Computing and Informatics, Vol. 42, 2023, 1213–1232, doi: 10.31577/cai_2023_5_1213

LOCATION ESTIMATION FROM AN INDOOR SELFIE

Mengqi Du, Yue Zhang

College of Computer Science and Technology
Zhejiang University of Technology
Hangzhou, China
e-mail: mengqidu@foxmail.com, 1111712013@zjut.edu.cn

Jianhua Zhang

School of Computer Science and Engineering
Tianjin University of Technology
Tianjin, China
e-mail: zjh@ieee.org

Honghai Liu

School of Mechanical Engineering and Automation
Harbin Institute of Technology
Shenzhen, China
e-mail: honghai.liu@icloud.com

Abstract. With the development of social networks and hardware devices, many young people have post a lot of high definition v-logs containing selfie images and videos to commemorate and share their daily lives. We found that the reflected image of corneal position in the high definition selfie image has been able to reflect the position and posture of the selfie taker. The classic localization works estimating the position and posture from a selfie are difficult because they lack the knowledge of the environment. The corneal reflection images inherently carry information about the surrounding environment, which can reveal the location, posture and even height of the selfie taker. We analyze the corneal reflection imaging process in the selfie scenario and design a validation experiment based on this process to

estimate the pose of the selfie in several scenarios to further evaluate the leakage of the pose information of the selfie taker.

Keywords: Corneal imaging system, location estimation, privacy disclosure, selfie, social network

Mathematics Subject Classification 2010: 65-D19

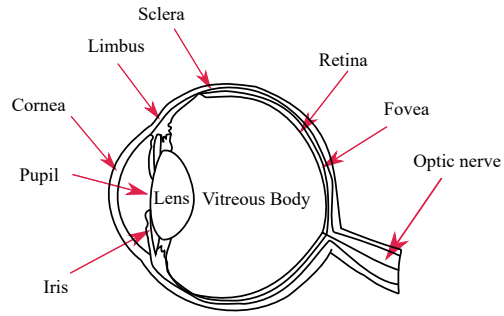
1 INTRODUCTION

Posting v-logs documenting and sharing life on social network sites is very popular among young people. However, selfie photos and videos often reveal the shooting time, environment, location, and even the habits of the shooter. In addition, as the quality of camera imaging has improved, the corneal reflection of the photographer's surroundings is much clearer than before. The existing works have demonstrated that corneal reflecting image can show the gaze and intention of the subject, and even analyze the environment in which the subject is located.

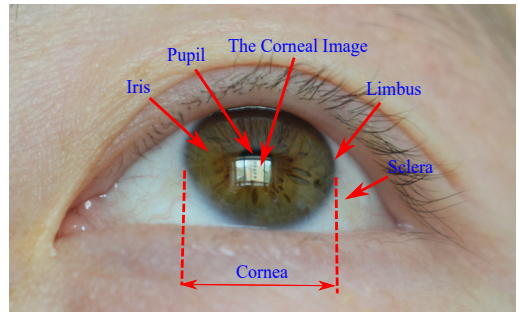
After a lot of observations, we found that the corneal reflecting images can even reveal the position and posture of the selfie taker. These information may reveal the selfie shooting habit even the height of the body. In this paper, we analyze the corneal imaging process during selfie, and experimentally verify and analyze that there is a certain correspondence between the corneal reflecting images and the position, posture of the selfie taker.

The cornea is a transparent semi-ellipsoidal structure located at the very front of the eye and plays a protective role for the eye, as shown in Figure 1 a). Incident rays shoot through the cornea, pass through the pupil, lens, and vitreous body to reach the fovea on the retina, where the light signal is converted into an electrical signal then transmitted by the optic nerve to the visual center of the brain, so that the body can perceive its surroundings visual information. The corneal surface is covered with a thin, reflexible tear film, when the incident rays pass through the cornea, a small amount of light will be reflected by the tear film. These reflected light can be captured by the human eye or optical device such as a camera, then form an image, this image is called corneal image (CI). The incident light – the corneal reflective surface – the reflected light capture device together form the corneal imaging system (CIS).

Since a CI can reflect the human surroundings, the CIS has extensive research and application prospects in the fields of human-machine interaction, computer graphics, disaster rescue, indoor and outdoor localization, and so on. However, there are still difficulties in the research of CIS. Because the cornea is a transparent structure, only a limited amount of incident light can be reflected back into the camera, creating a CI with little details and texture. In addition, the CI is always polluted by the color and texture of the iris in the eye image.



a) The anatomical view of eye



b) The frontview of eye

Figure 1. The structure of eye

Despite the poor imaging quality, however, it has been found from daily observation that the CIS is sensitive to high-light scenes: the high-light incident rays will be projected as the brightness pattern on the corneal surface, as shown in Figure 1 b). Therefore, many works based on the CIS require additional pieces of equipment, such as an infrared camera to determine the location of the high-light pattern in the CI, or a reliable light source, such as a bright, high-contrast light source as input to improve the quality of CI.

When taking indoor selfies, in order to obtain better imaging quality, the selfie taker often chooses a scene with better light distribution, such as facing a window or a screen being projected, so that an indoor high-light element can be projected as a bright pattern on the cornea. The projected pattern shape will change along the corneal pose transformation, we then can estimate the pose variation of the photographer. As an innovative and challenging work, in this paper, we will test this innovative hypothesis from both theoretical and experimental aspects.

2 RELATED WORK

2.1 Monocular Camera Based Localization

The monocular RGB camera positioning works enlighten our work. By recording the center displacement of the camera and simultaneously extracting and matching the affine invariant features between frames (e.g., SIFT [1], corner point [2]), or with the help of the scene geometry knowledge [3] (e.g. parallel lines, surfaces), to achieve the estimation of the change of the target object's pose. Monocular RGB cameras are widely used in AR [4, 5], visual SLAM [6, 7, 8], 3D reconstruction [9, 10], and other fields. The way the camera estimates the scene depth relies heavily on the feature matching between frames. The localization accuracy will be poor when the input light is weak. To improve the localization accuracy of monocular cameras, existing works try to find effective matching features among the matchable feature clusters with algorithms such as RANSAC [11], or artificially introduce distinctive markers in the scene, such as AprilTag [12], ARToolkit [13] to improve the feature extraction and matching accuracy. However, since the CI contain little texture, and details and lacks a sufficient number of effective feature points, the above methods cannot be applied in the localization method of this paper.

2.2 The CIS Researches and Applications

The subject's behavior and awareness can be inferred from the environmental images reflected from the cornea. This is the force that drives the research of the CIS. In order to analyze the light distribution around the subject, Tsumura et al. [14] calculate the source of light by analyzing the light spot reflected from the cornea to reconstruct the face model. Nishino and Nayar [15] view the cornea as a light probe to percept the light distribution of the scene then relighting the given 3D face.

To improve the imaging quality of CIS, Wang et al. [16] propose a CIS separation algorithm with two eye images as input. Nitschke and Nakazawa [17], based on the corneal and eye models through the super-resolution has proposed a method for CI enhancement.

Nishino and Nayar [18] reference and extend the work of Swaminathan et al. [19] to the field of CIS, and fully explain the relationship between scenes, corneas and cameras involved in CIS. As a complementary work to Nishino and Nayar [18], Nitschke et al. [20] propose a calibration method for CIS based on an infrared camera and an LED light array. Based on the above theory of CIS, Suda et al. [21] propose an algorithm for matching the CI to the scene in order to avoid the calibration process. Nakazawa et al. [22] propose a gaze-tracking algorithm based on the CIS using a bendable LED dot matrix system assisted by an infrared camera. Lander et al. [23] propose a work for computing 3D gaze from the 2D gaze with the help of infrared cameras and scene cameras. Ohshima et al. [24] try to match the CI with the scene pictures in the database by neural network, showing the possible privacy security risk and the prospect of human-machine interaction for CIS. Du et al. [25]

propose a gaze tracking method based on CIS with the help of an AprilTag marker, which makes it possible to use gaze as an AR interaction method.

As a catadioptric imaging system, the CIS has been developed in the past two decades, and its main objectives are focused on assisting gaze-tracking and the analysis and recognition of reflection scenes. As an important component of CIS, the posture of the cornea is critical to the imaging process, and in addition, the posture of the cornea is often closely related to the posture of the person. However, there is less existing related work. By analyzing the relationship between scene, corneal pose and head pose, we try to verify the feasibility of CIS-based indoor localization and show the possibility of privacy leakage risk of selfies in social network in this paper.

3 THEORETICAL ANALYSIS OF CORNEAL IMAGING PROCESS DURING SELF-TIMER

The imaging process of the CIS is shown in Figure 2. The incident ray $\boldsymbol{v}_i \in V_I$ is reflected by the cornea as reflected light $\boldsymbol{v}_r \in V_R$, and part of the reflected ray set $V_C \subset V_R$ can be captured by the camera to participate in the corneal imaging process. When the subject is indoors and facing the high-light L, the set of incident corneal rays $V_H \subset V_I$ from the L, of which the partially reflected rays $V_{HR} \subset V_C$, are captured by the camera and become the bright pattern of the CI. The rest of the CI comes from the partially reflected rays $V_{SR} \subset V_C$ which are the reflected rays of incident rays V_S in the scene.

There are two projection processes in CIS. Firstly, the incident rays V_I are projected to the corneal surface. Secondly, the reflected rays V_C are projected to the camera imaging plane. The corneal imaging process can be expressed as Equation (1).

$$I = P_c V_I, \tag{1}$$

where P_c is the projection matrix, which can be expressed as Equation (2):

$$P_c = K_c \left[R_c | \boldsymbol{t}_c \right], \tag{2}$$

where

$$\boldsymbol{t}_c = -R_c \tilde{C}_c,$$

where K_c, R_c and $\tilde{C}_c$ are the intrinsic parameter, rotation matrix and the inhomogeneous coordinate of the camera center position, respectively.

As the extrinsic camera parameters, R_c and $\tilde{C}_c$ described the posture and position of the CIS in the world coordinate system O_w. Without loss of generality, the head pose can be considered to represent R_c. In this paper, the subject posture and position we try to estimate can be represented by R_c as well as $\tilde{C}_c$.

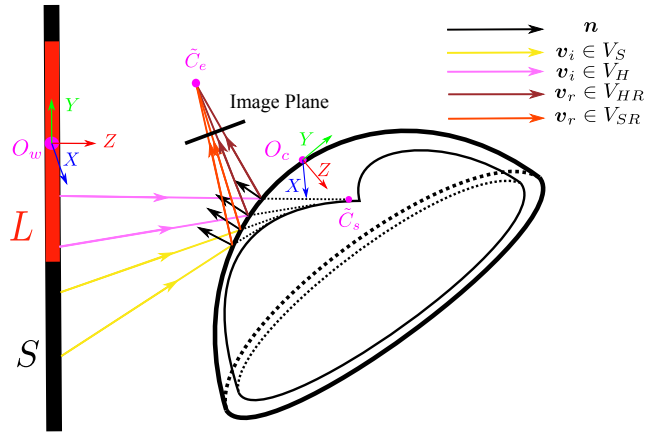


Figure 2. Imaging process of CIS. The incident ray set V_I is reflected by the cornea, and a portion of the reflected ray set V_C is captured by the camera to form a CI.

3.1 Intrinsic Parameter of CIS

The intrinsic parameter K_c of the CIS can be expressed as

$$K_c = P_e P_s R_g,$$

where P_e is the projection matrix of the camera, which takes the form $P_e = K_e\,[R_e|\boldsymbol{t}_e]$, where K_e is the intrinsic parameter of the camera, $\boldsymbol{t}_e = -R_e\tilde{C}_e$, R_e and $\tilde{C}_e$ are the extrinsic parameters of the camera.

The form of K_e is as in Equation (3), without loss of generality, assuming that the camera pixels are square $f_x = f_y = f$, skew parameter $s = 0$, and the camera center is at the center of the imaging plane $c_x = c_y = 0$.

Then K_e in the projective transformation mainly scales the incident light and reduces the dimension.

$$K_e = \begin{bmatrix} f_x & s & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{3}$$

R_e is the rotation matrix of the camera, and $\tilde{C}_e$ is the position of the CIS represented in inhomogeneous coordinate form in the world coordinate O_w.

The Z component of $\tilde{C}_e$ affects the size of the face in the image, which is mainly affected by the subject's shooting habits and arm length, while the X, Y components mainly affect the position of the bright pattern in the CI. The $\tilde{C}_e$ has less influence on the position of the bright pattern in the CI when the face area is often centered in the selfie.

In order to estimate the effect of R_e on the CIS when taking a selfie, we invite 9 subjects sit in front of the screen in turn, complete the selfie with the front lens of the phone and keep the phone position still, and then take the AprilTag marker photo which displayed on the screen with the rear lens, calculate the orientation of the phone at this moment, and repeat five times for each person, record and count the data, as shown in Figure 3. It can be seen that the camera imaging plane is basically parallel to the screen plane, and R_e has little effect on the CIS.

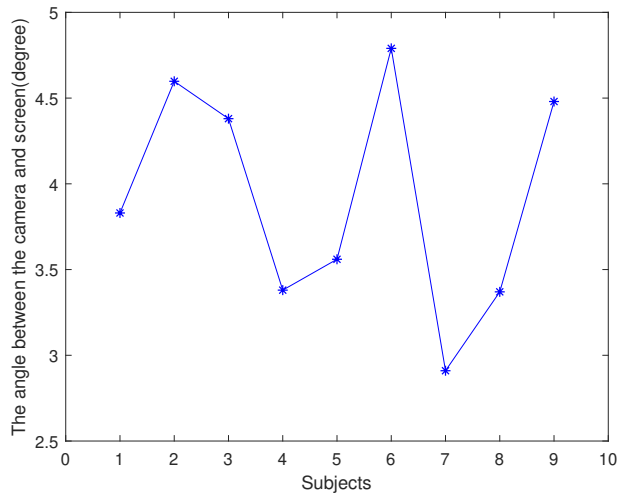


Figure 3. Camera pose distribution when taking selfies of different subjects

$R_g = \{yaw, pitch, roll\}$ is the rotation of the eyeball during corneal reflection imaging, *yaw*, *pitch*, and *roll* are the rotation angles of the eyeball around the Y, X, and Z axes under the coordinate system O_c, respectively. The eyeball rotation affects the pose of the corneal reflective surface P_s.

We refer to the work of Nishino and Nayar [18] to model the corneal geometry structure and analyze the influence of the corneal reflective surface posture P_s during the corneal imaging process, as shown in Figure 4. The corneal geometry can be described as

$$S(t, \theta) = (S_x, S_y, S_z) = (\lambda \cos\theta, \lambda \sin\theta, t)\,,$$

where $t \in [0, 2.18]$, $\theta \in [0, 2\pi]$ and

$$\lambda = \sqrt{-pt^2 + 2Rt}.$$

Based on the anatomical work of Kaufman and Alm [26], the shape of the cornea is found to be essentially the same in different adults. The radius of curvature at the vertex $R = 7.8\,\text{mm}$, the mean eccentricity $e = 0.5$, and $p = 1 - e^2$. The tangent

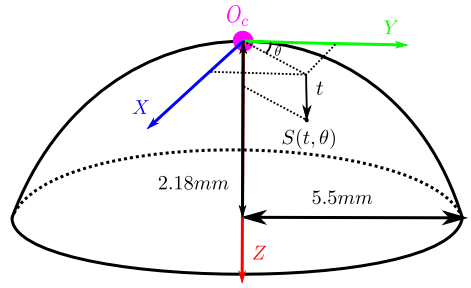


Figure 4. Modeling the geometric structure of the cornea. Corneal height was 2.18 mm, limbus as the outermost edge, which can be considered as a circle, has a radius of 5.5 mm.

$\boldsymbol{\tau}(t,\theta) = \left[\nabla_{S_x}, \nabla_{S_y}, \nabla_{S_z}\right]^T$ to any point $S(t,\theta)$ on X, Y, Z

$$\nabla_{S_x} = \frac{\cos\theta}{2\sqrt{-1.5t+15.6}} - \sin\theta\sqrt{-0.75t^2+15.6t},$$

$$\nabla_{S_y} = \frac{\sin\theta}{2\sqrt{-1.5t+15.6}} + \cos\theta\sqrt{-0.75t^2+15.6t},$$

$$\nabla_{S_z} = 1.$$

For any point $S(t,\theta)$ on the corneal surface, its normal vector $\boldsymbol{n}(t,\theta) = [F_x, F_y, F_z]^T$. The $\boldsymbol{n}(t,\theta)$ and $\boldsymbol{\tau}(t,\theta)$ satisfy

$$\boldsymbol{n}^T(t,\theta)\cdot\boldsymbol{\tau}(t,\theta) = F_x\nabla S_x + F_y\nabla S_y + F_z\nabla S_z = 0.$$

The projection of an incident ray on the corneal surface can be expressed by Equation (4):

$$P_s = K_s\left[R_s|\boldsymbol{t}_s\right], \tag{4}$$

where $\boldsymbol{t}_s = -R_s\tilde{C}_s$, K_s as the intrinsic parameter has similar values for different $\tilde{C}_s$. R_s and $\tilde{C}_s$ are the extrinsic parameters of P_s. The rotation matrix R_s can be obtained from $\boldsymbol{n}(t,\theta)$ by the Rodriguez rotation formula.

$\tilde{C}_s$ is the position of the optical center in O_w for a single corneal reflecting surface expressed in inhomogeneous form. $\tilde{C}_s$ is determined jointly by the incident light $\boldsymbol{v}_i$ and the corneal surface $S(t,\theta)$ involved in the reflection. Different surface involved in the corneal imaging process will have its viewpoint. All viewpoints will form the viewpoint trajectory envelope:

$$V(t,\theta,r) = S(t,\theta) + r\boldsymbol{v}_i(t,\theta),$$

where r is the distance between $S(t,\theta)$ and $\tilde{C}_s$, which can be obtained by

$$detJ(V(t,\theta,r)) = 0,$$

where J is the Jacobi matrix.

The factors act on P_s include three points, namely R_g, R_c and $\tilde{C}_c$. Eye rotation R_g has the same effect on R_s as the subject's head posture change R_c, but when R_g changes, it is accompanied by the corresponding change in the position of the cornea, iris, etc. in the eye. Therefore, we can analyze R_g by the position of cornea in the eye image, and then detach the effect of R_g from P_s to avoid its influence on the estimation of R_c. So far, the CI is unique for a given position R_c, $\tilde{C}_c$, and a given R_g, in the same scene. In fact, shooting CI is often accompanied by the freedom of eyeball rotation, so that for any R_c, $\tilde{C}_c$ corresponds to a set of CIs with different projection bright patterns and iris positions.

3.2 Pose and Position Estimation of CIS

The image quality of CI acquired during self-photography is poor. In addition, the high-light L may be projected incompletely on the corneal reflective surface due to the effect of eyeball rotation, even though the subject is facing L. These two aspects lead to difficulties in solving the exact R_c, $\tilde{C}_c$ by stripping the projection matrix K_c in Equation (1).

We have known that there exists a set of CIs corresponding to the subject in any one of the poses R_c and position $\tilde{C}_c$. These CIs are not the same due to the influence of P_e, R_g. As described in Section 3.1, P_e has less effect on the selfie imaging process, and in our experiments, we have focused on the role of R_g on indoor localization.

We try to test our work in different high-light scenes and take a set of CIs with different R_g for the given subject's pose and position. We will use these CIs with various positions and poses to train the neural network and ensure that the neural network can predict the positions and poses of the subjects in the CIs.

In addition, the size of the high-light L is another factor that affects the positioning performance. Different sizes of L will have different positioning accuracy and effective positioning range. Based on the above hypothesis, we try to design an experiment to verify our idea and also estimate the accuracy performance under different high-light source sizes.

To observe the performance of localization, we find the effective indoor ground range S, and divide S into 64 equal regions, each region is centered at $C_i \in C$. In the data acquisition process, we let the subject take a selfie in R_{ci} pose at three different locations in region i and record the location data $C_{ci} = \{C_{ix}, C_{iy}, H_s\}$ containing the subject's height H_s, the pose data R_{ci} and the eye part of the selfie image. Next, we record the test set data at the center C_i in the same data acquisition manner.

We use eye photos and high-light size of the D_{Train} as input and pose, position information as output to complete the training of VGG16 network. Using the eye photos and high-light size from the D_{Test} as the input of the trained VGG16, we then compare the output of VGG16 network with the ground truth data to

verify the feasibility of localization by CIs and analyze the effective localization range.

The indoor space S includes some indistinguishable regions, and we exclude the above locations from the effective space S in order to speed up the convergence of the training network and reduce the possibility of over-fitting. For locations very close to the high-light L, the bright pattern of the high-light projection completely occupies the CI, and in this case, the shape of the bright pattern does not reflect the position of the subject. When the subject is far away from L, the shape of the bright pattern does not reflect the change of pose and position, both of which are indistinguishable regions.

4 POSITION AND POSE ESTIMATION PERFORMANCE

We looked for 60 volunteers with different heights, ages and genders as subjects. We measured the height of the subject as shown in Figure 5. We conducted experiments in three experimental fields $S = \{S_1, S_2, S_3\}$ with three different high-light areas $L = \{L_1, L_2, L_3\}$, respectively, to observe the effect of different high-light areas on the localization. The sizes of L_1, L_2, L_3 are $280 \times 160\,\mathrm{cm}^2$, $130 \times 120\,\mathrm{cm}^2$ and $55 \times 33\,\mathrm{cm}^2$. Each experimental field was divided into 64 regions with 90 cm, 50 cm and 25 cm sides, as described in Section 3.2.

For subject s, the head pose is $R_{cs} = \{yaw, pitch, 0°\}$, where yaw and $pitch$ are the rotation angles of s around the Y and X axes in world coordinates, respectively. The yaw angle range is

$$yaw \in \{-60°, -45°, -30°, -15°, 0°, 15°, 30°, 45°, 60°\}$$

and pitch angle range is

$$pitch \in \{-30°, 0°, 30°\}.$$

At different R_c, $\tilde{C}_c$, the subject s holds a camera and takes a selfie video with different eyeball rotations for at least 15 seconds at 60 fps, 4 k resolution to collect data from the training and test dataset. To ensure that the eye image was less affected by the subject's motion and pose, we asked the subjects to maintain as stable a pose as possible, and also asked them to rotate their eyes at will when they could see the high-light plane to ensure the presence of a more complete light spot in the CI. We then segmented the video into images and split the eye images by frame.

Nevertheless, there were some unusable images, especially when subjects stand in the edge regions of S and were asked to collect experimental data in extreme poses (e.g., $yaw = 60°$, $pitch = 30°$). As in Figure 6, when the eye image has severe motion blur, or when the reflected spot is heavily obscured, such an eye image is invalid and needs to be removed from the dataset. After the experimental data collection and collation process, we get 34 209 294 valid images from 41 472 000 eye images in three experimental spaces, accounting for 82.48 %. The efficient distribution of data

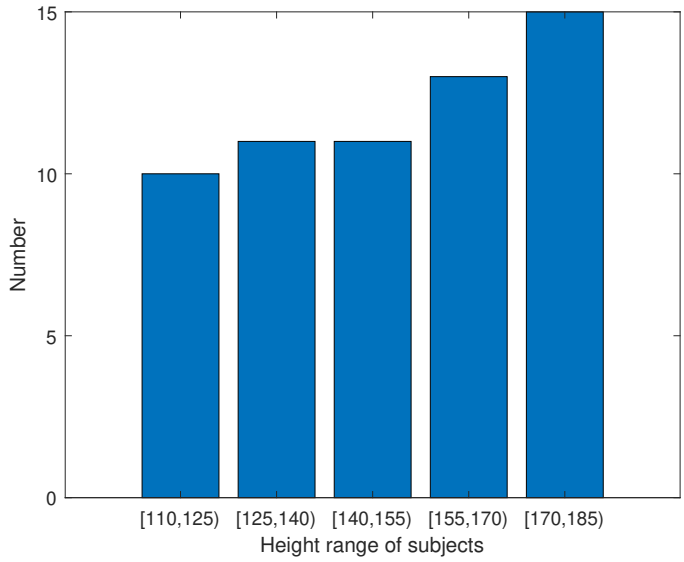
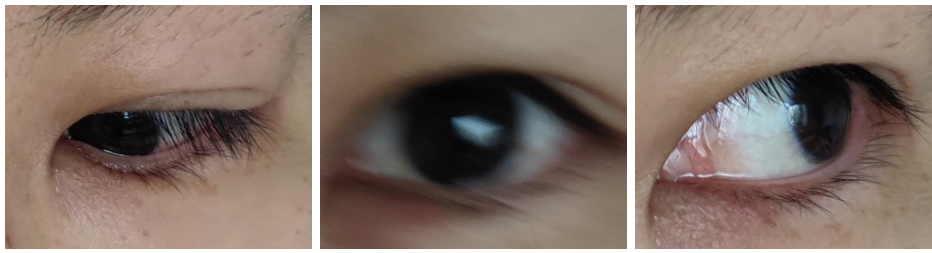


Figure 5. Height distribution of the subjects

in each region is shown in Figure 7. We can see that the closer to the center of S, the more efficient the data are, and the least efficient the data are located at the four corners.



a) Light spot is heavily covered b) Severe motion blur c) Limited spot reflected

Figure 6. Types of video frames that need to be removed from the dataset

After collecting the dataset, we trained and tested the VGG16 network according to Section 3.2.

During the experiment, we find that the key factor affecting the localization accuracy is the size of the spot. Small spots require a wide range of subject's movement and position changes to produce significant deformation. The vertical distance from the measured object to the high-light plane and the orientation angle of the

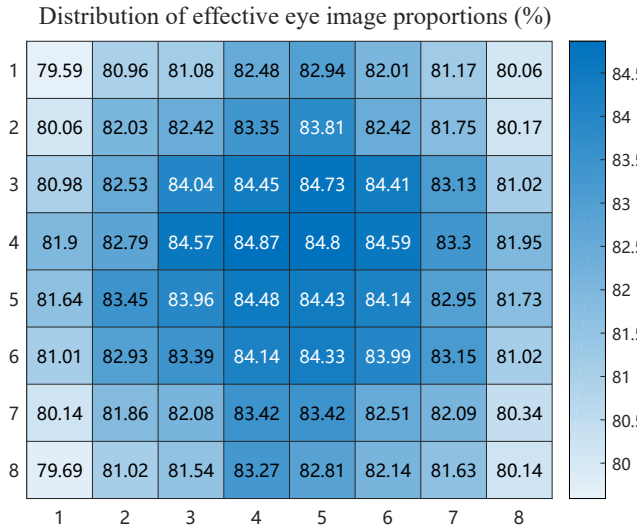


Figure 7. Distribution of available data in each region of the experimental space S as a percentage of the collected data

measured object relative to the high-light plane have a significant effect on the spot size. Similar to the performance measurements of Olson Edwin for AprilTag [12], we measured and demonstrated the experimental performance in terms of the vertical distance from the subject to the high-light plane and the angle between the subject direction and the normal vector of the high-light plane, respectively.

We take the average of the prediction results of valid test data collected from region $R(i,j)$ under pose p as the prediction result of $Pred(i,j,p)$

$$Pred(i,j,p) = \{Pred(i,j,p)_L, Pred(i,j,p)_P\},$$

where

$$Pred(i,j,p) = \frac{\sum_{n=0}^{60} Est(i,j,p)_n}{60},$$

$$Est(i,j,p)_n = \frac{\sum_{k=0}^{m} e(i,j,p)_k}{m},$$

where m is the number of valid video frames of the n^{th} video of test data acquired at $R(i,j)$ and $e(i,j) = \{e(i,j,p)_L, e(i,j,p)_P\}$ is the prediction result of video frames.

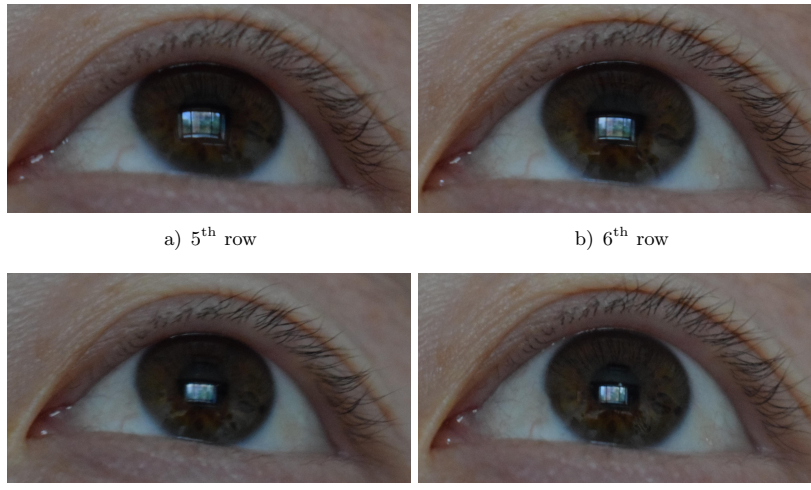
The localization offset range of region $R(i,j)$ under pose p can be decoupled into position estimation offset ranges and pose estimation offset ranges. We keep the pose of the subject equal to $0°$ when measuring the position estimation performance, while we analyze the pose prediction data collected from a fixed region when measuring the pose estimation performance.

We count the average range of the position localization offset ranges for all regions with distance i to the high-light plane as $LOffset(i)$.

$$LOffset(i) = \frac{\sum_0^8 |GT(i,j,0)_L - Pred(i,j,0)_L|}{8},$$

where $GT(i,j,p)_L$ is the location ground truth under the pose p in the $R(i,j)$. To illustrate the ability of our method to cope with different high-lighting environments, we have calculated the position localization offset ranges in L_1, L_2 and L_3 highlighting environment, respectively.

Figure 9 shows the relationship between i and $LOffset(i)$. It can be seen that the value of $LOffset(i)$ increases with i in the three different highlighting areas. As show in Figure 8, we found that the area of the corneal reflection spot decreased rapidly with increasing distance between the subject and the high light plane. When the subject was in row 7, the area of the spot was too small and insensitive to changes in position and posture. As a result, the prediction offset ranges located here becomes larger, the standard deviation of the predicted data increases, and the dispersion of the data becomes more pronounced.



a) 5th row b) 6th row

c) 7th row d) 8th row

Figure 8. The eye images are shot while the subject is standing in rows 5, 6, 7 and 8. By comparison, it can be found that the spot area in the eye image in rows 5 and 6 has a clear shape change, compared to the spot in rows 7 and 8, where the shape change is not obvious enough.

As shown in Figure 7, the proportion of valid data is highest at $R(4,4)$, which means that the performance of pose estimation is more representative than other regions. We convert the pose from the Euler angle to the orientation vector $\boldsymbol{v}$,

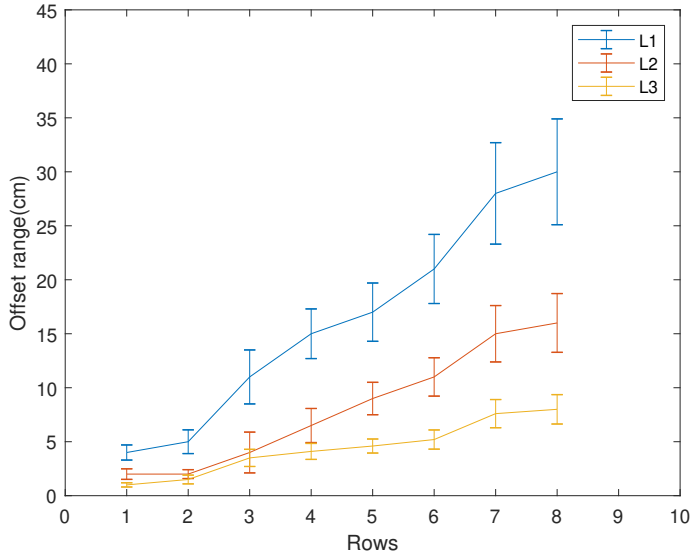


Figure 9. The average location positioning offset ranges increases when the subject is far away from the high-light plane

then we can get the angle between the normal vector of the high-light plane and $\boldsymbol{v}$ is

$$\{0°, 15°, 30°, 33.22°, 41.4°, 45°, 52.2°, 60°, 64.3°\}.$$

We count the pose estimation offset ranges $POffset(p)$

$$POffset(p) = |GT(4,4,p)_P - Pred(4,4,p)_P|,$$

where $GT(i,j,p)_P$ is the pose ground truth in $R(i,j)$. The performance is shown in Figure 10.

We still examined the performance in L_1, L_2 and L_3 high-light scenes, respectively. Similar to the offset ranges distribution of the position estimation, the offset range increases with the orientation angle. However, the incremental gradient is much smaller than the gradient of the position estimation. After examining the original data and the prediction results, we found that at larger orientations, some subjects' eye images were not complete enough, which could easily lead to significant offset ranges in predictions.

5 DISCUSSION

The privacy disclosure risk existing in the social network sites has become a hot topic of concern in recent years. In this paper, we analyze the corneal imaging

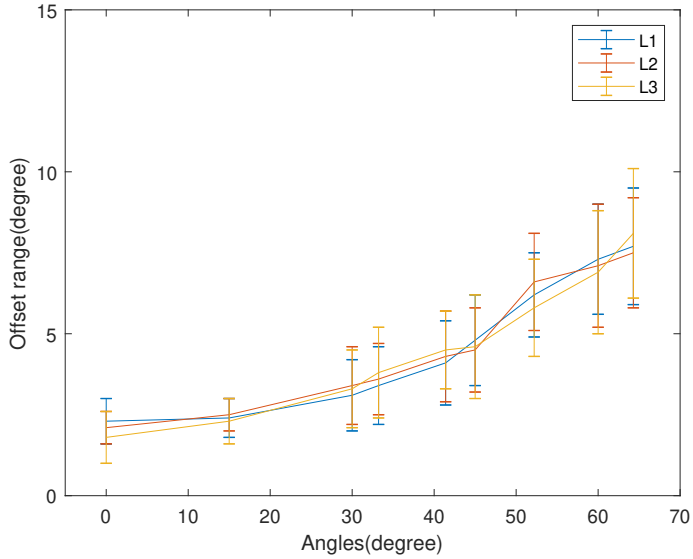


Figure 10. The estimated offset ranges of the postural in the region $R(4,4)$ indicated that the offset range grows slowly as the pose of the subjects increased

process in a selfie scenario, and design a validation experiment to the subjects with different heights in different indoor scenes. Experiments show that selfies and videos taken in high-light scenes on social media do carry some risk of privacy disclosure for location, posture, height, and lifestyle habits. Our work contributes to the privacy protection of v-loggers and also provides theoretical support for privacy-sensitive groups.

Every coin has two sides. Although the CI can cause the privacy disclosure trouble to selfie taker, the CIS based localization method can be applied in other fields, like disaster rescue, crime tracking, human-machine interaction and computer graphics modeling. In addition, thanks to the weak quality of corneal reflection imaging, a balance of privacy protection and convenience can be achieved when CIS based positioning methods are used in daily life.

Our research is innovative and enriches the research and application scenarios of the CIS. It has been experimentally demonstrated that corneal reflection can reveal the approximate pose and position of the photographer in the room. With the development of social media, there are a large number of selfie videos and photos on Youtube, Twitter and TikTok, including a large number of photos and videos of faces facing windows, screens, etc. These images and videos contain corneal reflection areas that contain information about the subject's pose, which can be used to improve human-machine interaction and expand the application scenarios of human-machine interaction under the constraints of the law. The corneal images

can be shot easily with non-aggressive to others, and as wearable hardware devices evolve, CIS-based positioning efforts can be senselessly integrated into existing smart devices, such as smart glasses, as well as AR and VR devices while being non-intrusive to personal space.

From the above experimental results, we can see that the performance of the CI in the indoor position and posture estimation is related to the area of the high-light and the position of the distance from the high-light. We depend on the solid angle to evaluate the influence of these two aspects.

The solid angle Ω_c presents the reflectable range of the cornea. Let the angle between the normal vector $\boldsymbol{n}_{(0,0)}$ of the corneal curvature vertex O_c and the direction of the optical axis of the camera be σ, Ω_c decreases with increasing σ. For any reflected infinitesimal space δA, the solid angle is

$$\Omega_A = \frac{\delta A \cos^3 \phi(i,j)}{z^2},$$

where $\phi(i,j)$ is the angle between $\boldsymbol{n}_{(0,0)}$ and the normal $\boldsymbol{n}_A$ of δA. Then the solid angle of the corneal reflectable scene C and the high-light L are

$$\Omega_C = \sum \frac{\delta C \cos^3 \phi(m,n)}{z^2},$$

$$\Omega_L = \sum \frac{\delta L \cos^3 \phi(i,j)}{z^2},$$

where $i \in [0,m]$, $j \in [0,n]$, then the scale of the bright pattern in the CI can be expressed as

$$p = \Omega_L / \Omega_C. \tag{5}$$

It is known that Ω_C is constant when σ is constant, and for the sake of calculation, it is assumed that $\delta L = 1$. In general, since the $\Omega_C > 2\pi$ of the cornea, the FOV of the corneal reflection is larger than that of the hemispherical reflective surface. However, due to the influence of the corneal geometry, the resolution at the edge of the cornea decreases sharply compared to that above the pupil. The projected image of the scene at the edge of the cornea piles up and compresses [18]. Therefore, in this paper, when measuring the proportion of the bright pattern in the CI by Equation (5), the value of Ω_C should be less than 2π, and the ratio of the area of the light pattern formed by the projection of the high-light on the cornea to the area of the whole cornea reflectable range is shown in Figure 11.

L_1, L_2 and L_3 are high-light scenes with different sizes that have been used in the experimental section. The size of the bright pattern in the CI changes slowly with increasing Z-axis compared to the position near the high-lights. Also, regions with the same Z value have different localization accuracy. When the distance d between the subject s and O_w increases, the rate of change of Ω_L decreases and p is too similar to be distinguished from nearby regions.

Our work can estimate the posture and position of the subject without additional devices. However, our method can only estimate the posture of head instead of

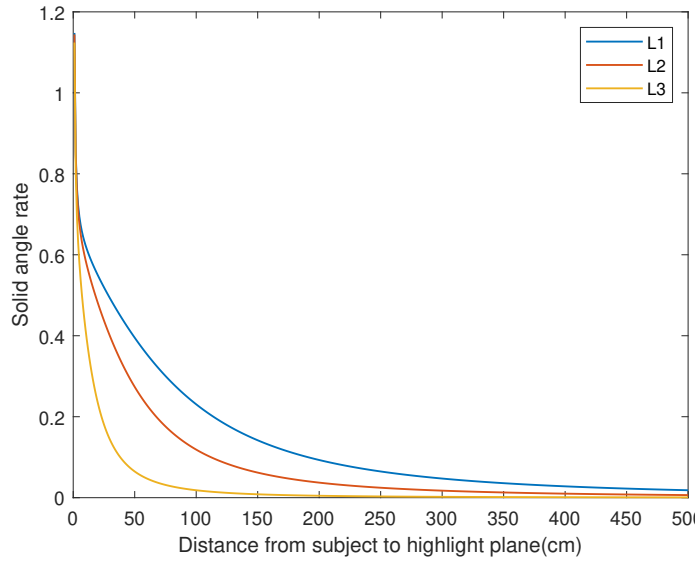


Figure 11. The effect of the subject position to the size of the bright pattern in CI

the whole body. Moreover, from the experimental results, our method cannot be applied to fine body motion tracking applications. Besides, the accuracy of pose and position estimation can hardly be improved without the assistance of the additional equipment.

6 CONCLUSION

In this paper, we demonstrate theoretically and experimentally that selfie images contain private information such as the position and pose of the selfie taker. Our work can help users of social networks to protect their privacy further. In addition, indoor localization by CI is promising in the field of low-accuracy indoor localization, which is less intrusive and less equipment friendly than existing methods. In our future work, we will focus on how to improve the accuracy of CIS for indoor positioning and try to further develop and expand its application.

REFERENCES

[1] Lowe, D. G.: Distinctive Image Features from Scale-Invariant Keypoints. International Journal of Computer Vision, Vol. 60, 2004, No. 2, pp. 91–110, doi: 10.1023/B:VISI.0000029664.99615.94.

[2] Harris, C.—Stephens, M.: A Combined Corner and Edge Detector. Proceedings of the Fourth Alvey Vision Conference (AVC 1988), Alvey Vision Club, 1988, pp. 147–151, doi: 10.5244/C.2.23.

[3] Hartley, R.—Zisserman, A.: Multiple View Geometry in Computer Vision. Cambridge University Press, 2003, doi: 10.1017/CBO9780511811685.

[4] Feiner, S.—MacIntyre, B.—Seligmann, D.: Knowledge-Based Augmented Reality. Communications of the ACM, Vol. 36, 1993, No. 7, pp. 53–62, doi: 10.1145/159544.159587.

[5] Chen, J.—Granier, X.—Lin, N.—Peng, Q.: On-Line Visualization of Underground Structures Using Context Features. Proceedings of the 17th ACM Symposium on Virtual Reality Software and Technology (VRST '10), 2010, pp. 167–170, doi: 10.1145/1889863.1889898.

[6] Mur-Artal, R.—Montiel, J. M. M.—Tardos, J. D.: ORB-SLAM: A Versatile and Accurate Monocular SLAM System. IEEE Transactions on Robotics, Vol. 31, 2015, No. 5, pp. 1147–1163, doi: 10.1109/TRO.2015.2463671.

[7] Wang, Z.—Zhang, J.—Chen, S.—Yuan, C.—Zhang, J.—Zhang, J.: Robust High Accuracy Visual-Inertial-Laser SLAM System. 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2019, pp. 6636–6641, doi: 10.1109/IROS40897.2019.8967702.

[8] Zhang, J.—Gui, M.—Wang, Q.—Liu, R.—Xu, J.—Chen, S.: Hierarchical Topic Model Based Object Association for Semantic SLAM. IEEE Transactions on Visualization and Computer Graphics, Vol. 25, 2019, No. 11, pp. 3052–3062, doi: 10.1109/TVCG.2019.2932216.

[9] Snavely, N.—Seitz, S. M.—Szeliski, R.: Photo Tourism: Exploring Photo Collections in 3D. ACM Transactions on Graphics, Vol. 25, 2006, No. 3, pp. 835–846, doi: 10.1145/1141911.1141964.

[10] Wu, C.: Towards Linear-Time Incremental Structure from Motion. 2013 International Conference on 3D Vision – 3DV 2013, IEEE, 2013, pp. 127–134, doi: 10.1109/3DV.2013.25.

[11] Fischler, M. A.—Bolles, R. C.: Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. Communications of the ACM, Vol. 24, 1981, No. 6, pp. 381–395, doi: 10.1145/358669.358692.

[12] Olson, E.: AprilTag: A Robust and Flexible Visual Fiducial System. 2011 IEEE International Conference on Robotics and Automation, 2011, pp. 3400–3407, doi: 10.1109/ICRA.2011.5979561.

[13] Kato, H.—Billinghurst, M.: Marker Tracking and HMD Calibration for a Video-Based Augmented Reality Conferencing System. Proceedings 2nd IEEE and ACM International Workshop on Augmented Reality (IWAR '99), 1999, pp. 85–94, doi: 10.1109/IWAR.1999.803809.

[14] Tsumura, N.—Dang, M. N.—Makino, T.—Miyake, Y.: Estimating the Directions to Light Sources Using Images of Eye for Reconstructing 3D Human Face. Color and Imaging Conference, Vol. 11, 2003, No. 1, pp. 77–81, doi: 10.2352/CIC.2003.11.1.art00014.

[15] Nishino, K.—Nayar, S. K.: Eyes for Relighting. ACM Transactions on Graphics (TOG), Vol. 23, 2004, No. 3, pp. 704–711, doi: 10.1145/1015706.1015783.

[16] Wang, H.—Lin, S.—Liu, X.—Kang, S. B.: Separating Reflections in Human Iris Images for Illumination Estimation. Tenth IEEE International Conference on Computer Vision (ICCV '05) Volume 1, Vol. 2, 2005, pp. 1691–1698, doi: 10.1109/ICCV.2005.215.

[17] Nitschke, C.—Nakazawa, A.: Super-Resolution from Corneal Images. Proceedings of the British Machine Vision Conference (BMVC 2012), BMVA Press, 2012, `https://bmva-archive.org.uk/bmvc/2012/BMVC/paper022/index.html`.

[18] Nishino, K.—Nayar, S. K.: Corneal Imaging System: Environment from Eyes. International Journal of Computer Vision, Vol. 70, 2006, No. 1, pp. 23–40, doi: 10.1007/s11263-006-6274-9.

[19] Swaminathan, R.—Grossberg, M. D.—Nayar, S. K.: Caustics of Catadioptric Cameras. Proceedings Eighth IEEE International Conference on Computer Vision (ICCV 2001), Vol. 2, 2001, pp. 2–9, doi: 10.1109/ICCV.2001.937581.

[20] Nitschke, C.—Nakazawa, A.—Takemura, H.: Eye Reflection Analysis and Application to Display-Camera Calibration. 2009 16$^{\text{th}}$ IEEE International Conference on Image Processing (ICIP), 2009, pp. 3449–3452, doi: 10.1109/ICIP.2009.5413852.

[21] Suda, S.—Yamagishi, K.—Takemura, K.: User Calibration-Free Method Using Corneal Surface Image for Eye Tracking. Proceedings of the 12$^{\text{th}}$ International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2017) – Volume 6: VISAPP, SciTePress, 2017, pp. 67–73, doi: 10.5220/0006100100670073.

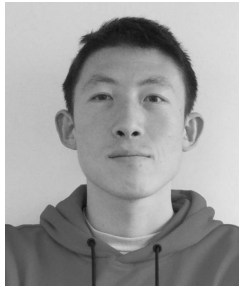
[22] Nakazawa, A.—Kato, H.—Nitschke, C.—Nishida, T.: Eye Gaze Tracking Using Corneal Imaging and Active Illumination Devices. Advanced Robotics, Vol. 31, 2017, No. 8, pp. 413–427, doi: 10.1080/01691864.2016.1277552.

[23] Lander, C.—Löchtefeld, M.—Krüger, A.: hEYEbrid: A Hybrid Approach for Mobile Calibration-Free Gaze Estimation. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, Vol. 1, 2018, No. 4, Art. No. 149, doi: 10.1145/3161166.

[24] Ohshima, Y.—Maeda, K.—Edamoto, Y.—Nakazawa, A.: Visual Place Recognition from Eye Reflection. IEEE Access, Vol. 9, 2021, pp. 57364–57371, doi: 10.1109/ACCESS.2021.3071406.

[25] Du, M.—Chen, K.—Zhang, J.—Liu, H.: A Swift Gaze Estimate Method Based on the Corneal Image System. 2022 IEEE 25$^{\text{th}}$ International Conference on Computer Supported Cooperative Work in Design (CSCWD), 2022, pp. 734–739, doi: 10.1109/CSCWD54268.2022.9776291.

[26] Kaufman, P. L.—Alm, A.: Adler's Physiology of the Eye: Clinical Applications. Mosby, 2003.



Mengqi Du received his Master's degree in computer science and technology from the Zhejiang University of Technology in 2018. He is currently pursuing his Ph.D. degree in the Department of Computer science, Zhejiang University of Technology. His research interests include computer graphics, computer vision, human attention analysis and human-computer interaction.

Yue Zhang received his Master's degree in computer science and technology from the Zhejiang University of Technology in 2017. He is currently pursuing his Ph.D. degree in the Department of Computer science, Zhejiang University of Technology. His research interests include intelligent system, biological signal processing, human-computer interaction and adaptive learning methods.

Jianhua Zhang received his Ph.D. from the University of Hamburg, Hamburg, Germany in 2012. He is currently Professor with the School of Computer Science and Engineering, Tianjin University of Technology, Tianjin, China. His current research interests include SLAM, 3D vision, reinforcement learning, and machine vision.

Honghai Liu received his Ph.D. degree in robotics from the Kings College London, London, U.K., in 2003. He is the Chair Professor of Human Machine Systems, University of Portsmouth, Portsmouth, U.K. His research interests include biomechatronics, pattern recognition, intelligent video analytics, intelligent robotics, and their practical applications with an emphasis on approaches that could make contribution to the intelligent connection of perception to action using contextual information.