15th IAPR International Conference on Machine Vision Applications (MVA)
Nagoya University, Nagoya, Japan, May 8-12, 2017.

Point of Gaze Estimation Using Corneal Surface Reflection and Omnidirectional Camera Image

Taishi Ogawa
Kyoto University
ogawa@ii.ist.i.kyoto-u.ac.jp

Atsushi Nakazawa
Kyoto University
nakazawa.atsushi@i.kyoto-u.ac.jp

Toyoaki Nishida
Kyoto University
nishida@i.kyoto-u.ac.jp

Abstract

We present a human point of gaze estimation system using corneal surface reflection and omnidirectional image taken by a fish eye. Only capturing an eye image, our system enables to find where a user is looking in 360° surrounding scene image. We first generates multiple perspective scene images from an equirectangular image and perform registration between corneal reflection and perspective images. We then compute the point of gaze using a 3D eye model and project the point to an omnidirectional image. We evaluated the robustness of registration and accuracy of PoG estimations using two indoor and five outdoor scenes, and found that gaze mapping error was 5.526[deg] on average. This result shows the potential to the marketing and outdoor training system.

1 Introduction

Human gaze information is popularly used in a number of research fields such as marketing, consumer research and child social development study. Eye gaze tracking (EGT) systems, such as Tobii[1], are the currently employed. These systems performs high accuracy in precisely setuped conditions, however, have several drawbacks.

First, they require system calibrations at everytime. If the configuration changes such as when the headmount devices drift occurs, we needs to perform tedious calibration again. Moreover, EGT systems only obtain the PoG in the range of a frontal scene and therefore have difficulty in obtaining out of a frontal scene image coordinate.

In addition to these technical problems, the systems also have a social problem about privacy concerns due to the frontal scene camera.

In this work, we present a novel EGT system which uses a corneal imaging technique[1][2][3] and an omnidirectional camera image. Figure 2 shows an overview of our system.

First we create perspective images from a omnidirectional camera image and then conduct registration between an eye image and each perspective image. At the same time, we estimate a 3D eye pose[2] and compute the gaze reflection point(GRP) in the eye image. Finally, we project GRP to the omnidirectional image using the result of registration.

The advantages of our system are as follows:

[1]http://www.tobiipro.com/

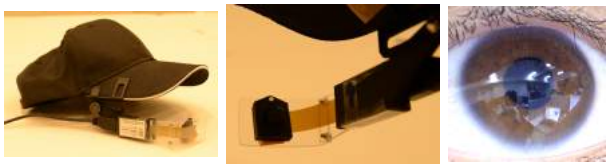

Figure 1. A corneal imaging camera and an eye image.

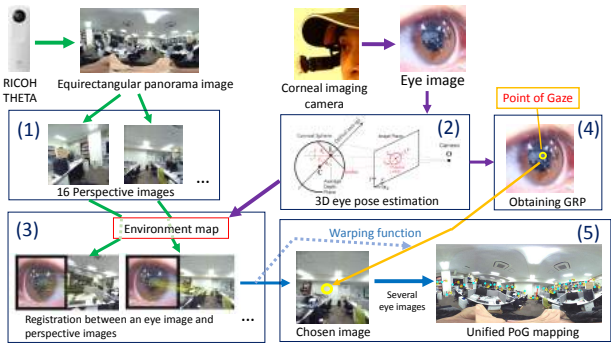

Figure 2. Overview of the system.

(1) Our system can obtain the PoG only from an eye image, therefore, does not require calibration step and headmount drift.

(2) Mapping the PoG to an omnidirectional camera image enables us to obtain the PoG in our 360° surrounding scene.

(3) Scene images can be prepared off-line and a corneal imaging camera does not take scene images. We therefore do not suffer from privacy issues.

2 Related Work

Corneal Imaging Techniques. Using corneal reflection for point of gaze estimation has been proposed in existing work[1][2][3]. Using the method proposed by [3], the relation between a cornela reflection and a perspective scene image can be obtained automatically. Therefore, it does not need to rely on calibration. Using this advantage, corneal reflection image can be directly used for scene observations what people are looking at[4] and real-time human view estimation[2].

Image Registration. There have been a lot of image registration algorithms. In recent years, feature-based registration is popular because of development of

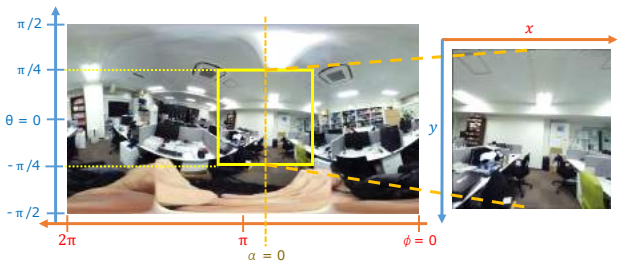

Figure 3. omnidirectional camera image and perspective images.

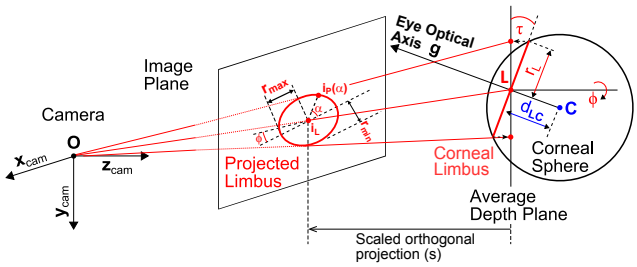

Figure 4. 3D eye pose estimation from the projected limbus[2]

local feature descriptors such as SIFT[5] and SURF[6]. RANSAC[7] and its extensions are popular techniques to robustly estimate parameters between noisy images. First, the hypothesis of a transformation is obtained by using randomly sampled pairs of points. Then counting how many other pairs are correctly warped (inliers), the hypothesis which has most inliers is chosen.

To obtain PoG in sphere scene images, we need to solve the *registration* problem between corneal surface images and sphere scene images.

3 Point of Gaze Estimation System

Figure 2 shows the overview of our system. The system consists of five components: **1)** Create perspective images from a omnidirectional camera image. **2)** 3D eye pose estimation from an eye image. **3)** Registration between an eye image and perspective images. **4)** GRP estimation in an eye image. **5)** Mapping the GRP to the omnidirectional image and obtain the PoG.

We will describe the details in followings.

1) Create perspective images from a omnidirectional camera image. We use RICOH THETA (Figure 2) which produces a equirectangular panorama image $\mathbf{I}_0$ where x-axis and y-axis corresponds to ϕ and θ respectively (Figure 3). Since people do not looks at the sky and the ground, we only use $-\pi/4 \leq \theta \leq \pi/4$ in the image. Regarding horizontal direction, we assume 16 virtual perspective cameras whose viewing angles are $\pi/2$ and resolutions are 600×600. We obtain the perspective images $\mathbf{I}_1, \mathbf{I}_2, \ldots, \mathbf{I}_{16}$ from $\mathbf{I}_0$ by using following equation,

$$\mathbf{I}_n(x,y) = (\ 600(1-\tan\alpha_n)/2,\ 600(1-\tan\theta)/2\ ), \quad (1)$$
$$(n = 1,2,\ldots,16,\ \ -\pi/4 \leq \alpha, \theta \leq \pi/4),$$
$$\text{where} \quad \alpha_n = \phi - (n-1)\pi/16.$$

Here, we assume the resolutions of the $\mathbf{I}_0$ are X, Y, and $\phi = 2\pi/(X-x)$, $\theta = \pi/(Y-y) + \pi/2$ in the pixel of $\mathbf{I}_0(x,y)$.

2) 3D eye pose estimation from an eye image. The 3D eye pose is computed by using the methods[1][2]. Then the eye optical axis $\mathbf{g}$ can be obtained from the elliptical contour in the projected limbus as $\mathbf{g} = [-\sin\tau\sin\theta \ \ \sin\tau\cos\theta \ \ -\cos\tau]^{\mathbf{T}}$, where angle $\tau = \pm\arccos(r_{\min}/r_{\max})$ corresponds to the tilt of the limbus plane with respect to the image, and angle θ is already known as it is the rotation of the limbus ellipse in the image plane (Figure 4). The average radius of the limbus r_{L} is approximately 5.6 mm.

Setting the average depth plane of the weak perspective projection at L, the scale parameter of the projection is given by $s = r_{\max}/r_{\mathrm{L}}$, and then the projected limbus center $\mathbf{i}_{\mathrm{C}}$ can be obtained as

$$\mathbf{i}_{\mathrm{C}} = \mathbf{i}_{\mathrm{L}} + s \cdot d_{\mathrm{LC}} \begin{bmatrix} \sin\tau\sin\theta \\ -\sin\tau\cos\theta \end{bmatrix}. \quad (2)$$

3) Registration between an eye image and perspective images. Next, we perform an image registration algorithm between an eye image and each of 16 perspective images and find the correct one. Here, we use RANSAC-based image registration algorithm.

We assume that the eye reflection and the virtual perspective camera share the 3D environment map showed as Figure 5[3]. First we obtain the function $\mathbf{L}(\mathbf{p})$, which transforms an eye image point $\mathbf{p}$ to the 3D environment map. The radius of the corneal sphere is r_{C}, thus the normal vector $\mathbf{n}_{\mathrm{p}}$ at the eye image point $\mathbf{i}_{\mathrm{p}}$ is obtained as

$$\mathbf{n}_{\mathrm{p}} = [\ n_{\mathrm{p}}^x \quad n_{\mathrm{p}}^y \quad n_{\mathrm{p}}^z\ ]^{\mathbf{T}}, \quad (3)$$
$$\mathbf{i}_{\mathrm{p}} = [\ p_x \quad p_y\ ]^{\mathbf{T}}, \quad \mathbf{i}_{\mathrm{c}} = [\ c_x \quad c_y\ ]^{\mathbf{T}},$$
$$n_{\mathrm{p}}^x = \frac{p_x - c_x}{s \cdot r_{\mathrm{C}}}, \quad n_{\mathrm{p}}^y = \frac{p_y - c_y}{s \cdot r_{\mathrm{C}}},$$
$$n_{\mathrm{p}}^z = \sqrt{1 - (n_{\mathrm{p}}^x)^2 - (n_{\mathrm{p}}^y)^2}.$$

Using the normal vector, the function $\mathbf{L}(\mathbf{p})$ is obtained as

$$\mathbf{L}(\mathbf{p}) = [\ 0 \quad 0 \quad 1\ ]^{\mathbf{T}} + 2(-[\ 0 \quad 0 \quad 1\ ] \cdot \mathbf{n}_{\mathrm{p}})\ \mathbf{n}_{\mathrm{p}}, \quad (4)$$

Next we obtain the function $\mathbf{A}_{\mathrm{s}}(\mathbf{q})$, which transforms a perspective image point $\mathbf{q}$ to the 3D environment map. Using a scene camera internal matrix K_{s}, $\mathbf{A}_{\mathrm{s}}(\mathbf{q})$ is obtained as

$$\mathbf{A}_s(\mathbf{q}) = \frac{\mathsf{K}_{\mathrm{s}}^{-1}[\ \mathbf{q}^{\mathrm{T}} \quad 1\ ]^{\mathrm{T}}}{||\mathsf{K}_{\mathrm{s}}^{-1}[\ \mathbf{q}^{\mathrm{T}} \quad 1\ ]^{\mathrm{T}}||}. \quad (5)$$

Thus, the registration problem can be formulated as obtaining the matrix R in the following equation: $\mathbf{L}(\mathbf{p}) = \mathsf{R}\mathbf{A}_{\mathrm{s}}(\mathbf{q})$.

R can be solved by using a single point registration algorithm shown in [josa]. In the end, the mapping function $\mathbf{W}$ which transforms a point $\mathbf{p}$ in an eye image to a point$\mathbf{q}$ in the perspective scene image $\mathbf{I}_t$ is computed using R as follows:

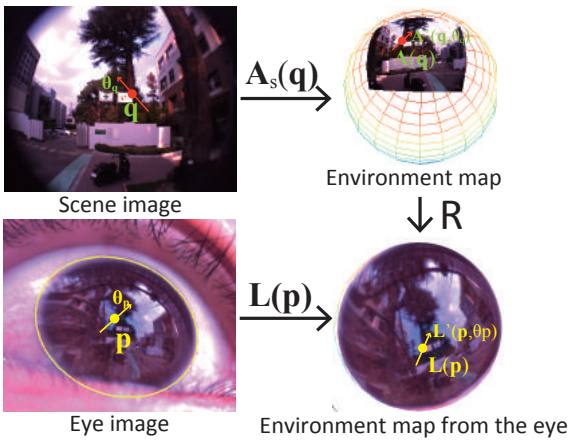

Figure 5. Relation of eye reflection and scene images and their environment maps (EM).

$$\mathbf{W}(\mathbf{p}) \equiv \frac{\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \mathsf{K}_s \mathsf{R}^{-1} \mathbf{L}(\mathbf{p})}{\begin{bmatrix} 0 & 0 & 1 \end{bmatrix} \mathsf{K}_s \mathsf{R}^{-1} \mathbf{L}(\mathbf{p})} \quad (t = 1, \ldots, 16). \tag{6}$$

For each perspective image $\mathbf{I}_1, \ldots, \mathbf{I}_{16}$, we obtain $\mathbf{W}_t(\mathbf{p})(t = 1, \ldots, 16)$ that maximize the number of inlier pairs of points between an eye and perspective image. Assuming the number of inlier pairs of point for $\mathbf{W}_t$ as E_t, we choose the t^* as follows:

$$t^* = \arg\max_{t=1,\ldots,16} E_t. \tag{7}$$

4) GRP estimation in an eye image. Using the obtained 3D corneal pose, we compute the GRP, which can be used to obtain the point in an eye image where the light from the PoG is reflected at the corneal surface[2] through following steps. First, we obtain the visual axis $\mathbf{g}'$, which is slightly different from the eye optical axis $\mathbf{g}$ (Figure 6(a)) ,$\mathbf{g}' = \mathbf{R}_{\text{offset}}\mathbf{g}$. $\mathbf{R}_{\text{offset}}$ can be described by the rotation about x-axis, y-axis, and z-axis. If we set the optimized value that minimizes the PoG error for each individual, we can estimate the PoG more accurately. However, we will use a constant value in section 4 since we show the effectiveness of the non-calibrated PoG estimation.

Figure 6(b) shows the light reflection on the corneal surface. The point $\mathbf{i_T}$ is the GRP $\mathbf{T}$ in the image plane. Using a weak perspective projection, the reflection of the light ray at $\mathbf{T}$ is formulated as

$$\mathbf{C} \cdot \mathbf{n_T} = \begin{bmatrix} \cos\tau' & \sin\tau' \end{bmatrix} \cdot \mathbf{n_T}, \quad \mathbf{C} = \begin{bmatrix} 1 & 0 \end{bmatrix},$$
$$\mathbf{n_T} = \begin{bmatrix} \cos\theta & \sin\theta \end{bmatrix}, \quad \tau' = \arccos(g'_z),$$
$$\mathbf{g} = \begin{bmatrix} g_x & g_y & g_z \end{bmatrix}^{\mathrm{T}}, \quad \mathbf{g}' = \begin{bmatrix} g'_x & g'_y & g'_z \end{bmatrix}^{\mathrm{T}}.$$

We then find the corneal angle θ by using the known eye gaze angle τ' as

$$\theta = \arctan((1 - \cos\tau')/\sin\tau'). \tag{8}$$

Using obtained corneal angle θ, GRP in the image plane $\mathbf{i_T}$ is obtained as

$$\mathbf{i_T} = \mathbf{i_L} + s\left(-d_{\text{LC}}\begin{bmatrix} g_x & g_y \end{bmatrix}^{\mathrm{T}} + r_{\text{C}} \sin\theta \begin{bmatrix} g'_x & g'_y \end{bmatrix}^{\mathrm{T}}\right). \tag{9}$$

where s is the scale factor of the weak perspective projection and $\mathbf{i_L}$ is the center of the limbus ellipse.

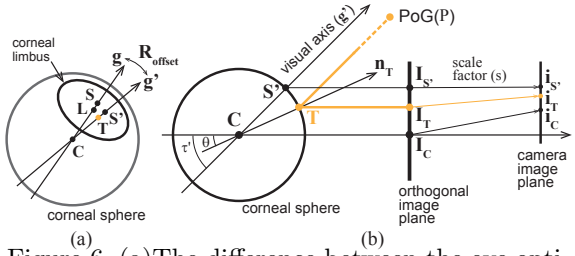

Figure 6. (a)The difference between the eye optical axis($\mathbf{g}$) and the visual axis($\mathbf{g}'$). Gaze reflection point $\mathbf{T}$ lies in a plane that passes the $\mathbf{g}'$ and the center of the corneal sphere $\mathbf{C}$. (b)Corneal reflection and gaze reflection point (GRP) $\mathbf{T}$[2]

5) Mapping GRP to the omnidirectional image and obtain the PoG. Now we have the GRP $\mathbf{i_T}$ and a warping function $\mathbf{W}_{t^*}$, thus, can compute the PoG in an image $\mathbf{I}_{t^*}$ by $\mathbf{j}_{t^*} = \mathbf{W}_{t^*}(\mathbf{i_T})$. We transform $\mathbf{j}_{t^*}$ to an omnidirectional image point $\mathbf{k} = (\phi_{\mathbf{k}}, \theta_{\mathbf{k}})$ as follows:

$$\begin{aligned} \phi_{\mathbf{k}} &= \arctan(1 - 2j_x/600) + (t^* - 1)\pi/16, \\ \theta_{\mathbf{k}} &= \arctan(1 - 2j_y/600). \\ \text{where } \mathbf{j}_{t^*} &= \begin{bmatrix} j_x & j_y \end{bmatrix}^{\mathbf{T}}. \end{aligned} \tag{10}$$

However, when the GRP $\mathbf{i_T}$ is transformed to out of the image region of $\mathbf{I}_{t^*}$, we perform scheme as follows:

$$t^* = \begin{cases} t^* - 2 & \text{if } j_x > N \\ t^* + 2 & \text{if } j_x < 0. \end{cases}. \tag{11}$$

Here, t^* is circulated ranging from 1 to 16.

4 Experiments

We conducted two experiments in two indoor scenes and five outdoor scenes (Figure 8). The first evaluates the robustness of the single point registration algorithm and the second examines the angular accuracy of the mapped PoG. For each scene, 1 - 3 subjects looked at 10 - 28 instructed points, and eye images were taken by an corneal imaging camera. The PoG estimation system was implemented on MATLAB R2015a and worked on an Intel Core i7-4790K 4.00GHz CPU and 32GB RAM PC.

The robustness of the single point registration algorithm. For each scene and eye corneal image, we evaluated how many images were correctly matched. Table 1 shows the results. In total, the registration robustness was 78.6% in indoor scenes and 85.8% in outdoor scenes. However, the performance in indoor scene 2 was especially low, which was caused by noisy eye images due to the low intensity.

The angular accuracy of the mapped PoG of the omnidirectional images. In successfully registered image pairs, we examined the angular accuracy. For each scene and subject, we obtained the PoG from their eye images and calculated the angular error using the ground truth (instructed points), where $\mathbf{R}_{\text{offset}}$ is -0.10 [rad] rotation about x-axis. Table 1 and Figure 7 shows the results. The angular errors of the PoG in indoor scenes were 5.953 [deg] on average, ranging from 4.039 [deg] to 6.929 [deg], and those in outdoor scenes were 5.391 [deg] on average, ranging from 3.101 to 11.252 [deg].

Figure 8. Examples of scene images and eye images for the experiment.

Table 1. experimental results

Scene		Subjects number	Successfully registered		Angular error M[deg]	SD
in-door	Sc1	1	26 / 26	(100.0%)	6.929	4.102
	Sc2	1	5 / 10	(50.0%)	4.039	2.235
	Sc2	2	7 / 10	(70.0%)	4.787	1.878
	Sc2	3	6 / 10	(60.0%)	4.678	2.342
out-door	Sc1	1	10 / 10	(100.0%)	5.928	2.431
	Sc1	2	8 / 10	(80.0%)	3.823	1.153
	Sc1	3	10 / 10	(100.0%)	8.358	3.761
	Sc2	1	7 / 10	(70.0%)	3.747	1.825
	Sc2	2	8 / 10	(80.0%)	7.549	3.464
	Sc2	3	8 / 10	(80.0%)	5.944	1.660
	Sc3	1	11 / 12	(91.7%)	3.736	1.954
	Sc3	2	11 / 12	(91.7%)	11.252	5.006
	Sc3	3	8 / 12	(66.7%)	6.307	3.006
	Sc4	1	7 / 10	(70.0%)	4.851	3.629
	Sc4	2	10 / 10	(100.0%)	3.608	1.639
	Sc4	3	10 / 10	(100.0%)	3.784	1.013
	Sc5	1	10 / 12	(83.3%)	3.181	1.518
	Sc5	2	12 / 12	(100.0%)	3.101	1.040
	Sc5	3	9 / 12	(75.0%)	5.515	2.488
Indoor scenes total			44 / 56	(78.6%)	5.953	
Outdoor scenes total			139 / 162	(85.8%)	5.391	
All scenes total			183 / 218	(83.9%)	5.526	

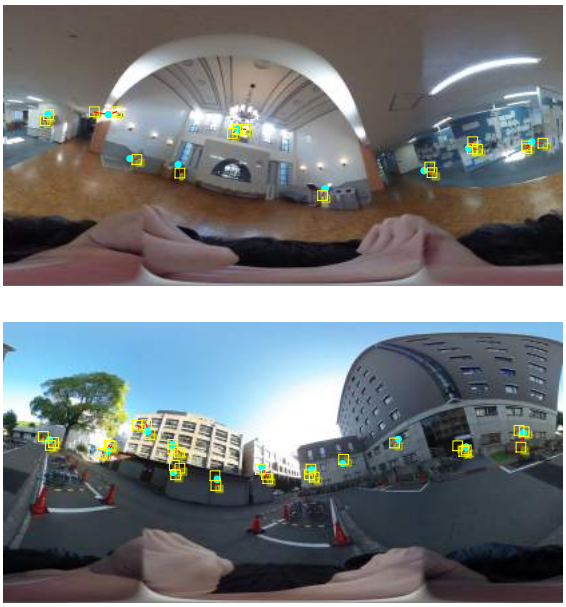

Figure 7. The results of PoG mapping in indoor scene 2 and outdoor scene 5. Yellow squares are estimated PoG (3 users), cyan circles are the ground truth and red lines are errors.

5 Discussion

Registration tendency. Local feature-based registration is used in our system. The registration was successful for more than 80 % of eye images, however, it failed for several eye images due to a lack of feature points in eye images and noise of eye images. Thus, the registration is tenderly more robust in scenes where there are many objects than scenes where we only can look at the sky, the ground, and walls.

The PoG mapping error. The angular error is caused by the error of the warping function through the single-point registration. In the experiments, the farther matched image points were from the GRP, the more error the mapped PoG had. The error can be diminished by the direction-revising scheme, however, this problem still remains.

6 Conclusion

We show a human point of gaze estimation system using eye corneal reflection and an omnidirectional camera image. The average of gaze estimation errors is 5.526 [deg], which are slightly larger than that of the current EGT systems. However, the EGT systems require the system calibration since they rely on geometric PoG estimation. The proposed system does not require calibrations and does not suffer from device drifting. Moreover, scene images can be prepared off-line, therefore, do not suffer from privacy issues. Using our system, we can observe all-round gaze in an unified scene image information. This shows the potential to the marketing and outdoor training system.

References

[1] Ko Nishino and Shree K Nayar. Corneal imaging system: Environment from eyes. *International Journal of Computer Vision*, 70(1):23–40, 2006.

[2] Atsushi Nakazawa, Christian Nitschke, and Toyoaki Nishida. Non-calibrated and real-time human view estimation using a mobile corneal imaging camera. In *Multimedia & Expo Workshops (ICMEW), 2015 IEEE International Conference on*, pages 1–6. IEEE, 2015.

[3] Atsushi Nakazawa, Christian Nitschke, and Toyoaki Nishida. Registration of eye reflection and scene images using an aspherical eye model. *Journal of the Optical Society of America A*, 33:2264–2276, 2016.

[4] Kentaro Takemura, Tomohisa Yamakawa, Jun Takamatsu, and Tsukasa Ogasawara. Estimating focused object using corneal surface image for eye-based interaction. In *3rd International Workshop on Pervasive Eye Tracking and Mobile Eye-Based Interaction*, 2013.

[5] David G Lowe. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2):91–110, 2004.

[6] Herbert Baya, Andreas Essa, Tinne Tuytelaarsb, and Luc Van Goola. Speeded-up robust features (surf). *Computer vision and image understanding*, 110(3):346–359, 2008.

[7] Martin A Fischler and Robert C Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395, 1981.