] S. A. Johansen, J. San Agustin, H. Skovsgaard, J. P. Hansen, and M. Tall, "Low cost vs. high-end eye tracking for usability testing," in *CHI'11 Extended Abstracts on Human Factors in Computing Systems*. ACM, 2011, pp. 1177–1182.

[6] N. Schneider, P. Bex, E. Barth, and M. Dorr, "An open-source low-cost eye-tracking system for portable real-time and offline tracking," in *Proceedings of the 1st Conference on Novel gaze-controlled applications*. ACM, 2011, p. 8.

[7] R. Mantiuk, M. Kowalik, A. Nowosielski, and B. Bazyluk, "Do-it-yourself eye tracker: Low-cost pupil-based eye tracker for computer graphics applications," in *International Conference on Multimedia Modeling*. Springer, 2012, pp. 115–125.

[8] R. G. Lupu, F. Ungureanu, and V. Siriteanu, "Eye tracking mouse for human computer interaction," in *E-Health and Bioengineering Conference (EHB), 2013*. IEEE, 2013, pp. 1–4.

[9] L. Sesma, A. Villanueva, and R. Cabeza, "Evaluation of pupil center-eye corner vector for gaze estimation using a web cam," in *Proceedings of the symposium on eye tracking research and applications*. ACM, 2012, pp. 217–220.

[10] Y.-m. Cheung and Q. Peng, "Eye gaze tracking with a web camera in a desktop environment," *IEEE Transactions on Human-Machine Systems*, vol. 45, no. 4, pp. 419–430, 2015.

[11] E. Skodras, V. G. Kanas, and N. Fakotakis, "On visual gaze tracking based on a single low cost camera," *Signal Processing: Image Communication*, vol. 36, pp. 29–42, 2015.

[12] C. H. Morimoto, D. Koons, A. Amit, M. Flickner, and S. Zhai, "Keeping an eye for hci," in *Computer Graphics and Image Processing, 1999. Proceedings. XII Brazilian Symposium on*. IEEE, 1999, pp. 171–176.

[13] J. J. Cerrolaza, A. Villanueva, M. Villanueva, and R. Cabeza, "Error characterization and compensation in eye tracking systems," in *Proceedings of the symposium on eye tracking research and applications*. ACM, 2012, pp. 205–208.

[14] J. San Agustin, H. Skovsgaard, J. P. Hansen, and D. W. Hansen, "Low-cost gaze interaction: ready to deliver the promises," in *CHI'09 Extended Abstracts on Human Factors in Computing Systems*. ACM, 2009, pp. 4453–4458.

[15] Z. Zhu and Q. Ji, "Eye gaze tracking under natural head movements," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, vol. 1. IEEE, 2005, pp. 918–923.

[16] J. J. Cerrolaza, A. Villanueva, and R. Cabeza, "Taxonomic study of polynomial regressions applied to the calibration of video-oculographic systems," in *Proceedings of the 2008 symposium on Eye tracking research & applications*. ACM, 2008, pp. 259–266.

[17] E. D. Guestrin and M. Eizenman, "General theory of remote gaze estimation using the pupil center and corneal reflections," *IEEE Transactions on biomedical engineering*, vol. 53, no. 6, pp. 1124–1133, 2006.

[18] P. Blignaut and D. Wium, "The effect of mapping function on the accuracy of a video-based eye tracker," in *Proceedings of the 2013 Conference on Eye Tracking South Africa*. ACM, 2013, pp. 39–46.

[19] P. Blignaut, "A new mapping function to improve the accuracy of a video-based eye tracker," in *Proceedings of the South African Institute for Computer Scientists and Information Technologists Conference*. ACM, 2013, pp. 56–59.

[20] P. Blignaut, "Mapping the pupil-glint vector to gaze coordinates in a simple video-based eye tracker," *Journal of Eye Movement Research*, vol. 7, no. 1, 2013.

[21] S. Rattarom, N. Aunsri, and S. Uttama, "Validation of the polynomial models in the interpolation based gaze estimation," in *Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), 2015 12th International Conference on*. IEEE, 2015, pp. 1–4.

[22] S. Rattarom, N. Aunsri, and S. Uttama, "Interpolation based polynomial regression for eye gazing estimation: A comparative study," in *International Conference on Digital Arts, Media and Technology (ICDAMT), 2016 1st International Conference on*, 2016.

[23] S. Saenmuang, A. Sirijariyawat, and N. Aunsri, "The effect of moisture content, temperature and variety on specific heat of edible-wild mushrooms: Model construction and analysis," *Engineering Letters*, vol. 25, no. 4, pp. 446–454, 2017.

[24] C. H. Morimoto and M. R. Mimica, "Eye gaze tracking techniques for interactive applications," *Computer vision and image understanding*, vol. 98, no. 1, pp. 4–24, 2005.

[25] A. J. Hornof and T. Halverson, "Cleaning up systematic error in eye-tracking data by using required fixation locations," *Behavior Research Methods, Instruments, & Computers*, vol. 34, no. 4, pp. 592–604, 2002.

[26] P. Blignaut and D. Wium, "Eye-tracking data quality as affected by ethnicity and experimental design," *Behavior research methods*, vol. 46, no. 1, pp. 67–80, 2014.

Suwitchaya Rattarom received his B. Eng. degree in computer engineering from Khon Kaen University, Thailand in 1996, and M. Sc. degree in computer science from Chiang Mai University, Thailand in 2002. He is now a Ph.D. candidate of Computer Engineering, Mae Fah Luang University, Thailand. His current research interests are Eye-Gaze Tracking, Interpolation Based Gaze Estimation and Human Computer Interaction.

Surapong Uttama received his B.S from Chulalongkorn University, Thailand in 1996, M.S. from Asian Institute of Technology, Thailand in 1998, and Ph.D. in Computer Science from University of La Rochelle, France in 2008. He has been working as a lecturer and a researcher at Mae Fah Luang University since 1999. Presently, he is engaged in research on Image Processing, Computer Vision, Content-based Image Retrieval and Mobile and Human Interface Technology.

Nattapol Aunsri received the B.Eng. degree and M.Eng. degree in Electrical Engineering from Khon Kaen University and Chulalongkorn University, Thailand in 1999 and 2003, respectively. He obtained M.Sc. degree in Applied Mathematics and Ph.D. degree in Mathematical Sciences from New Jersey Institute of Technology, Newark, NJ, in 2008 and 2014, respectively. He became a Member (M) of IAENG in 2014.

Since May 2017, he has been an Assistant Professor of Computer Engineering with the School of Information Technology, Mae Fah Luang University, Chiang Rai, Thailand. He is also with the Brain Science and Engineering Innovation Research Unit at Mae Fah Luang University. His research interests include ocean acoustics, Bayesian filtering, signal processing, biomedical signal processing, drug-drug interactions, eye gaze tracking, and mathematical and statistical modeling.