See What a Strabismus Patient Sees Using Eye Robots

Yidi Huang[1], Qi Wei[2], Joseph L. Demer[3], and Ningshi Yao[1*]

Abstract— Ocular mobility disorders such as strabismus affect millions of people. Patients' descriptions of their symptoms, such as what they see and how their vision has changed, are important for ophthalmologists to diagnose, monitor progression, and evaluate treatment effectiveness. However, such verbal depiction may be vague and subjective. A data-driven simulator that visualizes abnormal vision experienced by a strabismic patient can be helpful to objectively illustrate each individual's vision condition and thus to better understand and manage strabismus. To fulfill this technical void, this paper presents the first vision visualization robot that uses human eye movement data to simulate strabismic vision. We developed a robotic binocular eye platform, which is capable of displaying simulated visual scenes using its onboard cameras. Based on the hypothesis that a human's binocular vision fusion process can be mimicked as a homography transformation from one view to another view, we developed a pipeline to estimate the time-varying homography matrix, and generate the fused view of a human's binocular vision. The effectiveness of the proposed method is demonstrated through experiments with eye movement data from both healthy individuals and strabismic patients.

I. INTRODUCTION

Various vision disorders pertain to the human eye, such as strabismus, affecting more than 18 million people in the United States [1]. Such vision problems can be debilitating because people with strabismus develop double/blurred vision, eyestrain, or other symptoms impairing daily activities. The diagnosis of strabismus is currently based on ocular motility examination and the symptoms of a patient are subjectively conveyed through verbal communication. The descriptions of patients can be vague and reveal insufficient information [2], limiting their value in effective diagnosis and examination. A subject-specific visualization tool that simulates what a patient sees is needed for both patients and ophthalmologists.

Robotic eyes can be a useful tool to advance our understanding of human eye movement under both normal and abnormal conditions. Within the past two decades, several robotic eye systems have been developed. Different actuation models were explored to imitate extraocular eye muscles in robotic eyes [3]–[6]. A bio-mimetic eye that mimicked human nerve fibers [7] was also developed. To simulate human vision, binocular models and bionic eyes were designed to calculate the 3D position of the object of interest [8], [9]. In [10], bionic eyes used images of the environment to estimate the initial location and orientation of the cameras with respect to the object of interest. In addition to simulating the 3D localization functionality of human eyes, a spherical parallel mechanism has been used to coordinate eye movements to maintain a stable gaze during head motion [11]. Although these works have made contributions to binocular vision, they assumed that the two eye robots are aligned so that the visual scenes perceived by the robots are clear, mimicking normal vision conditions. However, strabismic patients usually experience blurry vision. Therefore, previous bionic eyes cannot be directly applied to simulate strabismus vision.

In this paper, we present the **first** human vision visualization robotic platform, which can enable others, especially ophthalmologists, to see what patients with strabismus see. We developed a robotic eye platform that includes 2 robotic eyes. The robotic eyes are able to independently rotate horizontally and vertically either to fixate on a target or to follow an eye movement trajectory. The target locations and eye rotations recorded from subjects with normal vision, as well as from strabismic patients were expressed as horizontal and vertical angles, which were used as target rotation angles for the robotic eyes to simulate. While imitating human eye movement, robotic eyes utilize two onboard cameras to record human's corresponding visual scenes. Based on the proposed hypothesis that human binocular vision processing can be represented as a homography transformation from one view to another, we developed a method to estimate the time-varying homography matrices that are used to fuse the different scene images viewed by the left- and right-sight cameras, simulating human binocular vision. The robotic eye-driven visualization platform is able to generate a fused video, which can vividly show double and/or blurry vision experienced by a strabismic patient. The effectiveness of the proposed methods is analyzed in experiments with 7 volunteers and 1 strabismic patient's binocular misalignment data.

Different from the current clinical diagnosis, which is an ocular motility test in 5 fixed eye positions and provides subjective and qualitative information [12], the proposed platform can generate objective, quantitative, and transient vision information. To the best of our knowledge, there are no other existing platforms/systems that provide such subject-specific binocular vision visualization.

[1]Yidi Huang (yhuang35@gmu.edu) and Ningshi Yao (nyao4@gmu.edu) are with the Department of Electrical and Computer Engineering, George Mason University, Fairfax, VA, 22030, USA.

[2]Qi Wei (qwei2@gmu.edu) is with the Department of Bioengineering, George Mason University, Fairfax, VA 22030, USA.

[3]Joseph L. Demer (jld@jsei.ucla.edu) is with the Departments of Ophthalmology and Stein Eye Institute, University of California, Los Angeles, CA 90054, USA.
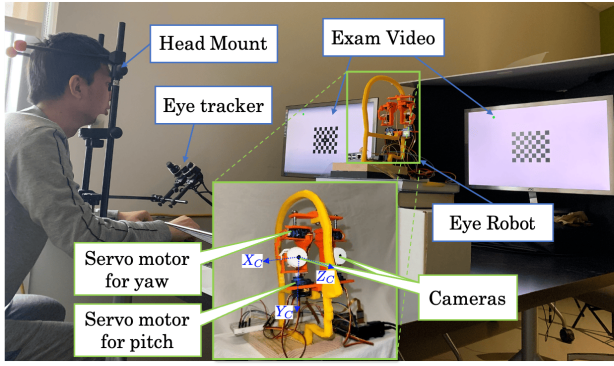
*Send correspondence to Ningshi Yao.

Fig. 1. Overall human vision visualization system.

II. ROBOTIC EYE PLATFORMS

The proposed human vision visualization system is shown in Fig. 1. A participant is side-by-side with the robotic eyes in front of two identical monitors so that the human and the robot can look at the same video displayed on the screens.

1) Robot Hardware Setup: The robotic eyes consists of four servo motors for controlling yaw and pitch angles, two camera and two micro-processors, as shown in Fig. 1. We use the SG90 Micro Servo Motors as the actuators to change the robotic eye's pitch and yaw angle. The change of roll angles is not considered in this work because the visual effects caused by changes in roll angle in human eyes are minor. Raspberry Pi Spy Cameras are utilized to capture images that mimic the optical imaging of the human eye's retina. Each robotic eye ball with an onboard camera is controlled by two servos to simulate eye movements in horizontal and vertical (yaw and pitch, respectively) directions. The processing boards are the Raspberry Pi Compute Module 4 B+ Chip. We employed two separate boards to independently control two robotic eyeballs and their corresponding servo sets. The platform frame is 3D printed. In Fig. 1, the camera is embedded within a fixed white eye socket holder, with the servo below directly controlling the camera's yaw rotation. The servo above controls the pitch rotation via a pull rod.

2) Robot Model: We denote the 3D position of each robotic eye as $\boldsymbol{D}_i^R$, where the subscription $i \in \{l, r\}$ distinguishes left and right eyes, and the superscription R remarks the variable for the robot. The euler angles of each eye at time t are denoted as $\boldsymbol{x}_i^R(t) = \left[\psi_i^R(t), \theta_i^R(t), \phi_i^R\right]^T$, where $\psi_i^R(t)$ and $\theta_i^R(t)$ are the time-varying yaw and pitch angles, respectively. The roll angle is denoted as ϕ_i^R, which is a constant.

The camera on each robotic eye is modeled as a pinhole camera. Denote the intrinsic matrix as C_{in}, which depends on the focal length and optical center of the camera on robot eye i, and extrinsic matrix $C_{ex} = \begin{bmatrix} \boldsymbol{R}(\boldsymbol{x}_i^R(t)) & \boldsymbol{D}_i^R \\ 0 & 1 \end{bmatrix}$, where $\boldsymbol{R}(\boldsymbol{x}_i^R(t))$ is a 3×3 rotation matrix. Then the projection of a 3D point to camera image plane can be represented as

$$\lambda[u, v, 1]^T = C_{in} \cdot C_{ex} \cdot [P_x, P_y, P_z, 1]^T \quad (1)$$

where $[P_x, P_y, P_z, 1]^T$ and $[u, v, 1]^T$ are the homogeneous coordinates of a 3D point in world frame and its projection 2D projection in image. The image captured by each robotic
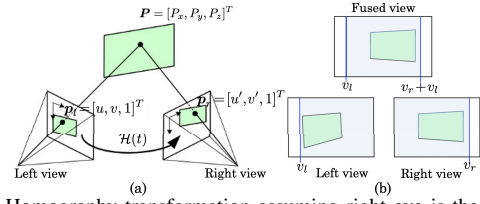

Fig. 2. Homography transformation assuming right eye is the prime eye. (a) Homography transformation in 3D; (b)Common regions of two views.

eye is denoted as $\boldsymbol{I}_i^R(\boldsymbol{x}_i^R(t), \boldsymbol{D}_i^R)$ with $\boldsymbol{I}_i^R(u, v; \boldsymbol{x}_i^R(t), \boldsymbol{D}_i^R)$ denoting the light intensity in grey scale at pixel (u, v).

III. PROBLEM FORMULATION

The aim of this work is to present the visual scene perceived by strabismic patients in the form of a video using robotic eyes. To facilitate human vision visualization, we make the following assumptions.

Assumption 1: The visual scene perceived by each eye can be approximated by digital images.
The retina consists of photoreceptor cells arranged in a grid-like pattern. Each cell corresponds to a specific area of the visual field and contributes to the formation of a complete image, which is similar to a digital image with pixels.

Assumption 2: The visual field size of both eyes is the same, regardless of whether a human has strabismus or not. The assumption is intuitively valid because patients with strabismus only have disorders with eye movement and their vision perception functions are healthy.

Denote the 3D position of the center of a human eye ball as $\boldsymbol{D}_i^H$, where the subscription $i \in \{l, r\}$ denotes left and right, and the superscription H marks the variables are humans'. At time t, the yaw, pitch, and roll angles of the eyes of a human are defined as $\boldsymbol{x}_i(t) = [\psi_i(t), \theta_i(t), \phi_i]^T$ for $i \in \{l, r\}$. Based on *Assumptions* 1 and 2, the visual scene of each eye can be represented as an image, denoted as $\boldsymbol{I}_i^H(\boldsymbol{x}_i(t), \boldsymbol{D}_i^H)$. Since each human eye is located at a different position with a different eye gaze angle, the image seen by each eye is not the same, i.e., $\boldsymbol{I}_l^H(\boldsymbol{x}_l(t), \boldsymbol{D}_l^H) \neq \boldsymbol{I}_l^H(\boldsymbol{x}_r(t), \boldsymbol{D}_r^H)$ for any t. The left and right views share a common region, which is illustrated in Fig. 2(b), which is the visual scene both eyes see. We define the left boundary of the common region of the left view as v_l, and the right boundary of the common region of the right view as v_r.

For healthy people without strabismus, the fused view from two eye views should be clear after human brain's processing. Based on this intuition, we assume that the human's binocular vision fusion can be approximated as the homographic transformation from the non-prime eye to prime eye such that transformed common regions completely overlap with each other. Without loss of generality, we will use right view as the prime in the following sections.

Assumption 3: For an arbitrarily small constant $\epsilon > 0$, there exists a homography matrix $\mathcal{H}(t)$ from left view to right view such that

$$\sum_{u=1}^{U} \sum_{v=v_l}^{V} \left| \boldsymbol{I}_l^H(u, v; \boldsymbol{x}_l(t), \boldsymbol{D}_l^H) - \boldsymbol{I}_r^H(u', v'; \boldsymbol{x}_r(t), \boldsymbol{D}_r^H) \right|^2 < \epsilon$$

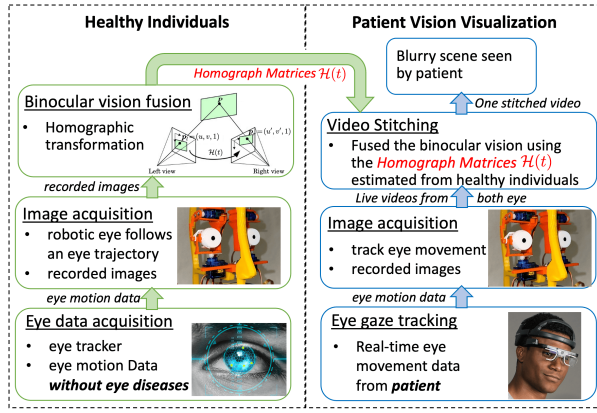$$\text{s.t. } [u', v', 1]^T = \mathcal{H}(t)[u, v, 1]^T, \quad (2)$$

Fig. 3. Overall flow diagram of the proposed pipeline. (left) Estimation process to compute the homography matrix of human's binocular vision; (right) Vision visualization process of a patient.

Fig. 4. Videos watched by humans.(a) Exam video, where the blue arrows represent the trajectory the green dot travels. (b) Position tuning video

Fig. 5. Flow diagram of robot control and image acquisition step.

where U and V represents the size of the image in pixels. Then the fuse view within the common region from the two eye views can be expressed as

$$\boldsymbol{I}^H(u', v'+v_l; t) = \frac{1}{2}\boldsymbol{I}_l^H(u, v; \boldsymbol{x}_l(t), \boldsymbol{D}_l^H) \quad (3)$$
$$+ \frac{1}{2}\boldsymbol{I}_r^H(u', v'; \boldsymbol{x}_r(t), \boldsymbol{D}_r^H) \text{ with } [u', v', 1]^T = \mathcal{H}(t)[u, v, 1]^T$$

for any u and v within the common region, where $\boldsymbol{I}^H(u', v'; t)$ represents the visual scene after our brains processes the images from two eyes.

However, the fused view represented in (3) cannot be presented to any other humans, due to two major problems:

Problem 1: There is no existing tool to measure/capture the images perceived by each eye, i.e., $\boldsymbol{I}_l^H(\boldsymbol{x}_l(t))$ and $\boldsymbol{I}_r^H(\boldsymbol{x}_r(t))$ of a human are not available to others. To overcome this challenge, we utilized the proposed robotic eye platform and designed controllers for the robotic eye such that the angles of the robotic eyes $\boldsymbol{x}_i^R(t)$ can track the human eye movement $\boldsymbol{x}_i(t)$ for any time t. And tune the positions of the robotic eye $\boldsymbol{D}_i^R$ such that the captured images from robotic eye can mimic what people see from each eye, i.e., $\boldsymbol{I}_i^R(\boldsymbol{x}_i^R(t), \boldsymbol{D}_i^R) \approx \boldsymbol{I}_i^H(\boldsymbol{x}_i(t), \boldsymbol{D}_i^H)$. The detailed designs will be introduced in Section IV-A.

Problem 2: The homography matrix $\mathcal{H}(t)$ of human vision is unknown.
To obtain the time-varying matrix $\mathcal{H}(t)$, we proposed and verified a hypothesis that the view fusion process in our brains does not differ much with or without strabismus, and presented a pipeline to estimate the homography matrix $\mathcal{H}(t)$ using data collected from healthy individuals, which can be used for strabismus patients vision visualization.

IV. METHODOLOGY

The major steps of the proposed visualization method are illustrated by Fig. 3. We use an eye gaze tracker to obtain human's eye angle data. Then each robotic eye is controlled to track the real human eye gaze data and record what a human sees. The data from healthy individuals is utilized to estimate the homography matrices of human's binocular vision. Based on the estimated homography matrices, a video stitching technique is applied to generate the fused view.
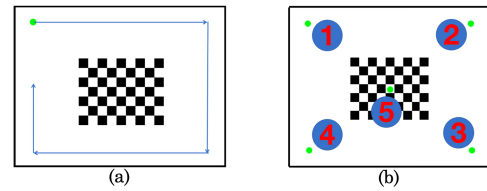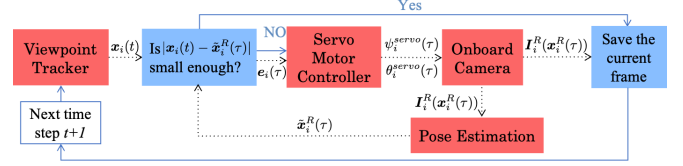
A. Exam Video and Eye Gaze Data Collection

We design an exam video for participants to watch, such that all participants look at the same scene. The same exam video is used by the robotic eyes to capture frame images. Participants are asked to focus on the moving green dot during the data collection, so that we can capture the transient behaviors of human eye movement. As shown in Fig. 4(a), a static chessboard is positioned at the center of the video, which is distinguishable for robotic eyes to process. A moving green dot travels clockwise around the chessboard at a constant speed, following the trajectory shown by the arrows in Fig. 4(a).

The Viewpoint Eye Tracker® [13] is utilized to collect human eyes' pitch and yaw angles $\boldsymbol{x}_l(t)$ and $\boldsymbol{x}_r(t)$. Once the human eye positions $\boldsymbol{D}_i^H, i \in \{l, r\}$ are fixed (the head mount shown in Fig. 1 of the eye tracker is installed at a fixed location), determining the proper initial position $\boldsymbol{D}_i^R$ of the robotic eyes platform such that $\boldsymbol{I}_i^R(\boldsymbol{x}_i^R(t), \boldsymbol{D}_i^R) \approx \boldsymbol{I}_i^H(\boldsymbol{x}_i(t), \boldsymbol{D}_i^H)$ is non-trivial, because the focal length of human eyes is unknown. The range of $\boldsymbol{I}_i^H$ is hard to decide. Therefore, we proposed an additional step to determine where to place the robotic eyes. It is based on the fact that when human's views focus on the green dot, the projection of the green dot should be at the center on each side of view.

Hence, before the exam video is displayed to a human, we first require the human to watch a calibration video where the green dot appears sequentially at the five pre-define points in the screen as stimuli for participants to fixate on, as demonstrated by Fig. 4(b). The corresponding eye angles $\boldsymbol{x}_i(t_1), ..., \boldsymbol{x}_i(t_5)$ are recorded during the observation of these five points, and used as rotation angles for the robotic eyes. The robotic eye platform's position $\boldsymbol{D}_i^R$ is adjusted so that the center point of each side view of the robotic eye overlaps with the green dots for every $\boldsymbol{x}_i^R(t_j) = \boldsymbol{x}_i(t_j)$, where $j = 1, \ldots, 5$, respectively.

B. Robotic Eye Rotation Control and Image Acquisition

The goal of this step is to ensure that $\boldsymbol{x}_i^R(t) \approx \boldsymbol{x}_i(t)$ and $\boldsymbol{I}_i^R(\boldsymbol{x}_i^R(t), \boldsymbol{D}_i^R) \approx \boldsymbol{I}_i^H(\boldsymbol{x}_i(t), \boldsymbol{D}_i^H)$ at any time t. To achieve this goal, we present an approach to collect correct images at the right time instant.

Due to the limited precision and dead band width of the micro servo motors, which is approximately between 0.06 and 0.1 degrees, the servo may not respond in a certain range near the target angle when it is close to the desired angle $\boldsymbol{x}_i(t)$. And the time for $\boldsymbol{x}_i^R$ to converge to the target angle $\boldsymbol{x}_i(t)$ is much larger than the sampling time of the eye tracker. Therefore, an offline robotic eye control method and image acquisition method are adopted, as illustrated by Fig. 5, instead of a real-time target angle tracking for the robotic eyes. For each instant t, when the human eye gaze angles are measured by the eye tracker as $\boldsymbol{x}_i(t)$, the angles $\boldsymbol{x}_i^R$ of the robotic eyes are adjusted in a slower time τ. When the difference of between $\boldsymbol{x}_i^R(\tau)$ and the tracking target $\boldsymbol{x}_i(t)$ is small enough, which we set to be $|\boldsymbol{x}_i(t) - \boldsymbol{x}_i^R(\tau)| \leq 0.25$ degree, then the current frame $\boldsymbol{I}_i^R(\boldsymbol{x}_i^R(\tau), \boldsymbol{D}_i^R)$ is stored as a proper approximation of $\boldsymbol{I}_i^H(\boldsymbol{x}_i(t), \boldsymbol{D}_i^H)$ and then move on to the next sampled data $\boldsymbol{x}_i^H(t+1)$ obtained by eye tracker. Some design details are explained further in the following.

1) Robotic Eye Angle Measurement: To measure the current Euler angles of each robotic eye, we first use standard camera calibration to obtain the intrinsic matrix [14]. Then after detecting the corner points on the chessboard through Harris corner detection algorithm [15], we use the perspective-n-point (or PnP) method to perform camera pose estimation [16] in order to calculate the camera's extrinsic matrix C_{ex}. The matrix C_{ex} includes an estimated rotation matrix $\tilde{\boldsymbol{R}}$. The current yaw and pitch angles of the camera, $\tilde{\psi}^R(\tau)$ and $\tilde{\theta}^R(\tau)$, can be calculated based on $\tilde{\boldsymbol{R}}_i$ as $\tilde{\psi}^R(\tau) = \mathrm{atan2}(\tilde{\boldsymbol{R}}_{21}, \tilde{\boldsymbol{R}}_{11})$ and $\tilde{\theta}^R(\tau) = \mathrm{atan2}(-\tilde{\boldsymbol{R}}_{31}, \sqrt{\tilde{\boldsymbol{R}}_{32}^2 + \tilde{\boldsymbol{R}}_{33}^2})$.

2) Robotic Eye Controller: The robotic eye control use a proportional-derivative (or PD) controller, which continues adjusting the servo motor until the difference between current camera angle $(\tilde{\psi}_i^R(\tau), \tilde{\theta}_i^R(\tau))$ and the target angle $(\psi_i(t), \theta_i(t))$ for $i \in \{l, r\}$ meets the error tolerance. Define $e_i^{\psi}(\tau) = \psi_i(t) - \tilde{\psi}_i^R(\tau)$ and $e_i^{\theta}(\tau) = \theta_i(t) - \tilde{\theta}_i^R(\tau)$. The control laws are designed as $\psi_i^{servo}(\tau) = K_P^{\psi} \cdot e_i^{\psi}(\tau) + K_D^{\psi} \cdot \dot{e}_i^{\psi}(\tau)$ and $\theta_i^{servo}(\tau) = K_P^{\theta} \cdot e_i^{\theta}(\tau) + K_D^{\theta} \cdot \dot{e}_i^{\theta}(\tau)$, where K_p^{ψ}, K_P^{θ}, K_D^{ψ}, and K_D^{θ} are PD control coefficients, and $(\psi_i^{servo}(\tau), \theta_i^{servo}(\tau))$ are the additional rotation angle of the servo on robotic eye i from the previous angle.

C. Homography Matrices Estimation

When a human fuses two views together, especially for strabismic patients, we propose the following hypothesis

Hypothesis 1: The homography matrix $\mathcal{H}(t)$ of healthy individuals and strabismic patients are similar, if the humans are looking at the same scene from the same distance.
This hypothesis allows us to leverage the *Assumption* 3 as a prior knowledge to estimate the $\mathcal{H}(t)$ using the eye movement data of people with normal vision, and transform the estimated $\mathcal{H}(t)$ to patients with strabismus.

After obtaining the two images of a human's reference angles at time t, the Harris corner detector is used again to obtain two sets of corner points in the chessboard. Denote $\boldsymbol{p}_i^c$ as the 2D coordinate of a corner point in image captured by the robot eye i. The homography matrix estimation can

participant #	interocular distance	myopia	astigmatism
1	68mm	none	none
2	66mm	350(l), 175(r)	125(l), 175(r)
3	70mm	350(l), 450(r)	0(l), 50(r)
4	70mm	125(l), 75(r)	25(l), 25(r)
5	61mm	none	none
6	65mm	none	none
7	63mm	none	none

be formulated as a least min-square optimization

$$\mathcal{H}^*(t) = \underset{\mathcal{H}(t)}{\mathrm{argmin}} \sum_{\boldsymbol{p} \in \{\boldsymbol{p}_q^c\}} \left| \boldsymbol{I}_l^R(\boldsymbol{p}, t) - \boldsymbol{I}_r^R(\boldsymbol{q}, t) \right|^2 \qquad (4)$$
$$\text{s.t. } [\boldsymbol{q}, 1]^T = \mathcal{H}(t)[\boldsymbol{p}, 1]^T$$

where $\boldsymbol{p}$ and $\boldsymbol{q} \in \mathbb{R}^2$ are the 2D coordinates of corner points. The homographic matrix can be obtained by solving (4), or using deep learning approaches [17].

D. Binocular Vision Fusion

Once $\mathcal{H}^*(t)$ is calculated, the perspective warping [18] can be applied to one of the images, typically the image that is being transformed to align with the other image. The two views can be stitched together to generate the fused view $\boldsymbol{I}^R$ as $\boldsymbol{I}^R(u, v; t) = \boldsymbol{I}_l^R(u, v; t)$ if $v < v_l$, $\boldsymbol{I}^R(u, v; t) = \boldsymbol{I}_r^R(u, v - v_l; t)$ if $v > v_l + v_r$, and $\boldsymbol{I}^R(u, v; t) = \frac{1}{2}\left[\boldsymbol{I}_l^R(u', v'; t) + \boldsymbol{I}_r^R(u, v - v_l + 1; t) \right]$ otherwise, where $[u, v, 1]^T = \mathcal{H}^*(t)[u', v', 1]^T$.

For strabismic patients, $\mathcal{H}^*(t)$ cannot be directly estimated because the *Assumption* 3 does not hold for them. However, if the *Hypothesis* 1 is valid, which will be verified in the experiment Section V-B, we can use the homography matrices $\mathcal{H}^*(t)$ calculated from healthy humans to patients. We input the eye angle data of the strabismic patient, denoted as $\boldsymbol{x}_i^P(t)$, into the robotic eyes as target angles, and save the resulting images $\boldsymbol{I}_i^R(\boldsymbol{x}_i^P(t), \boldsymbol{D}_i^R)$. However, instead of directly computing $\mathcal{H}^*(t)$ for the strabismic patient, we apply the homography matrix $\mathcal{H}^*(t)$ obtained from healthy participants to stitch the two views $\boldsymbol{I}_i^R(\boldsymbol{x}_i^P(t), \boldsymbol{D}_i^R)$ together, thus obtaining the fused views of the strabismic patient.

V. EXPERIMENT RESULTS

This section presents the experimental results, which verify *Hypothesis* 1, and the visualized double vision of patient. We recruited 7 volunteers with no ophthalmic diseases. Volunteers include 3 men and 4 women, aging from 20 to 32 year-old. Their specific information is listed in Table I, where symbol (l) and symbol (r) represents data for the left and right eyes, respectively.

A. Experiment Setup

Two monitors are used to show the exam videos at the same height relative to the same table and with a certain distance away from the eye tracker head mount, as shown in Fig. 1. The center of human eyes was positioned at the same height as the center of the monitor screen. The eye tracker was also positioned in front of the head mount. The
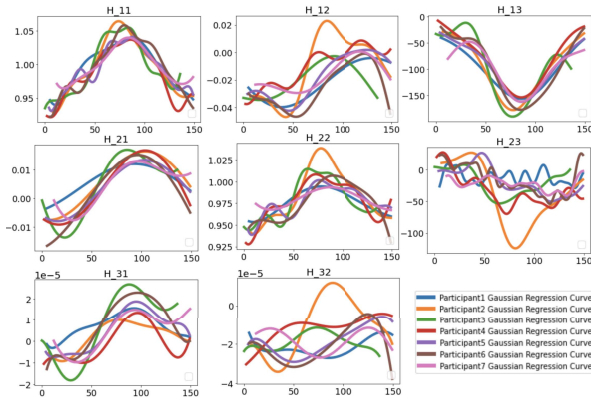
Fig. 6. The estimated $\mathcal{H}(t)$ of the 7 participants. The x-axis represent time and the y-axis represents the value of an entry of $\mathcal{H}(t)$.
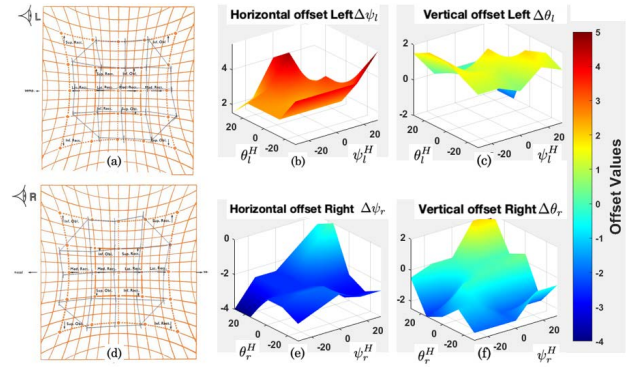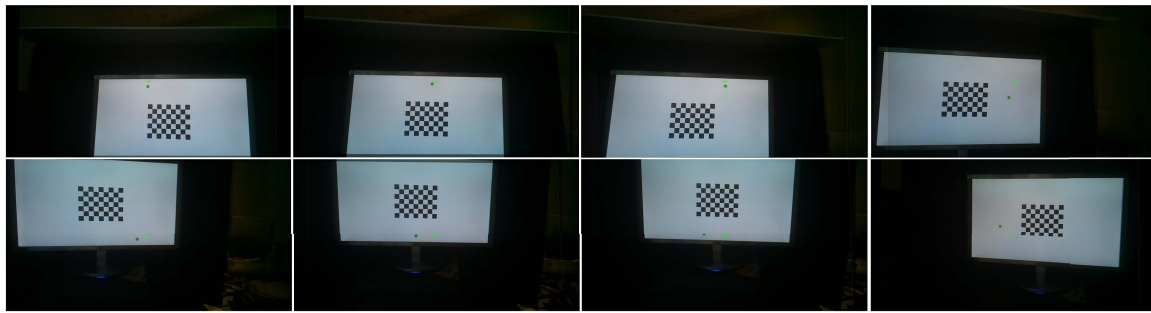

Fig. 7. Misalignment data of a strabismic patient: (a) ocular motility data of left eye; (b) interpolated yaw shift of left eye; (c) interpolated pitch shift of left eye; (d) ocular motility data of right eye; (e) interpolated yaw shift of right eye; (f) interpolated pitch shift of right eye.

centers of robotic eyes were aligned at the same height as the center of its corresponding screen. The exam video is designed to be 15 seconds long, recorded at a frame rate of 120 frames per second (or fps). The Viewpoint Tracker software is configured to acquire data at a frequency of 10 Hz, resulting in the capture of 150 sets of angles for each participant. The PD control parameters for the servo motor are $K_P^{\psi} = K_P^{\theta} = 0.25$, and $K_D^{\psi} = K_D^{\theta} = 0.05$.

During the eye gaze collection, all participants watched the same exam video under the same indoor lighting condition, to ensure that this environmental variable does not affect the results. We required each participant to focus on the green dot only by moving his/her eyes, while holding the head as static as possible.

B. Homography Transformation Verification

In this validation experiment, the data collected from 7 healthy individuals was used to calculate homography matrices corresponding to the 150 sampled data from the eye tracker. The estimated $\mathcal{H}^*(t)$ matrices of each volunteer were plotted over time. Each entry of $\mathcal{H}^*(t)$ (excluding $\mathcal{H}_{33}$, which is always 1) is shown by a sub-figure in Fig. 6. We can see from this figure that most entries of the estimated $\mathcal{H}^*(t)$ are the same among 7 participants. Even though the curves of $\mathcal{H}_{12}$, $\mathcal{H}_{31}$, and $\mathcal{H}_{32}$ look very different, the magnitudes are very small. Therefore, these entries vary within a very small range. The only noticeable difference occurs in $\mathcal{H}_{23}$, which represents the transnational movement of the view in the vertical directions. The difference is mainly due to errors in the Eye Tracker measurement. Some participants inevitably blinked their eyes, so the corresponding data from those participants is inaccurate in vertical translation. However, the similarity in the graphical trends of the estimated $\mathcal{H}^*(t)$, especially the similarity in $\mathcal{H}_{11}$, $\mathcal{H}_{12}$, $\mathcal{H}_{21}$, and $\mathcal{H}_{22}$ which reflects the rotation of view, supports our *Hypothesis*.

C. Binocular Vision Fusion Results

To analyze the effectiveness of the proposed system and view fusion methodology, we want to have a fair comparison between what a with strabismic individual sees and a "ground truth" what this individual should see if he/she does not have any strabismus. Therefore, instead of directly using

the eye movement data from a person with strabismus, we simulated a patient's eye movements $(\psi_i^P(t), \theta_i^P(t)), i \in \{l, r\}$, by combing a healthy participant eye movement data $(\psi_i^H(t), \theta_i^H(t))$ with eye gaze shifts $(\Delta\psi_i, \Delta\theta_i)$ of a strabismic patient's Hess Screen Test data [19]. By doing so, we can distinguish where a patient actually looks at, which is $(\psi_i^P(t), \theta_i^P(t))$, with where a patient should look at, which is the ground truth $(\psi_i^H(t), \theta_i^H(t))$. The strabismus patient's Hess Screen Test data were only measured point-wise at every 15 degrees within a $[-30, 30]$ degrees range vertically and horizontally, as shown by Fig. 7(a) and (d). However, the healthy participant data $(\psi_i^H(t), \theta_i^H(t))$ are not exactly located at those measured positions. Therefore, we bilinearly interpolated the shifts between where the patient looks at and the target grid points, as shown in Fig. 7(b), (c), (e) and (f), for yaw and pitch angles, respectively. Then the simulated patient eye gaze angles can be computed as $\psi_i^P(t) = \psi_i^H(t) + \Delta\psi_i(\psi_i^H(t), \theta_i^H(t))$ and $\theta_i^P(t) = \theta_i^H(t) + \Delta\theta_i(\psi_i^H(t), \theta_i^H(t))$, where $i \in \{l, r\}$.
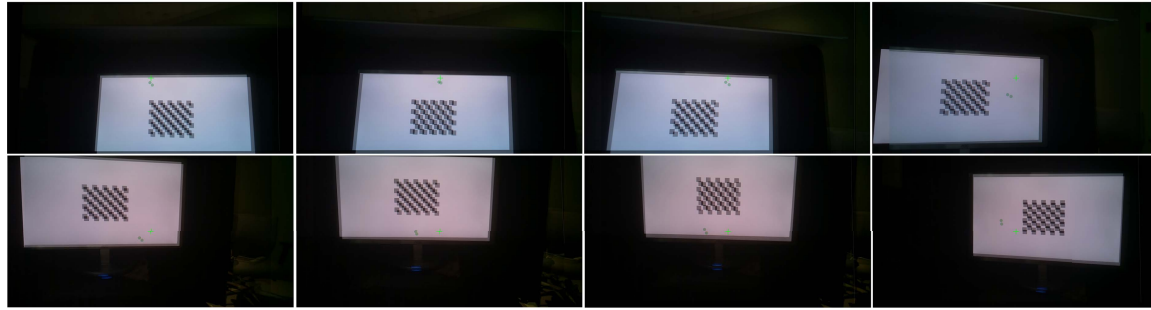
Figure 8(a) shows several video frames of Participant 1's fused view, using the estimated $\mathcal{H}^*(t)$ based on his/her true eye movement data $(\psi_i^H(t), \theta_i^H(t))$. All frames seem clear with one green dot appearing in these frames. The chessboard has no ghosting effect. Fig. 8(b) shows the corresponding video frames of the fused view of the simulated strabismus patient, using Participant 1' data added with the interpolated angle shifts shown in Fig. 7. We can see that each frame has two green dots due to misalignment of two eyes and the chessboard has a ghosting effect. Notice that the magnitude and direction of the green dots' shift are different in each frame. For example, the green dot shift in the sub-figure on the first row and second column of Fig. 8(b) is the least obvious one. The green dots shift horizontally in the sub-figure on the first row and last column, while the green dots shift vertically in the sub-figure on last row and last column. These results are consistent with what happen to strabismus patients, because the degree of misalignment in patients can vary at different eye positions.

VI. CONCLUSIONS AND FUTURE WORKS

We presented a novel solution based on 2-DOF robotic eyes to simulate and visualize human vision, including the

(a) Video frames of the fused vision for Participant 1

(b) Video frames of the fused vision for simulated strabismus patient

Fig. 8. Comparison of the fused view of a healthy individual and a simulated strabismus patient using real patient data.

vision of patients with strabismus. The proposed method can provide patient-specific, objective, and transient visual information to ophthalmologists. Such eye robots can contribute to the development of more effective and targeted treatments for strabismus and other binocular vision disorders, and are also essential for gaining a comprehensive understanding of the complex processes underlying human perception behavior and processing. For future work, the servo motors used on the current robot will be replaced by artificial muscles, so that the actuation of the robotic eyes can be more realistic. Furthermore, more participants will be recruited and more data from strabismic patients will be tested.

VII. ACKNOWLEDGMENTS

REFERENCES

[1] M. Y. Chang, F. G. Velez, J. L. Demer, S. J. Isenberg, A. L. Coleman, and S. L. Pineles, "Quality of life in adults with strabismus," *American journal of ophthalmology*, vol. 159, no. 3, pp. 539–544, 2015.

[2] P. Ichhpujani, G. L. Spaeth, and M. Yanoff, *Expert techniques in ophthalmic surgery*. Jaypee Brothers Medical Publishers, 2019.

[3] A. Lenz, S. R. Anderson, A. G. Pipe, C. Melhuish, P. Dean, and J. Porrill, "Cerebellar-inspired adaptive control of a robot eye actuated by pneumatic artificial muscles," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 39, no. 6, pp. 1420–1433, 2009.

[4] L. Li, H. Godaba, H. Ren, and J. Zhu, "Bioinspired soft actuators for eyeball motions in humanoid robots," *IEEE/ASME Transactions on Mechatronics*, vol. 24, no. 1, pp. 100–108, 2018.

[5] S. K. Rajendran, Q. Wei, and F. Zhang, "Two degree-of-freedom robotic eye: design, modeling, and learning-based control in foveation and smooth pursuit," *Bioinspiration & Biomimetics*, vol. 16, no. 4, p. 046022, 2021.

[6] S. K. Rajendran, Q. Wei, N. Yao, and F. Zhang, "Observability analysis and reduced-order observer design for a super-coiled polymer-driven robotic eye," in *2022 IEEE 61st Conference on Decision and Control (CDC)*. IEEE, 2022, pp. 1385–1391.

[7] L. Gu, S. Poddar, Y. Lin, Z. Long, D. Zhang, Q. Zhang, L. Shu, X. Qiu, M. Kam, A. Javey *et al.*, "A biomimetic eye with a hemispherical perovskite nanowire array retina," *Nature*, vol. 581, no. 7808, pp. 278–282, 2020.

[8] D. Fan, Y. Liu, X. Chen, F. Meng, X. Liu, Z. Ullah, W. Cheng, Y. Liu, and Q. Huang, "Eye gaze based 3d triangulation for robotic bionic eyes," *Sensors*, vol. 20, no. 18, p. 5271, 2020.

[9] S. Boonkwang and S. Saiyod, "Distance measurement using 3d stereoscopic technique for robot eyes," in *2015 7th International Conference on Information Technology and Electrical Engineering (ICITEE)*. IEEE, 2015, pp. 232–236.

[10] Q. Wang, W. Zou, F. Zhang, and D. Xu, "Binocular initial location and extrinsic parameters real-time calculation for bionic eye system," in *Proceeding of the 11th World Congress on Intelligent Control and Automation*. IEEE, 2014, pp. 74–80.

[11] H. Li, J. Luo, C. Huang, Q. Huang, and S. Xie, "Design and control of 3-dof spherical parallel mechanism robot eyes inspired by the binocular vestibule-ocular reflex," *Journal of Intelligent & Robotic Systems*, vol. 78, pp. 425–441, 2015.

[12] K. B. Gunton, B. N. Wasserman, and C. DeBenedictis, "Strabismus," *Primary Care: Clinics in Office Practice*, vol. 42, no. 3, pp. 393–407, 2015.

[13] K. Arrington, "Viewpoint eye tracker," *Arrington Research*, 1997.

[14] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 22, no. 11, pp. 1330–1334, 2000.

[15] C. Harris, M. Stephens *et al.*, "A combined corner and edge detector," in *Alvey vision conference*, vol. 15, no. 50. Citeseer, 1988, pp. 10–5244.

[16] D. Nister, "Camera pose estimation using image features and multi-view geometric constraints," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2. IEEE, 2004, pp. 265–272.

[17] T. Nguyen, S. W. Chen, S. S. Shivakumar, C. J. Taylor, and V. Kumar, "Unsupervised deep homography: A fast and robust homography estimation model," *IEEE Robotics and Automation Letters*, vol. 3, no. 3, pp. 2346–2353, 2018.

[18] M. Brown and D. G. Lowe, "Automatic panoramic image stitching using invariant features," *International Journal of Computer Vision*, vol. 74, no. 1, pp. 59–73, 2007.

[19] Q. Wei, R. A. Clark, and J. L. Demer, "Can binocular alignment distinguish hypertropia in sagging eye syndrome from superior oblique palsy?" *Investigative Ophthalmology & Visual Science*, vol. 63, no. 10, pp. 13–13, 2022.