CamType: Assistive Text Entry Using Gaze with an Off-the-shelf Webcam

Yi Liu
Interdisciplinary Graduate School, Nanyang Technological University, Singapore

Bu-Sung Lee, Deepu Rajan
School of Computer Science and Engineering, Nanyang Technological University, Singapore

Andrzej Sluzek
Department of Electrical and Computer Engineering, Khalifa University, Abu Dhabi, UAE

Martin J. McKeown
Department of Medicine, The University of British Columbia, Vancouver, Canada

As modern assistive technology advances, eye-based text entry systems have been developed to help a subset of physically challenged people to improve their communication ability. However, speed of text entry in early eye-typing system tends to be relatively slow due to dwell time. Recently, dwell-free methods have been proposed which outperform the dwell-based systems in terms of speed and resilience, but the extra eye-tracking device is still an indispensable equipment. In this article, we propose a prototype of eye-typing system using an off-the-shelf webcam without the extra eye tracker, in which the appearance-based method is proposed to estimate people's gaze coordinates on the screen based on the frontal face images captured by the webcam. We also investigate some critical issues of the appearance-based method, which helps to improve the estimation accuracy and reduce computing complexity in practice. The performance evaluation shows that eye typing with webcam using the proposed method is comparable to the eye tracker under a small degree of head movement.

Keywords: Assistive technology, Eye-typing system, Dwell-free methods, Appearance-based method

INTRODUCTION

As the human lifespan increases with advances in healthcare technology, the notion of a healthy body also includes the ability to communicate with society in a meaningful way. However, such communication may be impaired due to conditions such as Amyotrophic lateral sclerosis (ALS) and locked-in Syndrome. (Liu, Lee, McKeown, & Lee, 2015). People with these severe motor disabilities may not be able to use even speech properly, and control of the eyes might be the only means to communicate. Eyes play a very important role in our daily life, and receive eighty percent of a human's sensory information (Lu, Sugano, Okabe, & Sato, 2014). The ability to track eye movements is essential for revealing people's physiological states, such as their desires, cognitive processes, and interpersonal relations. Although eye tracking has been applied to marketing, advertisement, and human behaviour analysis, it can now be employed to aid disable people. One of outstanding achievements is that eye-tracking devices have helped some disable people to enter text by controlling a typing system, i.e. eye typing (Majaranta, Ahola, & Špakov, 2009; Sarcar, Panwar, & Chakraborty, 2013; Urbina & Huckauf, 2010; Ward & MacKay, 2002), which provides an effective means of communication. Eye-based typing systems usually apply a virtual keyboard on the screen, take letters that a person has fixated at as inputs, and determine the intended word based on some pattern.

Early eye-typing systems required the person to select each individual letters by fixating at them for a dwell time, so that the speed of text entry is limited (Kotani, Yamaguchi, Asao, & Horii, 2010; MacKenzie & Zhang, 2008; Räihä & Ovaska, 2012). Then, in order to speed up the rate of text entry, dwell-free systems were developed to eliminate dwell time (Kristensson & Vertanen, 2012; Pedrosa, Pimentel, & Truong, 2015; Pedrosa, Pimentel, Wright, & Truong, 2015), where the systems determine the intended word based on use the sequential letters of gaze trace instead of pausing at each individual letter. However, Without dwell time, the input contains a lot of redundant and irrelevant letters due to "The Midas Touch" problem, which causes some common typing errors in practice (Liu, Zhang, Lee, Lee, & Chen, 2015). In order to address the critical issues, a robust eye-typing recognition method was proposed, which improves the robustness of eye typing in practice dramatically (Liu, Lee, & McKeown, 2016). However, the commercial eye-tracking equip-

Correspondence concerning this article should be addressed to Yi Liu, Nanyang Institute of Technology in Health and Medicine, Interdisciplinary Graduate School, Nanyang Technological University, 50 Nanyang Avenue, Singapore. E-mail: yliu028@e.ntu.edu.sg

ment is still indispensable in current eye-typing systems, which is burdensome in some cases, e.g. high cost, inconvenience. Thus, replacing the commercial eye-tracking device with a simple device will increase the flexibility and usability of eye-typing systems. Recently, with advances in computer vision technology, eye tracking with low-resolution face images has been proposed (Tan, Kriegman, & Ahuja, 2002). In addition, since most computers, especially mobile devices, are configured with the webcam, the eye-typing system using webcam would be a promising option.

Therefore, in this article, we proposed an eye-based typing system using standard webcam, in which the appearance-based method shows the feasibility of being used to estimate gaze points of people in the system. We then investigate two critical issues of appearance-based gaze estimation method. Since the captured eye images at calibration session contain invisible artefacts that cause large estimation error in subsequent gaze estimation, we propose an averaging filter to multiple captured images, so that the performance in terms of estimation accuracy improves through reducing the impact of artefacts. We study the effective features of the eye area that are highly related to the determination of gaze points, which achieves dimensionality reduction of the appearance space with maintaining the estimation accuracy at decreasing computational cost and time complexity. We then conduct eye-typing experiments of actual users. The performance evaluation in terms of word recognition rate indicates that it is feasible to develop a webcam-based eye-typing system using the proposed gaze estimation method with the robust recognition algorithm.

RELATED WORK

Eye Typing

Eye tracking research can be traced back to the late 18th century that the saccadic eye movements of people were investigated while reading by using the first eye-tracking device (Huey, 1908). At that time, the major task of eye-tracking research was physiological cognition analysis (Jacob & Karn, 2003; Ohno, 2007). More recently, with the development of computer system and eye-tracking technology, researchers have gradually focused on human-computer interaction, in which the system uses the eye-tracking device as an interaction interface, and takes eye movements as input signals to control the applications (Murata, 2006; Zander, Gaertner, Kothe, & Vilimek, 2010). Eye tracking then started to play an important role in assisting physically challenged people (Adjouadi, Sesin, Ayala, & Cabrerizo, 2004; Kocejko, Bujnowski, & Wtorek, 2009; Su, Wang, & Chen, 2006), and also help to conduct a more efficient communication for people with severe impairment who cannot even speak properly, e.g. Amyotrophic Lateral Sclerosis (ALS) or Locked-in Syndrome (Caligari, Godi, Guglielmetti, Franchignoni, & Nardone, 2013; Spataro, Ciriacono, Manno, & La Bella, 2014). As a promising assistive technique, eye typing allows people to enter text by just moving their eyes, so that it provides an effective means of communication (Majaranta & Räihä, 2002).

Due to the disadvantage of the dwell-based technique, the speed of early eye-typing systems is a crucial limitation (Kotani et al., 2010; Räihä & Ovaska, 2012). With adjustable dwell time (Majaranta et al., 2009) or modified keyboard layout (Sarcar et al., 2013; Urbina & Huckauf, 2010), the dwell-based typing systems have improved a lot already, but they still suffer from the bottleneck of fixation time. In order to solve the critical limitation, the dwell-free concept was proposed (Kristensson & Vertanen, 2012) which allows the person to enter an intended word by just looking at the letters of the word in order instead of letter-by-letter entry with dwell time, which is much faster than other dwell-based systems (Kristensson & Vertanen, 2012; Pedrosa, Pimentel, & Truong, 2015; Pedrosa, Pimentel, Wright, & Truong, 2015).

However, although the users do not need to pause at individual letters for a while, they are still required to look at all letters of the typing word. If some letter of the typing word is missing, the systems fail to recognize it. To solve the problem, Liu *et al.*(2015,2016) proposed two robust recognition algorithms in the dwell-free system, Moving Window String Matching (MoWing) and Longest Common States Mapping algorithm (LCSMapping) algorithms, respectively. With the robust algorithms, the eye-typing system can recognize the typing word according to the pattern of sequential letters that the users have gazed at, which is able to handle the common typing errors successfully. In addition, using a low-accuracy eye tracker, the dwell-free typing system still showed the good performance to text entry error. Nevertheless, it still requires a commercial eye tracker which prevents eye typing from becoming a pervasive technology. Nowadays, since most computers, especially mobile devices, are configured with the webcam, using webcam would increase the flexibility and usability of eye-typing systems. In addition, with advances in computer vision technology in face detection and recognition, eye tracking directly using original eye appearance images has been proven to be feasible with considerable performance accuracy (Lu, Sugano, Okabe, & Sato, 2011; Tan et al., 2002).

Gaze Estimation

In general, image-based eye tracking research refers to two areas: *eye localization* and *gaze estimation* (Hansen & Ji, 2010). The main task of eye localization is detecting the existence of eye features in the image, and interpreting the positions of the eye areas. The eye (centre) localization also usually refers to the pupil or iris centre detection, and the traditional eye localization (eye areas) can be a prerequisite. A lot of techniques have been proposed (Kim, Choi, Shin,

& Ko, 2015; Skodras & Fakotakis, 2015; Valenti & Gevers, 2012; P. Wang, Green, Ji, & Wayman, 2005; Zhou & Geng, 2004; Zhu & Ji, 2005), which are robust in complicated environment, under various eye appearances and illumination, and even with occlusion. Instead of only localizing where the eyes are in face images, gaze estimation is determining what a person is looking at using detected eyes in images or videos (Hansen & Ji, 2010), 3-D gaze direction or 2-D gaze point. Generally, the two terms “eye tracking" and “gaze tracking" are inter-changeable, where gaze tracking is a process of consecutive gaze estimations from frame to frame in a video or real time.

The gaze of a person is determined by both head pose (i.e. position, orientation) and eyeball orientation. To change gaze direction, the person can move the head while keeping the eye stationary, or only move the eyes. Usually eye movement and head movement occur simultaneously, where head pose determines the coarse-grained gaze area, and the eyeball orientation determines the fine-grained gaze point. Therefore, the model of gaze estimation needs to take both head pose and eye position into account, which is very challenging in the research filed. In our study, due to the limited head motion of the physically disabled people, we ignore head movement modelling but only focus on the eye area. Basically, there are two categories of gaze estimation, feature-based and appearance-based estimation (Hansen & Ji, 2010).

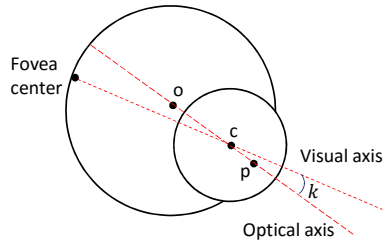


Figure 1. The 3D eye structure.

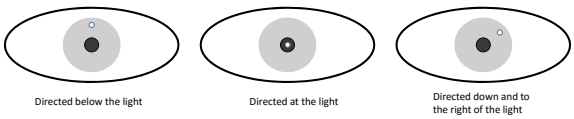


Figure 2. The relation of pupil and glints

The feature-based gaze estimation methods are modelling the relation between local eye features and gaze points, e.g. eye corners, iris/pupil centre, and corneal reflections (glints) from the eye image. There are two subtypes in feature-based approaches, 3D-model-based (geometric) methods and interpolation-based (regression-based) methods. the 3D-model-based methods (Chen & Ji, 2011; Villanueva, Cabeza, & Porta, 2007; J.-G. Wang, Sung, & Venkateswarlu, 2005) directly estimate the centres of eyeball and iris based on a geometric eye model (figure 1), and then obtain the visual axis (gaze direction) of the person. Since the angular offset of visual and optical axes is invariant to each person, gaze direction is usually determined by the estimated optical axis instead. However, in practice, it is necessary to obtain the 2-D gaze point. To avoid additional calculation of point of regard (PoR), the interpolation-based methods (Brolly & Mulligan, 2004; Ebisawa & Satoh, 1993; Morimoto & Mimica, 2005; Williams, Blake, & Cipolla, 2006) directly find the underlying mapping from the eye image features to the gaze coordinates. The relation of pupil and glints (cornea reflection) is the most popular and widely used for gaze estimation under active light models (Cerrolaza, Villanueva, & Cabeza, 2008) shown as figure 2. In addition, the pupil centre-eye corners vector is also regarded as a good feature with acceptable accuracy (Sesma, Villanueva, & Cabeza, 2012). However, the feature-based methods usually require highly-accurate detection of local features, and thus a high-quality camera or even infra-red light is necessary, which is infeasible in outdoor environment.

Instead of explicitly extracting local eye features, appearance-based approaches directly take the eye-area image as a high-dimensional vector, and then resemble the interpolation-based methods to train a regression function that maps it to 2-D gaze coordinates (Lu et al., 2011, 2014; Tan et al., 2002). The features to train a regression function are image features (pixel intensity, or other features which can describe the whole image), rather than the position of exact eye features. Instead of high accurate local eye features, appearance-based methods take low-resolution images as inputs, where an off-the-shelf webcam can be used which is more flexible, so that the appearance-based methods have been becoming a popular gaze estimation technique.

GAZE ESTIMATION FROM EYE APPEARANCE

Without highly-accurate features detection, appearance-based gaze estimation does not require the high quality of the input source, but usually utilizes a low-resolution eye-area image as the input. Each pixel of the image is described as one dimension in the appearance space, so that the entire image can be described as a feature point in the high-dimensional space, where the dimensionality is the same as the number of pixels in the image. The basic assumption of the appearance-based methods is that although these image points are in the high-dimensional space, they still follow the distance rule in Euclidean space. The two similar images would be two corresponding points close to each other in the high-dimensional space, and vice versa. Based on the assumption, the relative similarity among images will be represented by the relative distance of corresponding points in the high-dimensional space, as shown in figure 3, where we ignore the minor variance of iris and pupil size at different directions.

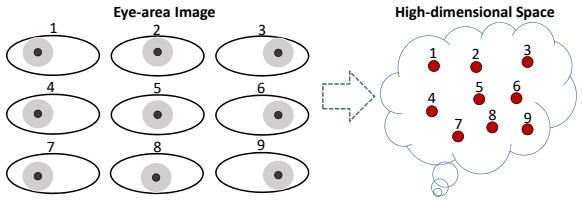


Figure 3. The relative relation of images and points. left-hand side indicates the eye-area images that looking at nine different positions, while right-hand side indicates the corresponding points in the high-dimensional space.

The "movement" of iris area is the main factor of variation of the eye appearance. The eye movement is rotating around the eyeball centre, so that iris area is moving in the 2-D surface which is embedded into the 3-D space. According to the properties of the Euclidean space above, the corresponding points of the eye images will form a manifold in the eye appearance space that has an approximately 2-D surface, so that the relative similarity would be maintained inside, and the 2D gaze points of eye images are sharing the same local interpolation with corresponding image points in the appearance space (figure 4).

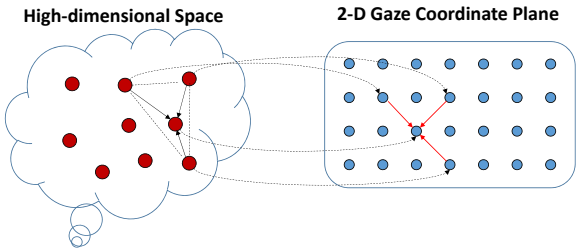


Figure 4. The mapping from high-dimensional space to gaze coordinate plane.

Eye-Appearance Feature Vector Extraction

The eye-appearance feature vector is extracted in the following steps.

1. *Eye-region image cropping* (figure 5). The first step is to extract the eye-region images. For the efficiency of computing and image processing, it is common to convert RGB scale image to gray scale. Initial eye areas (left and right) are detected by using Haar-cascade classifier (Viola & Jones, 2004). In order to guarantee the correct eye-area extraction, we use two methods to validate the detection result. 1). Eye should be contained in the face area. The whole face area is detected using Haar-cascade classifier, and the margin between eye area and face area should be in the predefined interval. 2). The average intensity of eye area should be lower than the average of the face area. The valid detection result is taken as the rough eye region.

2. *Feature vector generation* (figure 6). The extracted rough eye-region image usually contains the eyebrow, which is irrelevant to the gaze. Eyeball and eyebrow areas are two darkest parts in the image. For eyebrow removal, we apply the horizontal integrated projection of the image, which is adding up the pixel intensity of each row, and removing the upper-peak area. Then, we further crop the eye area in terms of the outer and inner eye corners detected by Canny edge filter (Lu et al., 2011), and the eye image is resized at the fixed aspect ratio, 60×36 pixels as suggested in (X. Zhang, Sugano, Fritz, & Bulling, 2015). Since the pixels of an image is represented by numerical values in computer system, the image can also be considered as a matrix of pixel intensity. Through a raster scan of the matrix, we obtain the final feature vector of the eye appearance.

Each element in the feature vector is corresponding to a pixel of the cropped eye-area image, i.e. the image is fully described by a 2160-dimensional vector, and the image can also be represented by a point in the appearance space. Due to two degrees of freedom of iris-area movement in the images, the corresponding points would move across a 2-D surface embedded into the high-dimensional space, which constitutes a manifold, called eye-appearance manifold.

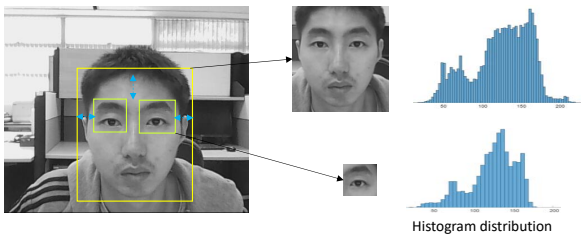


Figure 5. The rough eye-region image extraction

Eye-Appearance Manifold Verification

As we mentioned above, although the "image points" are in a high dimensional space, the constituted manifold of these points has an approximately 2D surface. Each point of an eye appearance on the 2D-surface will correspond to a gaze point that a person is looking at on the screen. Therefore, if the person is looking at distinct letters on a virtual keyboard, appearance points of eye images should be also differentiable, which is a prerequisite of recognizing which letters the person intends to enter using the eye image captured by the webcam.

To verify the practicability, we conducted a simple experiment. A participant was asked to sit in front of a monitor

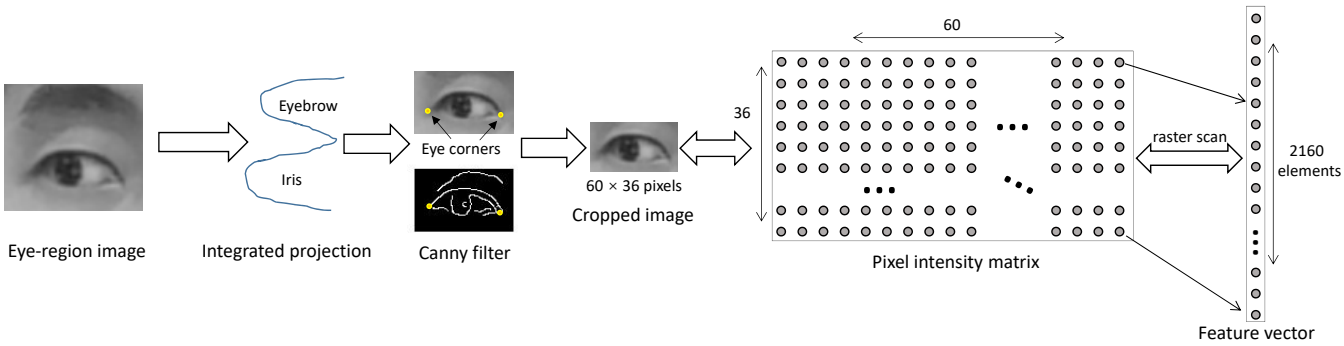


Figure 6. Eye-appearance feature vector generation.

Figure 7. The 26-letter keyboard layout

with webcam, where the monitor displays an on-screen virtual keyboard as shown in figure 7 that was used in (Liu, Lee, et al., 2015; Liu et al., 2016; Liu, Zhang, et al., 2015). To guarantee limited head movement, he was asked to move the head to a comfortable position first and then only rotate eyeball to gaze at every letter sequentially on the virtual keyboard. while looking at a key exactly, he clicked a “confirm" button, and then the webcam would capture fifty consecutive frontal face images within a short time. All images were be processed by using the above extraction steps, to generate the eye-appearance feature vectors. For a better visualization, we then used principal component analysis (PCA) to project these high-dimensional vectors into 2-D/3-D space.

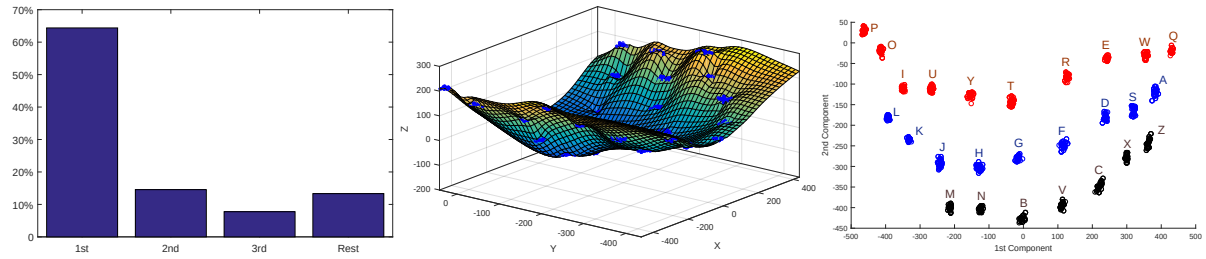
Figure 8a shows the percentage of eigenvalues of the transformed data using PCA. The first three components contain around 90% of accumulated eigenvalues, which means only using the three dimensions will well represent the data distribution in original 2160-dimensional space with little information loss. This observation verified the previous statement that the image points are embedding into a low-dimensional manifold, though they are in the high-dimensional space. Figure 8b plots the data distribution in first three principal components (3-D space), and the low-dimensional manifold of points has an approximately 2-D-surface shape. Figure 8c plots the first two principal components, where there are 26 clearly-separated clusters. The eye-appearance points looking at different letters are differentiable in 2-D space, which can be taken as a classification problem. In addition, the more important observation is that the relative positions of the clusters are the same as the keyboard layout, which means the linearity is existing in the manifold. The classification problem can be further converted to a regression problem, and gaze estimation becomes feasible in practice.

Instead of explicitly extracting the mapping function between local features (e.g. pupil centre, eye corners, or glints) and gaze points, the appearance-based methods implicitly extract the hidden mapping relations between the eye images and gaze coordinates. As we can see, the pattern that the eye appearance changes in high dimensional space is like the variation of gaze points on the 2-D screen, and the local linearity of the two spaces shows the relevance.

Manifold Learning Method

Appearance manifold is a continuous set of eye-appearance feature points embedding in the high dimensional space, and any point in the manifold can be linearly interpolated by its local neighbours. The task of manifold learning is reducing the dimensionality, i.e. extracting the low-dimensional manifold from the high-dimensional space, and the relative local geometric position should be maintained. In the practical learning problem, training samples are a set of high-dimensional data (eye appearance vectors) with corresponding low-dimensional data (2-D gaze coordinates), and estimating the gaze coordinate of a new eye appearance vector. Basically, there are two types of the learning, *global* and *local* learning.

Global learning is directly learning the global parameter between eye appearance vectors and corresponding gaze coordinates, and using the same parameter to determine the gaze coordinate of a new eye appearance. This method only estimates the global parameter using training samples once, but it cannot guarantee the invariance of relative position of data points in the two spaces. Therefore, we work with the local learning in appearance-based gaze estimation. Rather than directly learning the global transformation parameter



(a) (b) (c)

Figure 8. The PCA projection. (a) shows the percentage of eigenvalues. (b) plots image points distribution with the corresponding key labels in 2-D space formed by the first two principal components. (c) plots the 2-D-surface manifold in 3-D space formed by the first three principal components.

from appearance data to the gaze coordinates, local learning is finding the local parameter that reconstructs the given new appearance vector $\widehat{\boldsymbol{x}}$ using training appearance vectors $\{\boldsymbol{x}_i\}$, and then applying the same local parameter to infer the gaze coordinate $\widehat{\boldsymbol{p}}$ using $\{\boldsymbol{p}_i\}$ of $\{\boldsymbol{x}_i\}$:

$$\begin{aligned} \widehat{\boldsymbol{x}} &= \boldsymbol{X}\boldsymbol{w} \\ \widehat{\boldsymbol{p}} &= \boldsymbol{P}\boldsymbol{w} \end{aligned} \tag{1}$$

Where $\boldsymbol{X} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_k\}$, $\boldsymbol{P} = \{\boldsymbol{p}_1, \boldsymbol{p}_2, ..., \boldsymbol{p}_k\}$, and $\boldsymbol{w} = (w_1, w_2, ..., w_k)^T$ denote the locally linear combination, which avoids directly estimating the mapping between $\boldsymbol{X}$ and $\boldsymbol{P}$. The local combination will guarantee the relative positions of $\widehat{\boldsymbol{x}}$ and $\widehat{\boldsymbol{p}}$ are the same in their respective spaces. However, since the equation is overdetermined, we have the optimization function with minimizing the estimation error:

$$\begin{aligned} \widetilde{\boldsymbol{w}} &= \arg\min_{w} \left\| \widehat{\boldsymbol{x}} - \sum_{i}^{k} w_i \boldsymbol{x}_i \right\| \quad s.t. \sum_{i}^{k} w_i = 1 \\ \widehat{\boldsymbol{p}} &= \sum_{i}^{k} w_i \boldsymbol{p}_i \end{aligned} \tag{2}$$

Where $\widetilde{\boldsymbol{w}}$ is optimal locally linear combination parameter minimizing the linear combination error.

Then without loss of generality, let $\widehat{\boldsymbol{x}}^L$, $\widehat{\boldsymbol{x}}^R$ denote the left and right eye appearance vectors of the frontal face image; $\boldsymbol{X}^L$, $\boldsymbol{X}^R$ denote the neighbouring appearance vectors of $\widehat{\boldsymbol{x}}^L$ and $\widehat{\boldsymbol{x}}^R$; $\boldsymbol{w}^L$, $\boldsymbol{w}^R$ denote corresponding linear combination parameters. Thus, the final reconstruction is given by:

$$\begin{aligned} \widetilde{\boldsymbol{w}}^L &= \arg\min_{w} \left\| \widehat{\boldsymbol{x}}^L - \boldsymbol{X}^L \boldsymbol{w}^L \right\|, \ s.t. \ \mathbf{1}^T \boldsymbol{w}^L = 1 \\ \widetilde{\boldsymbol{w}}^R &= \arg\min_{w} \left\| \widehat{\boldsymbol{x}}^R - \boldsymbol{X}^R \boldsymbol{w}^R \right\|, \ s.t. \ \mathbf{1}^T \boldsymbol{w}^R = 1 \end{aligned} \tag{3}$$

The gaze estimation is determined by:

$$\begin{aligned} \widehat{\boldsymbol{p}}^L &= \boldsymbol{P}^L \widetilde{\boldsymbol{w}}^L, \widehat{\boldsymbol{p}}^R = \boldsymbol{P}^R \widetilde{\boldsymbol{w}}^R \\ \widehat{\boldsymbol{p}} &= (\widehat{\boldsymbol{p}}^L + \widehat{\boldsymbol{p}}^R)/2 \end{aligned} \tag{4}$$

Where $\widehat{\boldsymbol{p}}$ is average estimated gaze points of the linear combination of both left and right eye images. Note that $\boldsymbol{X}^L$ and $\boldsymbol{X}^R$ are in different sets, so the functions are optimized in respective spaces. The accuracy is evaluated by the mean estimated angular error:

$$error = \frac{1}{n} \sum_{i=1}^{n} \arctan\left(\frac{\left\|\widehat{p}_i - p_i\right\|_2}{d}\right) \tag{5}$$

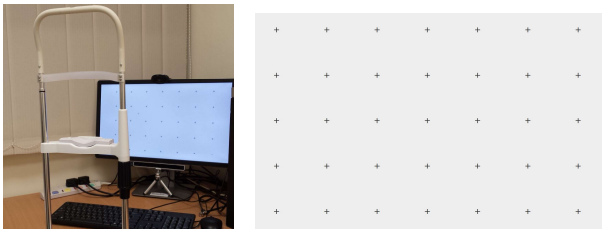
where $\left\|\widehat{p}_i - p_i\right\|_2$ denotes the Euclidean distance between real gaze coordinate p_i and estimated gaze coordinate $\widehat{p}_i$, and d denotes the average distance from user's eyes and the screen.

PRACTICAL ISSUES

Training data of interpolation-based methods are the most important to infer a new sample. In the gaze estimation method, training data are obtained from the calibration phase. Thus, guaranteeing a high-quality calibration is very crucial for the accuracy of gaze estimation. In addition, dimensionality reduction is also important part for a learning method to avoid the "curse of dimensionality". Therefore, in this work, we will study the above two issues: 1). improving the gaze estimation by reducing artefacts of calibration images. 2). reducing the image size by detecting the effective areas of eye image in determining the gaze point.

We conducted a calibration experiment with a desktop computer which displayed 35 cross-hair markers as shown in figure 9. There are two devices in the experiment, 1). the webcam (Logitech C170, 30fps) was used to capture the frontal face image. 2). The eye tracker (The EyeTribe, 30Hz) was used to provide estimated gaze points as control samples. In order to fixate the head, the chin rest was put in front of the monitor. Three university students participated in the experiment. They were asked to look at each marker on the screen. To better fixate at the marker centre, they moved the mouse cursor (the same shape with marker) to align with the marker. The participants clicked on the mouse key once the mouse cursor overlaps with the marker. The system then recorded three types of data, the marker's position, the estimated gaze

coordinates reported by the EyeTribe, and ten consecutive images captured by the webcam.



(a) (b)

Figure 9. (a) shows the experiment setup. (b) shows the layout of 35 cross-hair markers.

Artefacts Reduction

In the appearance-based method, the images of calibration are determining gaze points the subsequent new images. If the calibration images have the bias caused by artefacts while doing the calibration, it would lead to large estimation error in gaze estimation. In previous work (Lu et al., 2011; Tan et al., 2002), the system captured only one face image of participants while looking at each calibration point. Although the images are captured in a well-controlled environment, there are inevitable factors that degrade the quality of the calibration, such as instant fine-grain eye motion and illumination variance. The artefacts might cause a bias of captured images while looking at the same calibration point. In order to improve the quality of calibration, reducing the impact of artefact is necessary at the calibration stage.

In the experiment, our system captured multiple face images (10 images) while the participant was looking at each of 35 calibration points as shown in figure 10a. In order to compare with the performance of capturing one image of calibration, we select only one image from each set. This selection method would generate 10^{35} image combinations which is too large, and for computing efficiency, we choose 10,000 image combinations of them, i.e. randomly selecting 35 images from 35 sets respectively (figure 10b), and repeating 10,000 times. For reducing the impact of artefacts, we proposed an averaging filter to each set of images (figure 10c).

To evaluate the performance of the two methods, we used the leave-one-out cross validation. The average estimation error (angular) was taken as the performance metric in equation 5. Figure 11 shows the evaluation results of three participants. The first and second bars indicate the average error of 10,000 image combinations and minimum error of using best individual images, and the third bar indicates applying the averaging filter. We also calculated the error of gaze points estimated by the eye tracker. As we can see, the average estimation error of using individual image is 0.2 degree higher than using the proposed averaging filter. The minimal error of using individual images is still 0.1 degree higher. After applying the averaging filter, in subject 1 and 3, the proposed appearance-based method even outperforms the commercial eye tracker. The observation also verified that the artefacts of calibration images have great impact in the gaze estimation. Although there is no guarantee that the averaging filter is better than any "best" one of 10^{35} image combinations, the proposed filter is reducing the error by 20% from the average combination, and thus it is still a good method to minimize the impact of artifacts, and improve the estimation accuracy.

Effective Area Detection

In the appearance-based gaze estimation methods, there is a common sense that the gaze status is determined by the appearance of eye area, rather than other face features, e.g. nose, mouth. In general, the eye-area image is cropped at a predefined aspect ratio in terms of outer and inner eye corners. However, there is a fundamental issue that has never been studied, which is how the different sizes of eye images affect gaze estimation. Without addressing this issue, the current cropping method might not be efficient. To further crop the eye area, we propose three cropping operations as shown in figure 12. The horizontal operation is cropping both top and bottom rows of the eye image, and the vertical operation is cropping both left and right columns. The full operation is cropping pixels of four directions.

In the experiment, we still used the leave-one-out cross validation to evaluate the performance of each cropping. In figure 13, the x-coordinate indicates the number of cropping times. For example, the number of horizontal cropping is 1, which means removing a row of both top and bottom pixels of the image respectively. The y-coordinate denotes the angular estimated error. As we can see, while horizontally cropping the images, the error lines remain smooth until the "knee" point that the upper eyelid is removed, and then they are increasing dramatically. For vertical cropping, although the eye corners and even a part of the iris area are removed, the error lines do not increase dramatically.

The observations indicate the different effective features determining the gaze in appearance-based methods. In feature-based gaze estimation methods, eye corners are taken as the second important features besides the pupil centre. However, in appearance-based methods, eye corners are not important features, and the upper eyelid plays an important role that affects the gaze estimation. The importance of different eye features also reveals the mechanism of feature-based and appearance-based methods. Feature-based methods mainly focus on the relation of pupil centre and eye corners, and the eye corner is taken as a reference. However, appearance-based methods make use of the appearance variance of the eye area. Besides the movement of the iris part, the eyelid openness also affects the appearance, but the appearance of eye corner is relatively invariant. These observa-

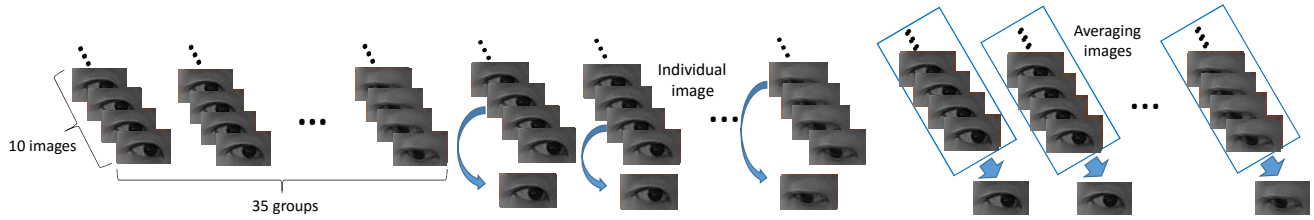


Figure 10. Individual images and averaging images

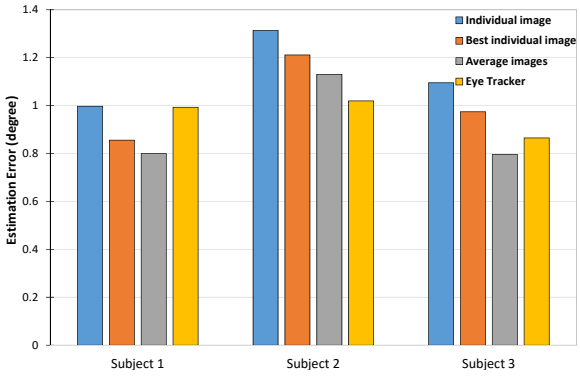


Figure 11. Evaluation results

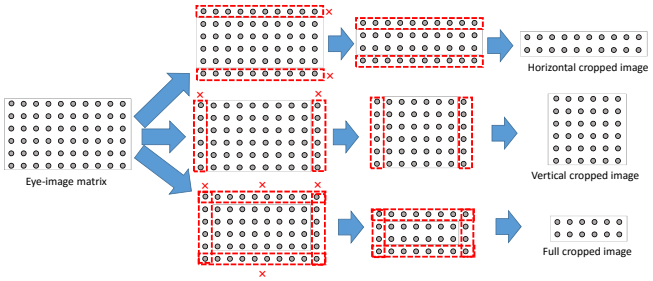


Figure 12. Cropping operations

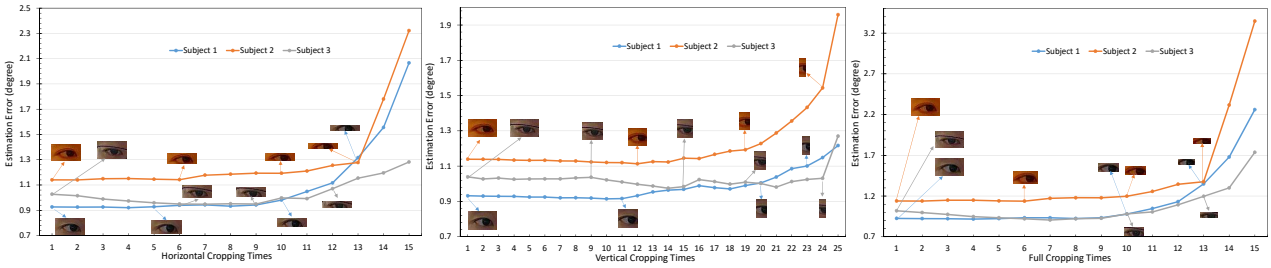
tions also help to improve the current cropping method that the inner and outer eye corners are not necessary to be contained, which is to further reduce the size of eye-area image. In figure 13c, despite ten-time full cropping, i.e. the size of eye image becomes 40×16 pixels from 60×36 pixels, the estimation error does not increase comparing with the original image. Meanwhile, the dimension of the appearance feature vector is also reduced from 2160 to 640, which decreases computational cost and time complexity.

PRACTICAL EYE TYPING EXPERIMENT

In the experiment, we developed an eye-typing system in desktop computer, which is configured with a 23-inch monitor. The placement of equipments is as follows: 1). The off-the-shelf webcam was attached on the top of the monitor; 2). the eye tracker was put under the monitor; 3). the chin rest was mounted on the desktop, and the distance between the chin rest and the monitor is 50cm. The layout of on-screen virtual keyboard is the same as previous work. Five volunteers (4 male, 1 female, 25 - 28 years old) who participated in the eye-typing experiment (Liu et al., 2016) were invited in the experiment, and thus all of them had prior experience with eye typing.

The experiment procedures are as follows: 1). the participants put the head upon the chin rest and moved to a comfortable position. They were asked to try to keep the head stationary, and only rotate the eyeball. 2). Then they were instructed to do the nine-point calibration with the eye tracker. 3). They were required to type 120 words using gaze with the eye-typing system. The words were randomly selected one by one from 5000 commonly-used words (Davies, 2011), and displayed in the top of keyboard which is the same as the previous work (Liu et al., 2016). The 120-words typing was divided into six consecutive sessions. In each session, the participants first did the calibration with the 26-letter keys, where the system recorded the pairs of the face images captured by the webcam while they were looking at each key, and the corresponding keys' coordinates. After doing the calibration, the participants would type 20 random words by sequentially gazing at the letter keys of the given word. while typing each word, the webcam captured the images at 30 frames per second, which is the same as the sampling frequency of the eye tracker. After completing all six sessions, there are six sets of calibration images, and 120 sets of images typing words.

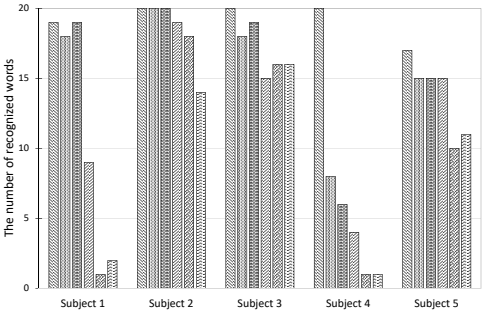
All gaze points of word-typing images were estimated by the calibration images. For the purpose of comparison, we applied two scenarios to estimate the gaze points of the images, 1). the calibration images of the only first session were used to estimate all 120 sets of word-typing images; 2). the calibration images of each session were used to estimate word-typing images in the same session (20 sets), respectively. These estimated points were collected in accordance with the time series. To evaluate the performance of eye typing using the webcam, we take recognition accuracy of intended words as the metric. A robust eye-typing algorithm,



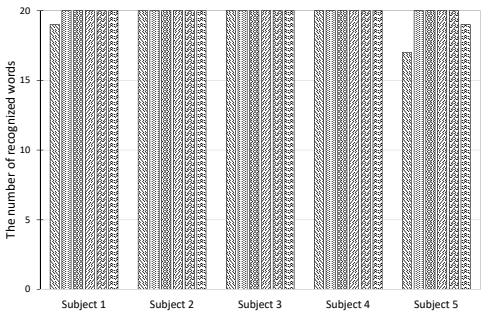
(a) Horizontal cropping (b) Vertical cropping (c) Full cropping

Figure 13. Evaluation results in different cropping times.

LCSMapping (Liu et al., 2016), was used to recognize the word in terms of the sequential gaze points, where it would rank all possible words in the dictionary by the probability. In the eye typing system, the top-k rate is usually used as the accuracy metric, i.e. the intended word is ranked in the top-k words (Liu, Lee, et al., 2015; Liu, Zhang, et al., 2015). In our experiment, we used the top-5 rate as suggested in (Liu et al., 2016), i.e. if top five candidate words contain the intended word, it is regarded as successfully recognition.



(a) using the first calibration



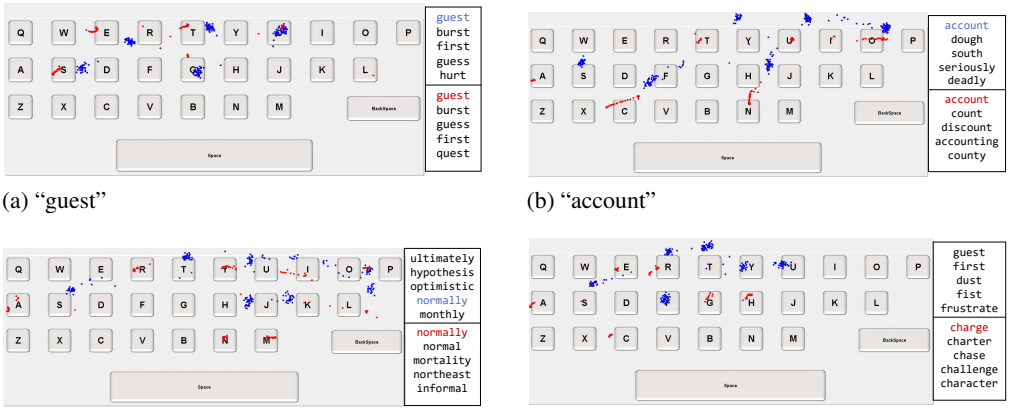
(b) using individual calibration

Figure 14. Recognition results.

Figure 14 shows the recognition results of 120 words of five subjects. In figure 14a, there are the results that calibration images in the only first session are used to estimate gaze coordinates of images in subsequent five sessions. The x-axis indicates the five subjects, where in each subject, the six bars indicate the six sessions of typing in order. The y-axis indicates the number of words that are successfully recognized. As we can see, in the first session, almost all twenty words can be recognized, the number of recognized words decreases over time (sessions). While using the calibration images of each session to estimate gaze coordinates of images in the same session, almost all words are successfully recognized. In the experiment, using the gaze coordinates estimated by the eye tracker, all words can be successfully recognized. For the better understanding of the quality of gaze estimation, we plot the estimated gaze points of typing several typical words with the virtual keyboard in figure 15. The blue dots and red dots represent the estimated points of our proposed method and the eye tracker, respectively. Along with the virtual keyboard, the upper and lower right box list the top five words that are recommended by the LCSMapping algorithm using “blue" points and “red" points respectively, and the intended words are highlighted.

Due to the smoothing algorithm of the eye tracker, the neighbouring points within a predefined threshold would be integrated into the same coordinates, and thus the red dots look concentrated and fewer. As we can see, the estimated points of the eye tracker is more accurate that the red dots cover all letters of the four words which are ranked at the first place. The point distribution (blue dots) of the four words represents four typical cases using the appearance-based method, 1). the nearest letters of these points contain all letters of the intended word; 2). there is a small drift between the points and the intended letters, i.e. some nearest letters of the points are neighbours of the intended letters. Since the LCSMapping algorithm can handle the neighbour-letter error well (the three types of errors refer to Liu et al. (2016).), the intended word is still ranked at the first place; 3). along with the neighbour-letter error caused by the drift, many other irrelevant letters are also included which cause the extra-letter error, and thus the priority of the intended word decreases, but it is still in the top five words. 4), as the points keep drifting, the missing-error is also existing which might make the recognition algorithm fail to recommend the intended word.

The above observations help us understand that the drift of



(a) “guest” (b) “account”

(c) “normally” (d) “charge”

Figure 15. Estimated gaze points of typing words

estimated points (or calibration drift) is the critical issue of word recognition in eye typing using the appearance-based method. When we only use the first calibration, the performance of recognition gets poorer over time, in which the main reason is the calibration drift, i.e. the variation of the status between typing words and the calibration phase is becoming larger. While doing the calibration every once in a while (recalibration), the drift can be controlled within limits (figure 15b, 15c), and the performance of recognition approximates the eye tracker as shown in figure 14a. Therefore, eye typing with the webcam is becoming feasible even under some degree of the calibration drift.

DISCUSSION

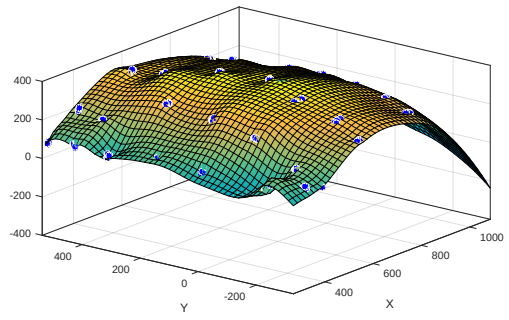
The basic idea of appearance-based gaze estimation method is to infer the gaze coordinate of a new face image based on the calibration images and their corresponding gaze coordinates. In machine learning filed, it is a common supervised learning algorithm to solve the regression problem. In general, the class (continuous value) of a new sample is usually predicted in terms of the global mapping parameter from the sample space to the target space, in which the dimension of the sample space depends on the number of attributes, and the target space is usually one or two dimensions. The global mapping emphases the inter-relation of two spaces, but misses the intra-correlation within the sample space.

In the appearance-based gaze estimation method, all attributes are generated from pixels of eye image (figure 6), and they have a strong spatial correlation. As we mentioned before, these “image” points constitutes a 2D-surface manifold in the sample space (figure 8b), where the local linear combination is existing within the manifold. Moreover, the relative positions of points are consistent with the degree of eyeball rotation. The points in the sample space (high-dimensional space) might share the same local relation with the target space (low-dimension space, or 2-D gaze coordinate), which has been verified well already. In order to guarantee the same local relation, recently, Yu et al.(2016) have further unified the structures of sample space and target space using a metric learning. Therefore, given a new eye image, the core step is to obtain the local combination parameter of its neighbouring training images, and then the optimal local parameter is employed to infer the new gaze coordinate using the combination of corresponding gaze coordinates of training images.

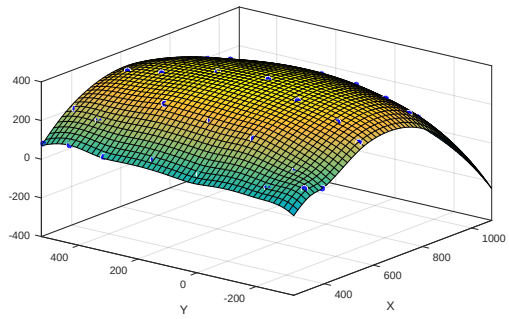
In general, the training images are collected through a calibration session. Since the subsequent gaze estimation replies on the training set, the quality of calibration is very crucial for estimation. The different number of calibration points has been already studied (Lu et al., 2011, 2014), in which the more calibration points (training images) usually improve the accuracy in subsequent gaze estimation. From the view of the sample space, the structure of manifold formed by training samples determines the gaze process. If we have enough training samples including all possible degrees of eyeball rotation, there would be no estimation error. In practice, it is infeasible to require the person to do the long-term calibration, which is burdensome. Since the range of the target space is usually finite, e.g. a screen, the boundary of the manifold is definite. Thus, the reference points of calibration are uniformly selected across the plane (figure 9b), which is a common method in calibration including existing commercial eye trackers.

Although the reference points in the target plane define the size and shape of manifold, the final manifold is formed by the appearance feature vectors of captured eye images. Thus, the quality of the eye images might also change the structure of manifold and then to affect the accuracy of gaze estimation. The traditional method is capturing the only one face image while the person is looking at each reference point, which might be deteriorated by uncontrollable artefacts. In order to reduce the impact of the artefacts, the averaging filter is employed to the multiple images captured at each refer-

ence point, which can improve the accuracy of estimation using the individual image (figure 11). As we mentioned above, the accuracy of the new sample estimation is the performance of the interpolation using training samples, which depends on the structure of the manifold. In figure 16, we plot the 3-D projection of the manifolds formed by individual images and average images using PCA. Although the structure of both manifolds is homogeneous, local texture is different. The manifold of the average images show a better approximation of the 2-D surface, and the interpolation becomes more accurate. Thus, the averaging filter is improving the estimation performance through smoothing the manifold, so that the linear combination of neighbouring sample is better supported. While the reference points in the target plane determine the coarse-grained shape of manifold, the eye images determine the fine-grained shape.



(a) all individual images



(b) the average images

Figure 16. Eye-appearance manifolds in 3D space.

As we mentioned before, each dimension (attribute) of the sample space is described by the pixels of the eye image. In fact, since the eye image has the specific pattern, e.g. the relative distance between eye features, these dimensions have a strong spatial correlation. Lu et al.(2011) and Yu et al.(2016) proposed the method of dimensionality reduction that is dividing the eye image into several small blocks, and taking the sum of pixel intensities in each block, in which the dimension of the target space decreases to the number of blocks. Although the method might require a shorter calibration session, the accuracy of estimation decreases because of information loss. Therefore, to achieve dimensionality reduction without information loss, i.e. reducing the dimensions that are irrelevant to determine the gaze point, the intuition of appearance-based method need to be further studied. In figure 17, we plot the weight of directly mapping from the sample space (eye image) to the target space (gaze point). Although we do not use the global mapping to estimate gaze points, it might help us understand the rough relation between the appearance and gaze point. We randomly select 10 to 500 eye images ($\boldsymbol{X}$) of word typing with their corresponding gaze coordinates ($\boldsymbol{P}$, estimated points using the proposed method above). Since the dimension (2160) of the sample space is much larger than the number (k) of selected eye images, the global mapping function $\boldsymbol{wX=P}$ is a underdetermined system, and L1 norm is used to obtain the weight $\boldsymbol{w}$ with k non-zero components which is highly related to determine gaze points $\boldsymbol{P}$. We will find that the area of eye orbit is contributing to the mapping, especially the edge area, e.g. eye lid and the boundary of iris. Thus, cropping other irrelevant pixels might not cause information loss, and also maintain the accuracy.

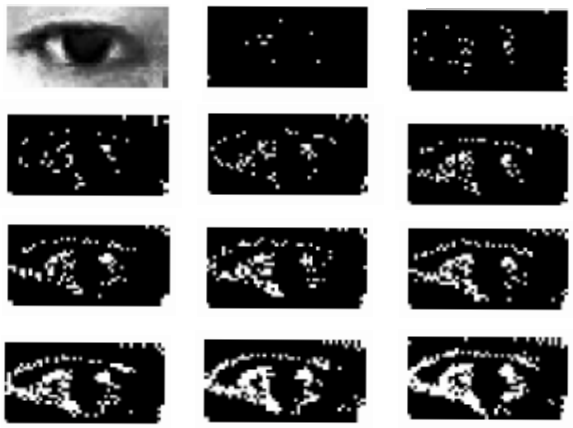


Figure 17. The effective area

When we use the calibration images to estimate a new image, the condition is that these images should be in the same manifold, i.e. sharing the same local structure. However, if the head pose changes, the condition is not satisfied. The same reference point on the plane will correspond to different eye appearance images, and vice versa. Thus, under each different head pose, the eye images will constitute a new manifold, in which the head pose is determining the location of manifold in the sample space, and the eye image is determining the position of the corresponding appearance vector in the manifold (Sugano, Matsushita, Sato, & Koike, 2008). While doing practical eye typing using the webcam, the critical issue is the calibration drift. Although the six sessions were consecutive and participants were required to keep the head stationary all the time, there might be also slight head

motion. For the better understanding of head movement, We use the head tracker in OpenFace toolkit (Baltru, Robinson, Morency, et al., 2016) to estimate the head pose of each face image, which outputs 6-dimensional head pose vectors. After processing the head pose vector, we can obtain the average head displacement of participants while typing each word in order as shown in figure 18. The displacement is increasing over the time, which causes the variation of eye images looking at the same coordinate. Since the relative position of the eye images in the manifold is still maintained, the new estimated points would demonstrate a drift of the actual points (figure 15). In the eye-typing system, there are three common typing errors, extra-letter, neighbour-letter, and missing-letter errors (Liu et al., 2016; Liu, Zhang, et al., 2015). As the points are keeping drifting, more missing-letter errors will deteriorate the accuracy of the recognition algorithm. The purpose of recalibration (calibrate in each session) is to control the drift of the points updating the manifold under a new head pose. Thus, eye typing using webcam is feasible under a small degree of head movement.

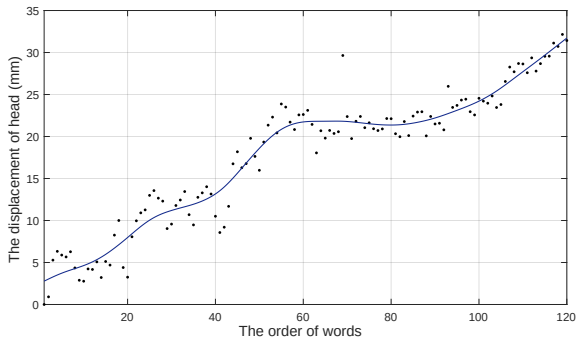


Figure 18. The displacement of head

However, there is still a challenging issue of eye typing using the webcam. To control the degree of the calibration drift, the system requires the person to do periodic recalibration, which brings extra burden. Zhang et al.(2014) and Vadillo et al.(2015) proposed the automatic recalibration methods based on the eye tracking data, but the methods are not suitable for the real-time system. Recently, along with the prevalence of deep learning, some research work also attempted to introduce deep neural network into gaze estimation training with millions of face images (Krafka et al., 2016; X. Zhang et al., 2015). Since the underlying head pose information has been automatically extracted from the big data set, it is possible that people do not even have to do calibration any more. However, it requires the high-performance machine, and our daily computer is not practical to train the model.

Besides the gaze estimation method, the recognition algorithm still need to improve to reduce the impact of the points drift. Since the relative position of points is maintained, the shape of gaze trace is homogeneous. Thus, it is possible to use shape-based approach which is comparing the shape of gaze trace of typing with the actual word shape (Hoppe, Löchtefeld, & Daiber, 2013; Kristensson & Zhai, 2004). Moreover, the recognition algorithm we used in our work is only focusing on the input of the person. If a prediction engine can be integrated into the eye-typing system with semantic analysis and user historical typing behaviour, it would also improve the accuracy of word recommendation in practice.

CONCLUSION AND FUTURE WORK

Eye-based text entry systems have been proposed to help the communication ability of a subset of physically challenged people. With the proposal of the dwell-free concept, the eye-typing systems allow people to type a word by gazing at the intended letter keys sequentially, i.e. focusing on the whole word entry like swipe typing rather than letter-by-letter selection, so that the speed and usability have improved significantly. Then robust recognition algorithms in eye-typing system have been proposed, e.g. MoWing (Liu, Zhang, et al., 2015) and LCSMapping (Liu et al., 2016), which make the system more resilient to common typing errors even using the low-accuracy eye-tracking device. However, the additional eye tracker is still a necessary device of the eye-typing system, which might be burdensome in some case.

Therefore, in our work, we have proposed an eye-typing system using an off-the-shelf webcam to replace the eye tracker. First, we compared two types of gaze estimation methods using the webcam, in which the appearance-based method is more promising, because of its advantages in processing low-resolution images. We then found that the eye-appearance feature vectors would constitute a 2-D manifold embedded into the appearance space, and these feature vectors are clearly differentiated in the manifold. We further studied practical issues of the appearance-based method, where an averaging filter was proposed to improve the gaze estimation accuracy through reducing the impact of uncontrollable artefacts. Through detecting the effective eye area determining the gaze variance, we achieved the purpose of dimensionality reduction without information loss. In the eye-typing experiment of actual users, although the proposed method is not as accurate as the commercial eye tracker, the robust typing algorithm still successfully recognized intended words according the estimated gaze points. With periodic recalibration controlling a small degree of head movement, the performance of word recognition using the webcam is approximating the eye tracker.

However, although the recalibration can effectively handle the impact of head movement, people are still required with extra efforts to do it regularly which might limit the speed of typing. In addition, the proposed appearance-based method is only modelling the eye area without other face features, so

that it might lose potential information of head pose. Therefore, in the future work, we will first focus on the challenging issue that is reducing the number of recalibration. Since the relative shape of gaze trace does not change despite of the calibration drift, the shape-based recognition methods can be integrated into the exiting eye-typing system. With semantic analysis in terms of the context, the prediction engine will also improve the quality of word recommendation by narrowing the search space. Regarding the appearance-based method, besides the eye area, we will also study the impact of the face area determining the gaze point with natural head movement. Furthermore, head pose estimation using the face image might help to handle large head movement and remove the calibration session, which is out of scope of this article, but still suggests a good future direction for our work. Due to the medical policy restrain, currently we are not allowed to invite the target patients in the experiment. For a more efficient system, some target users should be recruited to participate in the future work to explore new demands.

ACKNOWLEDGEMENT

This work is a collaboration with the Joint NTU-UBC Research Centre of Excellence in Active Living for the Elderly (LILY).

References

Adjouadi, M., Sesin, A., Ayala, M., & Cabrerizo, M. (2004). *Remote eye gaze tracking system as a computer interface for persons with severe motor disability*. Springer Berlin Heidelberg.

Baltru, T., Robinson, P., Morency, L.-P., et al. (2016). Openface: an open source facial behavior analysis toolkit. In *2016 ieee winter conference on applications of computer vision (wacv)* (pp. 1–10).

Brolly, X. L., & Mulligan, J. B. (2004). Implicit calibration of a remote gaze tracker. In *Computer vision and pattern recognition workshop, 2004. cvprw'04. conference on* (pp. 134–134).

Caligari, M., Godi, M., Guglielmetti, S., Franchignoni, F., & Nardone, A. (2013). Eye tracking communication devices in amyotrophic lateral sclerosis: Impact on disability and quality of life. *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration*, *14*(7-8), 546–552.

Cerrolaza, J. J., Villanueva, A., & Cabeza, R. (2008). Taxonomic study of polynomial regressions applied to the calibration of video-oculographic systems. In *Proceedings of the 2008 symposium on eye tracking research & applications* (pp. 259–266).

Chen, J., & Ji, Q. (2011). Probabilistic gaze estimation without active personal calibration. In *Computer vision and pattern recognition (cvpr), 2011 ieee conference on* (pp. 609–616).

Davies, M. (2011). *Word frequency data from the corpus of contemporary american english (coca).*

Ebisawa, Y., & Satoh, S.-I. (1993). Effectiveness of pupil area detection technique using two light sources and image difference method. In *Engineering in medicine and biology society, 1993. proceedings of the 15th annual international conference of the ieee* (pp. 1268–1269).

Hansen, D. W., & Ji, Q. (2010). In the eye of the beholder: A survey of models for eyes and gaze. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, *32*(3), 478–500.

Hoppe, S., Löchtefeld, M., & Daiber, F. (2013). EypeâĂŤusing eye-traces for eye-typing. In *Workshop on grand challenges in text entry (chi 2013).*

Huey, E. B. (1908). *The psychology and pedagogy of reading*. The Macmillan Company.

Jacob, R., & Karn, K. S. (2003). Eye tracking in human-computer interaction and usability research: Ready to deliver the promises. *Mind*, *2*(3), 4.

Kim, S.-T., Choi, K.-A., Shin, Y.-G., & Ko, S.-J. (2015). A novel iris center localization based on circle fitting using radially sampled features. In *Consumer electronics (isce), 2015 ieee international symposium on* (pp. 1–2).

Kocejko, T., Bujnowski, A., & Wtorek, J. (2009). Eye-mouse for disabled. In *Human-computer systems interaction* (pp. 109–122). Springer.

Kotani, K., Yamaguchi, Y., Asao, T., & Horii, K. (2010). Design of eye-typing interface using saccadic latency of eye movement. *International Journal of Human-Computer Interaction*, *26*(4), 361–376.

Krafka, K., Khosla, A., Kellnhofer, P., Kannan, H., Bhandarkar, S., Matusik, W., & Torralba, A. (2016). Eye tracking for everyone. In *Proceedings of the ieee conference on computer vision and pattern recognition* (pp. 2176–2184).

Kristensson, P. O., & Vertanen, K. (2012). The potential of dwell-free eye-typing for fast assistive gaze communication. In *Proceedings of the symposium on eye tracking research and applications* (pp. 241–244).

Kristensson, P.-O., & Zhai, S. (2004). Shark 2: a large vocabulary shorthand writing system for pen-based computers. In *Proceedings of the 17th annual acm symposium on user interface software and technology* (pp. 43–52).

Liu, Y., Lee, B.-S., McKeown, M., & Lee, C. (2015). A robust recognition approach in eye-based dwell-free typing. In *Proceedings of 2015 international conference on progress in informatics and computing* (pp. 5–9).

Liu, Y., Lee, B.-S., & McKeown, M. J. (2016). Robust eye-based dwell-free typing. *International Journal of Human-Computer Interaction*, *32*(9), 682-694. doi: 10.1080/10447318.2016.1186307

Liu, Y., Zhang, C., Lee, C., Lee, B.-S., & Chen, A. Q. (2015). Gazetry: Swipe text typing using gaze. In *Proceedings of the annual meeting of the australian special interest group for computer human interaction* (pp. 192–196).

Lu, F., Sugano, Y., Okabe, T., & Sato, Y. (2011). Inferring human gaze from appearance via adaptive linear regression. In *Computer vision (iccv), 2011 ieee international conference on* (pp. 153–160).

Lu, F., Sugano, Y., Okabe, T., & Sato, Y. (2014). Adaptive linear regression for appearance-based gaze estimation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, *36*(10), 2033–2046.

MacKenzie, I. S., & Zhang, X. (2008). Eye typing using word and

letter prediction and a fixation algorithm. In *Proceedings of the 2008 symposium on eye tracking research & applications* (pp. 55–58).

Majaranta, P., Ahola, U.-K., & Špakov, O. (2009). Fast gaze typing with an adjustable dwell time. In *Proceedings of the sigchi conference on human factors in computing systems* (pp. 357–360).

Majaranta, P., & Räihä, K.-J. (2002). Twenty years of eye typing: Systems and design issues. In *Proceedings of the 2002 symposium on eye tracking research & applications* (pp. 15–22). New York, NY, USA: ACM. Retrieved from http://doi.acm.org/10.1145/507072.507076 doi: 10.1145/507072.507076

Morimoto, C. H., & Mimica, M. R. (2005). Eye gaze tracking techniques for interactive applications. *Computer Vision and Image Understanding*, *98*(1), 4–24.

Murata, A. (2006). Eye-gaze input versus mouse: Cursor control as a function of age. *International Journal of Human-Computer Interaction*, *21*(1), 1–14.

Ohno, T. (2007). Eyeprint: Using passive eye trace from reading to enhance document access and comprehension. *International Journal of Human–Computer Interaction*, *23*(1-2), 71–94.

Pedrosa, D., Pimentel, M. d. G., & Truong, K. N. (2015). Filteryedping: A dwell-free eye typing technique. In *Proceedings of the 33rd annual acm conference extended abstracts on human factors in computing systems* (pp. 303–306).

Pedrosa, D., Pimentel, M. D. G., Wright, A., & Truong, K. N. (2015). Filteryedping: Design challenges and user performance of dwell-free eye typing. *ACM Transactions on Accessible Computing (TACCESS)*, *6*(1), 3.

Räihä, K.-J., & Ovaska, S. (2012). An exploratory study of eye typing fundamentals: dwell time, text entry rate, errors, and workload. In *Proceedings of the sigchi conference on human factors in computing systems* (pp. 3001–3010).

Sarcar, S., Panwar, P., & Chakraborty, T. (2013). Eyek: an efficient dwell-free eye gaze-based text entry system. In *Proceedings of the 11th asia pacific conference on computer human interaction* (pp. 215–220).

Sesma, L., Villanueva, A., & Cabeza, R. (2012). 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* (pp. 217–220). New York, NY, USA: ACM. Retrieved from http://doi.acm.org/10.1145/2168556.2168598 doi: 10.1145/2168556.2168598

Skodras, E., & Fakotakis, N. (2015). Precise localization of eye centers in low resolution color images. *Image and Vision Computing*, *36*, 51–60.

Spataro, R., Ciriacono, M., Manno, C., & La Bella, V. (2014). The eye-tracking computer device for communication in amyotrophic lateral sclerosis. *Acta Neurologica Scandinavica*, *130*(1), 40–45.

Su, M.-C., Wang, K.-C., & Chen, G.-D. (2006). An eye tracking system and its application in aids for people with severe disabilities. *Biomedical Engineering: Applications, Basis and Communications*, *18*(06), 319–327.

Sugano, Y., Matsushita, Y., Sato, Y., & Koike, H. (2008). An incremental learning method for unconstrained gaze estimation. In *European conference on computer vision* (pp. 656–667).

Tan, K.-H., Kriegman, D. J., & Ahuja, N. (2002). Appearance-based eye gaze estimation. In *Applications of computer vision, 2002.(wacv 2002). proceedings. sixth ieee workshop on* (pp. 191–195).

Urbina, M. H., & Huckauf, A. (2010). Alternatives to single character entry and dwell time selection on eye typing. In *Proceedings of the 2010 symposium on eye-tracking research & applications* (pp. 315–322).

Vadillo, M. A., Street, C. N., Beesley, T., & Shanks, D. R. (2015). A simple algorithm for the offline recalibration of eye-tracking data through best-fitting linear transformation. *Behavior research methods*, *47*(4), 1365–1376.

Valenti, R., & Gevers, T. (2012). Accurate eye center location through invariant isocentric patterns. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, *34*(9), 1785–1798.

Villanueva, A., Cabeza, R., & Porta, S. (2007). Gaze tracking system model based on physical parameters. *International Journal of Pattern Recognition and Artificial Intelligence*, *21*(05), 855–877.

Viola, P., & Jones, M. J. (2004). Robust real-time face detection. *International journal of computer vision*, *57*(2), 137–154.

Wang, J.-G., Sung, E., & Venkateswarlu, R. (2005). Estimating the eye gaze from one eye. *Computer Vision and Image Understanding*, *98*(1), 83–103.

Wang, P., Green, M. B., Ji, Q., & Wayman, J. (2005). Automatic eye detection and its validation. In *Computer vision and pattern recognition-workshops, 2005. cvpr workshops. ieee computer society conference on* (pp. 164–164).

Ward, D. J., & MacKay, D. J. (2002). Artificial intelligence: Fast hands-free writing by gaze direction. *Nature*.

Williams, O., Blake, A., & Cipolla, R. (2006). Sparse and semi-supervised visual mapping with the sˆ 3gp. In *2006 ieee computer society conference on computer vision and pattern recognition (cvpr'06)* (Vol. 1, pp. 230–237).

Yu, P., Zhou, J., & Wu, Y. (2016). Learning reconstruction-based remote gaze estimation. In *Proceedings of the ieee conference on computer vision and pattern recognition* (pp. 3447–3455).

Zander, T. O., Gaertner, M., Kothe, C., & Vilimek, R. (2010). Combining eye gaze input with a brain–computer interface for touchless human–computer interaction. *International Journal of Human-Computer Interaction*, *27*(1), 38–51.

Zhang, X., Sugano, Y., Fritz, M., & Bulling, A. (2015). Appearance-based gaze estimation in the wild. In *Proceedings of the ieee conference on computer vision and pattern recognition* (pp. 4511–4520).

Zhang, Y., & Hornof, A. J. (2014). Easy post-hoc spatial recalibration of eye tracking data. In *Proceedings of the symposium on eye tracking research and applications* (pp. 95–98).

Zhou, Z.-H., & Geng, X. (2004). Projection functions for eye detection. *Pattern recognition*, *37*(5), 1049–1056.

Zhu, Z., & Ji, Q. (2005). Robust real-time eye detection and tracking under variable lighting conditions and various face orientations. *Computer Vision and Image Understanding*, *98*(1), 124–154.

ABOUT THE AUTHORS

Yi Liu is a Ph.D. student of computer science at Interdisciplinary Graduate School, Nanyang Technological University. His research spans Brain-Computer Interaction, Human-Computer Interaction, Eye Tracking, and Computer Vision. He obtained his B.Eng. from School of Software, Harbin Institute of Technology, China.

Bu-Sung Lee is currently an Associate Professor with the School of Computer Science and Engineering, Nanyang Technological University. He held a joint position as Director(Research) HP Labs. Singapore from 2010 till 2012. His major research interests are in the area of Mobile and Pervasive Network, Distributed systems and Cloud/Grid Computing Technology.

Andrzej Sluzek is currently an Associate Professor with the Department of Electrical and Computer Engineering, Khalifa University (Abu Dhabi, UAE). From 1992 to 2011, he worked at Nanyang Technological University (School of Computer Engineering and Robotic Research Centre). His research interests include Machine Vision, Intelligent Robotics and Applications of Signal Processing.

Deepu Rajan is an Associate Professor in the School of Computer Science and Engineering at Nanyang Technological University, Singapore. He received his Bachelor of Engineering degree in Electronics and Communication Engineering from Birla Institute of Technology, Ranchi (India), M.S. in Electrical Engineering from Clemson University (USA) and Ph.D from Indian Institute of Technology, Bombay (India). He is a member of IEEE. His research interests include image processing, computer vision and multimedia signal processing.

Martin J. McKeown is a Professor of Medicine and ECE (Adjunct), Director of the Pacific Parkinson's Research Centre at UBC, and holds the PPRI/UBC Chair in Parkinson's Research. In addition to seeing patients with Movement Disorders, is developing new analytical methods to assess brain imaging data, and investigating novel treatments for Parkinson's Disease.