Visual Servo Control of COVID-19 Nasopharyngeal Swab Sampling Robot

Guebin Hwang, Jongwon Lee, and Sungwook Yang, *Member, IEEE*

Abstract— In this study, we present a visual servo control framework for fully automated nasopharyngeal swab robots. The proposed framework incorporates a deep learning-based nostril detection with a cascade approach to reliably identify the nostrils with high accuracy in real time. In addition, a partitioned visual servoing scheme that combines image-based visual servoing with axial control is formulated for accurately positioning the sampling swabs at the nostril with a multi-DOF robot arm. As the visual servoing is designed to minimize an error between the detected nostril and the swab, it can compensate for potential errors in real operation, such as positioning error by inaccurate camera-robot calibration and kinematic error by unavoidable swab deflection. The performance of the visual servo control was tested on a head phantom model for 30 unused swabs, and then compared with a method referring to only the 3D nostril target for control. Consequently, the swabs reached the nostril target with less than an average error of 1.2±0.5 mm and a maximum error of 2.0 mm via the visual servo control, while the operation without visual feedback yielded an average error of 10.6±2.3 mm and a maximum error of 16.2 mm. The partitioned visual servoing allows the swab to rapidly converge to the nostril target within 1.0 s without control instability. Finally, the swab placement at the nostril among the entire procedure of fully automated NP swab was successfully demonstrated on a human subject via the visual servo control.

I. Introduction

Coronavirus disease 2019 (COVID-19) has become a global pandemic with over 430 million cases worldwide, causing a social and economic crisis [1]. Early diagnosis is regarded as the most critical factor for preventing the rapid spread of COVID-19. However, extensive swab sampling may pose a high risk of cross-infection from close contact with patients. In addition, the heavy workload of medical professionals increases the healthcare burden. Furthermore, the diagnosis result and comfort level during the sampling procedure may be influenced by the degree of skills and fatigue of the medical staffs.

Recent studies have suggested that automation of COVID-19 sample collection can address these issues by performing it in a safe, standardized, and efficient manner [2], [3]. For example, a self-administered swab robot has been
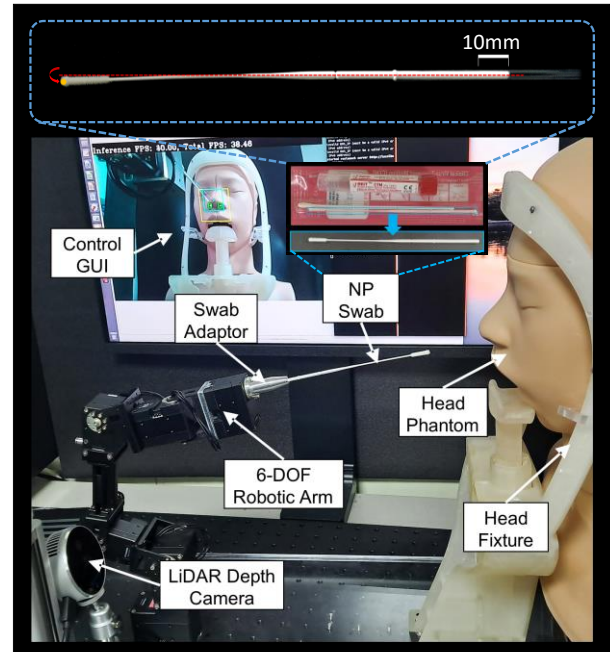
Figure 1.   Test setup for the visual servo control of a nasopharyngeal swab sampling robot with a testing swab.

reported by a research team in Singapore, in which swab insertion and withdrawal operations are only automated, whereas the rest of the procedures are controlled by patients [4]. A remotely operated swab robots have also been presented [5], [6]. Although such teleoperation can reduce the risk of cross-infection and leverage the experience of medical staffs during control, the healthcare burden led by the staffs' intervention still remains.

Fully automated systems have also been introduced, which can perform oropharyngeal (OP) or nasopharyngeal (NP) swabs with minimal intervention by the staffs [7], [8]. For instance, researchers from the University of Southern Denmark developed a fully automatic throat swab robot that identifies the swabbing area in the throat using computer vision, and conducts gentle swabs with force control [7]. Xie *et al*. proposed a robot-assisted OP swab sampling system with a rigid–flexible coupling robot for safe operation [8]. This robot adopts a deep learning-based approach to recognize and place the swab on target positions, such as the tonsil and uvula. Furthermore, it incorporates optimization control with the oral cavity center constraint and force data for sample collection.
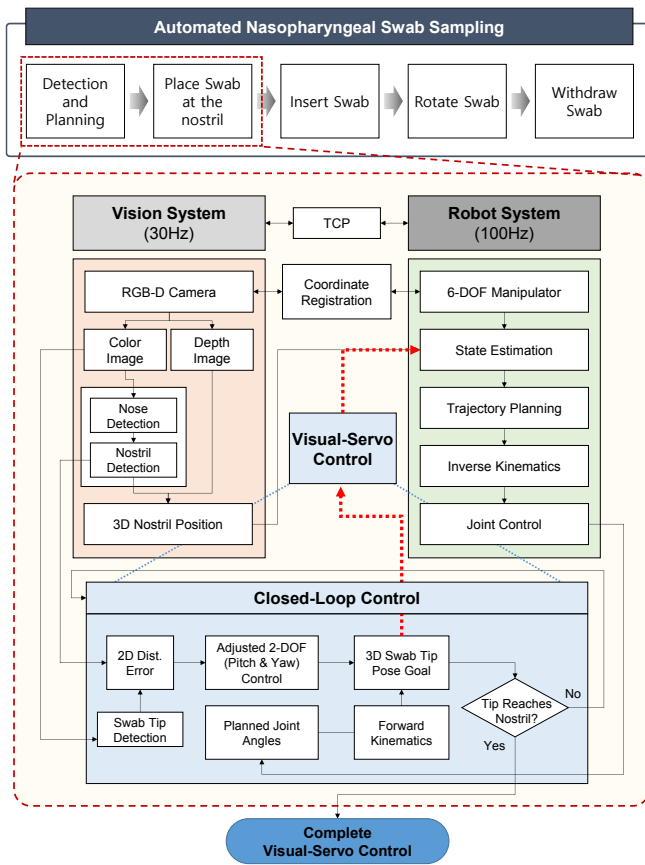
Figure 2. Control flow of automated nasopharyngeal swab sampling.

While nasopharyngeal (NP) swabs are the reference sampling method to detect COVID-19 [9], only a few studies of fully automated NP swabs have been reported. For example, Taiwan's biotechnology startup Brain Navi Co., Ltd. launched a fully automated NP swab robot that can identify the patient's facial structure and precise nostril location for collecting specimens from the nasopharyngeal target position [10]. However, the robot system requires an extra guiding tool; a patient should put on a nasal clip to provide key reference points for the machine operated. However, it may raise issues such as the replacement of the clip and patients' discomfort.

Unlike the OP swabs, the full automation of the NP swabs still poses a challenge in accurately detecting the small nostril and inserting the swab into its small orifice; the average size of the nostrils is 10–12 mm [11], whereas the mean of maximum mouth opening (MMO) in healthy subjects is about 47.0 mm [12]. Consequently, a flocked swab with a diameter of approximately 3 mm must be accurately located at the nostril within an error of 3–4 mm for NP swabs. Although the accurate positioning of the swab can be achieved by precise control of robot systems, any registration error between the camera and robot coordinates may hinder the accurate positioning of the swab on the target. Furthermore, unavoidable kinematic errors caused by deflected swabs may be involved in the control, which considerably hampers fully automated NP swabs in practice. A straight swab model is typically used for solving kinematics. However, the sampling swabs made of slender plastic shafts are fairly flexible, thus unused swabs may

undergo initial deflection during packaging and/or handling the swabs in peel packs even before sample collection as shown Fig. 1.

In this paper, we tackle these issues that arise primarily when placing the swab at the nostril as part of the full automation of the NP swabs. To accurately place the sampling swab at the nostril without an extra guiding tool, a visual servo control framework is proposed with deep learning-based nostril detection. The proposed visual servo control is validated using a custom-built 6-degrees-of-freedom (6-DOF) robotic arm and a LiDAR depth camera in real time. For visual servoing of the multi-DOF robot, we introduce a partitioned visual servoing scheme applicable to automated NP swabs. The visual servoing scheme was tested on a head phantom model using unused swabs with a variety of initial deflections, and then compared with position control that simply places the swab on the detected nostril target in 3D. Finally, swab placement at the nostril via visual servo control was demonstrated on a human subject.

II. MATERIALS AND METHODS

A. Overall System

The testbed for the proposed visual servo control comprises a custom-built serial robotic arm, LiDAR depth camera (Intel® RealSense™ LiDAR Camera L515, Intel Corp., USA), and swab adaptor to accommodate a sampling swab in the robot. The robotic arm provides six degrees of freedom (6-DOF) in motion for automated NP swab sampling. The 3-DOF motion is primarily used to position the tip of the swab in 3D, and the 2-DOF motion is used to adjust the insertion angle into the nostril—pan and tilt. The remaining DOF is employed to rotate the swab for sample collection. A trajectory from a current swab's tip poses to a target pose—the nostril location in 3D and insertion angles–is planned inline for a specified time to reach the target. Considering the trajectory, the corresponding joint angles for every control time tick are calculated using the inverse kinematics of the 6-DOF robot arm. The robotic arm is controlled by a host PC (Jetson AGX Xavier, Nvidia Corp., USA) running at 100 Hz on the robot operating system (ROS) [13]. In addition, the camera is attached to the base of the robot in an eye-to-hand configuration, as considering the minimum depth distance to reliably detect the nostril (~25 cm).

To determine the nostril target position in 3D, the vision system first detects the nostril in a 2D image, and then extracts the corresponding 3D position from a depth map acquired by the LiDAR. Prior to controlling the robotic arm, coordinate registration between the RGB camera and robotic arm (hand-eye camera calibration) is conducted using a checkboard attached to the distal end of the robotic arm. These two systems exchange data with each other via TCP/IP communication: the nostril position in 2D and 3D, current swab position, and current robot joint angles.

For visual servo control of the swab, the tip of the swab is detected, and then controlled to minimize the error between the tip and the nostril in the 2D image. Correcting the tip position via visual servoing results in the re-planning of the trajectory inline, while approaching to the nostril target. The

Figure 3. Cascade detection of nose and nostril and testing results.

Figure 4. Nostril detection for various head directions.

details of the visual servoing scheme are described in detail later. A schematic of the overall system and the control flow is shown in Fig. 2.

B. Nostril Detection

Accurate detection of the nostril is critical for the seamless operation of automated NP swab sampling, in which the robot automatically obtains the nostril target and inserts the swab into the nostril. However, it is challenging to detect a large variety of human nostrils with high precision. The shape and size of nostrils vary from person to person. Moreover, the angle and shadow of the nose affect the appearance of the nostrils in the images. Therefore, nostril detection algorithms based on geometric shapes or colors are prone to failure. We also found that the nostril detection accuracy of the Viola-Jones algorithm—a common machine learning algorithm for face detection is only 24%. Specifically, it shows substantial failure in detecting wide or snub noses.

Therefore, we introduce a deep learning-based method with a cascade scheme for accurately detecting the nostrils. First, the proposed method finds the subject' nose and sets it as a region of interest (ROI) to detect the nostrils. The nostrils are then found in the given ROI. This cascade scheme thus significantly improves the accuracy of nostril detection compared with directly searching the nostrils in an entire image. For example, it mitigates false-positive cases, such as eyes or mouth detected as nostrils.

Compared with various object detection algorithms, the YOLOv4 tiny—a lightweight version of YOLOv4 [14] is adopted for the real-time detection of the nose and nostrils with high accuracy. To learn the network, we considered a training dataset of 2150-face photos from the Flickr-Faces-HQ Dataset (FFHQ), and then used a total of
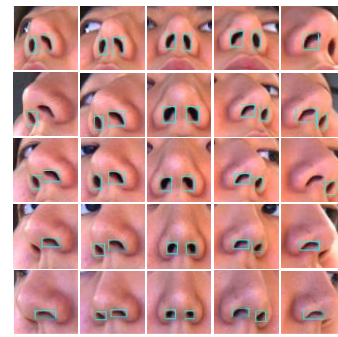
8600 photos generated by data augmentation including horizontal flipping, color transformation (saturation and exposure), random rotation, shearing, and cropping. We trained the network with an input image size of 640×640 and a batch size of 64. The results showed a detection accuracy of 99.0% with a 0.022 s-processing time on Jetson AGX Xavier with GPU acceleration in detecting the nose for a testing set of 100-face photos. A single false-positive result was accidently detected in the partial image of the the finger (with black-polished nail) stroking one's head. For the 200 nostrils of the 100 testing images, a detection accuracy of 96.0% was obtained, and the processing time for a particular ROI was 0.016 s (see Fig. 3). Eight false-negative detections were found owing to the extreme head orientation. In both tests, a detected bounding box was classifed as true-positive only if IoU (intesection over union) is greater than 0.8. Moreover, we investigated the detection performance for images with various head orientations captured using our automated swap sampling system. The nostrils were successfully detected in most of the head directions, as shown in Fig. 4.

Because the successive detection of both the nose and nostril within 30 FPS could be limited, the nose detection is only executed when necessary. For example, the nose detection is conducted once at the beginning of the automated procedure, and the detected nose offers the ROI for searching the nostrils. Once we detect the nostil as the form of a
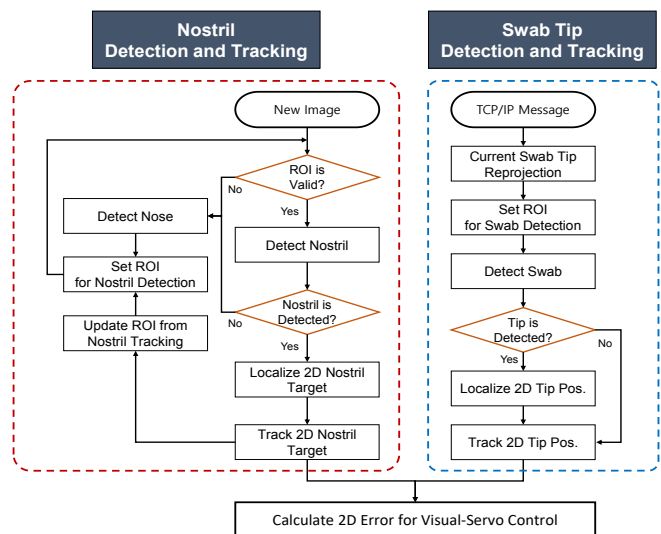

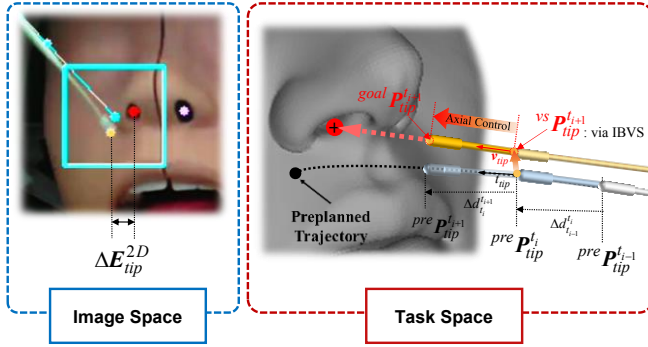Figure 5.   Detection flow of the nostril and swab tip for the visual servo control.

Figure 6.   Schematic of patitioned visual servoing for the NP swab.

bounding box given the ROI, a target point for placing the swab is specified by the moment center of the dark nostril region in the bounding box. The center location of the ROI is also kept updated via the Kalman filtering of the target point without running the nose detection; either the right or left nostril can be used selectively. If the nostril detecion fails, it repeats the nose detection until a new ROI is found. The overall procedure for the nostril detection is illustrated in Fig. 5.

C. Swab Tip Detection

For visual servoing of the sampling swab, we detect and track the tip of the swab in addition to the nostril detection. The swab is an easily bendable stick, 150 mm long, which includes a flocked head with a diameter of 3 mm at the tip. To reliably detect the swab in real time, an ROI is set by the projection of the swab tip's 3D position into the 2D image, as shown by the cyan square box in Fig. 6. The 3D position of the tip is estimated from the kinematic chain of the robot and swab, assuming that it is a straight and rigid shaft attached to the swab adaptor. Thereafter, the 3D tip position is streamed to the vision system via the TCP/IP communication. The 3D tip position is then mapped onto the image via a projection matrix given by the camera-robot calibration, which creates a 2D ROI for searching the tip in the image. Considering the 2D ROI, the swab is identified and binarized by HSV-color thresholding to obtain a bright and white swab. To eliminate spurious pixels, we use the RANSAC algorithm while assuming that the swab was a first-order polynomial with a certain width threshold. Considering a set of pixel points for the swab, the endpoint among the points projected onto the line is regarded as the tip location in the image. We apply Kalman tracking to the 2D swab tip position, as the tip detection may occasionally become unstable or undetectable.

D. Visual Servo Control

As mentioned earlier, it is challenging to accurately place the swab at the 3D nostril target within an error range of a few millimeters because of potential error sources: 1) the positioning error of the end-effector of the serial robot manipulator, 2) the detection error of the LiDAR depth camera, 3) the registration error between the camera and robot coordinates, and 4) kinematic model error owing to swab deflection. Among these error sources, our study focus on registration error and swab deflection, which are unavoidable at a certain level.

Registration errors may occur in the process of finding a spatial transformation between the robot and camera coordinates, which is described by the rotation (i.e., orientation) and translation of the camera with respect to the robot (world) frame. The translational error in the registration causes a proportional error in positioning the swab in 3D. In addition, the orientation error found in the rotation further exacerbates the positioning error of the end effector. Because the camera should be at a certain distance (herein, approximately 300 mm) from an object to detect depth using the ToF principle, even a small angular error causes a large displacement error while positioning the swab. Furthermore, the kinematic model error caused by the bent swab is more pronounced during automated swab control. An intact swab is prone to deflection (see Fig. 1). For instance, if the error of the camera's facing angle with respect to the robot base is greater than 2.0°, a positioning error of 5.2 mm at the tip may occur, where the camera's translational error and tip deflection is neglected. Likewise, a 5.2-mm targeting error at the swab's distal end can also occur even with the flawless robot-camera calibration if the swab tip is bent by 2.0° from its central axis. Either case would result in failure of inserting the swab into the nostril with a diameter of 10 mm. Fig. 6 shows the actual swab and virtual swab created by the projection of the ideal swab location (see the cyan line), while approaching the nostril; the projection of the tool tip is calculated via the robot-camera calibration assuming the ideal robot kinematics and the straight swab.

Therefore, we propose a visual servoing scheme that can compensate for the fundamental errors involved in the automated NP swab procedure. Thus, the visual servoing aims to gradually locate the swab onto the nostril using the 2D camera image until the swab is inserted into the nostril. However, a new visual servoing framework for swab control is required, unlike general visual servoing applications that rely on common feature points detected in successive image frames for controlling multi-DOF robots. Because we can obtain the error of a single feature point only (the nostril target), the 2-DOF control of the 6-DOF robot is allowed only via image-based visual servoing (IBVS). Thus, we adopt a partitioned visual servoing scheme introduced for beam control in automated laser microsurgery [16]. The 6-DOF motion of the swab tip is decoupled into a 2-DOF pan-tilt motion at the tip, 3-DOF translational motion of the robot head, and 1-DOF motion along the swab axis. The decoupled 2-DOF motion is then controlled via IBVS to locate the swab tip onto a detected nostril position using the camera.

For IBVS, we first formulate an analytical image Jacobian as introducing a 3D task plane—a virtual plane parallel to the head's frontal plane, which results in an interaction matrix for differential motions, as in (1).

$$\Delta x^{image} = J_p \Delta \Theta^{task}, \qquad (1)$$

where $\Delta x^{image} \in \mathbb{R}^{2 \times 1}$ and $\Delta \Theta^{task} \in \mathbb{R}^{2 \times 1}$ are differential motions in the image and task planes, respectively.

To derive the interaction matrix $J_p \in \mathbb{R}^{2 \times 2}$, two differential motions ΔP_u and ΔP_v are defined at the task plane containing the bases, u and v as in (2).
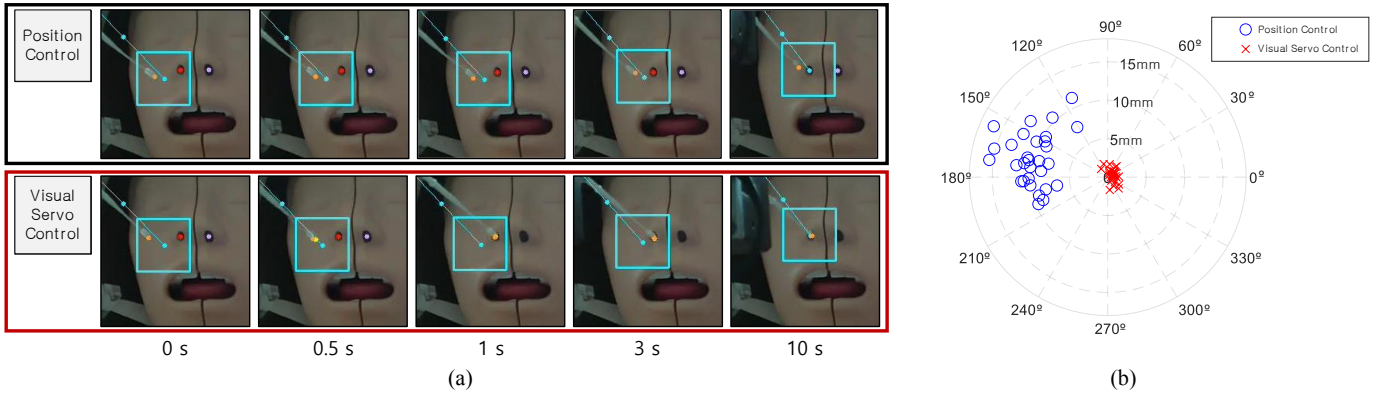
Figure 7. Comparison of the visual servo control vs. the open-loop control in the head phantom model: (a) still-cut images during the test and (b) the final locations of the testing swabs with spect to the nostril target (0, 0).

$$\Delta \boldsymbol{P}_u = \frac{{}^{task}\boldsymbol{P}_u^{target} - {}^{task}\boldsymbol{P}_u^{tip}}{\left| {}^{task}\boldsymbol{P}_u^{target} - {}^{task}\boldsymbol{P}_u^{tip} \right|} \quad \text{and} \quad \Delta \boldsymbol{P}_v = \frac{{}^{task}\boldsymbol{P}_v^{target} - {}^{task}\boldsymbol{P}_v^{tip}}{\left| {}^{task}\boldsymbol{P}_v^{target} - {}^{task}\boldsymbol{P}_v^{tip} \right|} \quad (2)$$

Their corresponding differential motions $\Delta \boldsymbol{p}_u$ and $\Delta \boldsymbol{p}_v$ in the image plane are also defined using a projection matrix, $\boldsymbol{M}_p$ provided by the camera-robot calibration:

$$\Delta \boldsymbol{p}_u = \boldsymbol{M}_p \Delta \boldsymbol{P}_u \quad \text{and} \quad \Delta \boldsymbol{p}_v = \boldsymbol{M}_p \Delta \boldsymbol{P}_v . \quad (3)$$

Because the differential motions, $\Delta \boldsymbol{P}_u$ and $\Delta \boldsymbol{P}_v$, are subject to the canonical bases of the task plane coordinates, the matrix $\boldsymbol{J}_p$ is composed by the two vectors, $\Delta \boldsymbol{p}_u$ and $\Delta \boldsymbol{p}_v$, as in (4).

$$\left[\Delta \boldsymbol{p}_u \quad \Delta \boldsymbol{p}_v \right] = \boldsymbol{J}_p \boldsymbol{I}^{2 \times 2} , \quad (4)$$

where $\boldsymbol{I}^{2 \times 2}$ denotes the identity matrix. Thereafter, we obtain the inverse of the interaction matrix, $\boldsymbol{J}_p^{-1}$, by considering the inverse of the matrix $\left[\Delta \boldsymbol{p}_u \quad \Delta \boldsymbol{p}_v \right]$ that has a full rank. Accordingly, the 3D position error between the nostril target and the current swab tip at the task plane, $\Delta \boldsymbol{E}_{tip}^{task} \in \mathbb{R}^{2 \times 1}$, can be estimated from the corresponding 2D error at the image plane, $\Delta \boldsymbol{e}_{tip}^{image} \in \mathbb{R}^{2 \times 1}$ as in (5).

$$\Delta \boldsymbol{E}_{tip}^{task} = \boldsymbol{J}_p^{-1} \Delta \boldsymbol{e}_{tip}^{image} \quad (5)$$

The joint angles to correct the error are approximated as $\Delta \boldsymbol{\theta}_{pitch,yaw}$ by assuming the pivotal motion of the swab with small angles around the swab adapter and the length of the swab specified as l_{swab}:

$$\Delta \boldsymbol{\theta}_{pitch,yaw} = 1/l_{swab} \Delta \boldsymbol{E}_{tip}^{task} , \quad (6)$$

Herein, the deviation of the swab length that may occur by installation or by its deflection is negligible.

Finally, we obtain an image Jacobian $\boldsymbol{J}$ to control the joint angles $\Delta \boldsymbol{\theta}_{pitch,yaw}$ by substituting (5) into (6) for a 2D

error $\Delta \boldsymbol{e}_{tip}^{image}$ between the nostril target and the current tip positions on the image plane as in (7).

$$\Delta \boldsymbol{\theta}_{pitch,yaw} = \boldsymbol{J}^{-1} \Delta \boldsymbol{e}_{tip}^{image} , \quad (7)$$

where $\boldsymbol{J}^{-1} = 1/l_{swab} \boldsymbol{J}_p^{-1}$. The joint angles for the pitch and yaw control are regulated by a PD controller (8), which minimizes the error between the current swab tip and nostril positions.

$$\boldsymbol{\theta}_{pitch,yaw}^{i+1} = \boldsymbol{\theta}_{pitch,yaw}^{i} + p_k \Delta \boldsymbol{\theta}_{pitch,yaw}^{i} + p_d \Delta \dot{\boldsymbol{\theta}}_{pitch,yaw}^{i} \quad (8)$$

Subsequently, we define a target position ${}^{vs}\boldsymbol{P}_{tip}^{i+1}$ regulated by visual servo control by combining the joint angles $\boldsymbol{\theta}_{pitch,yaw}^{i+1}$ with the other pre-planned joint angles via the forward kinematics of the robot arm:

$${}^{vs}\boldsymbol{P}_{tip}^{i+1} = \text{FK}(\boldsymbol{\theta}_{xyz}^{i}, \boldsymbol{\theta}_{pitch,yaw}^{i+1}, \boldsymbol{\theta}_{roll}^{i}) . \quad (9)$$

Unless the swab is perpendicular to the task plane, the advancement of the tilted swab toward the nostril can induce an error at the next image frame. Therefore, we introduce the axial control of the swab during the visual servoing. The axial control enforces the swab to move along an axis described by the nostril and the target positions, $\boldsymbol{P}_{nostril}$ and ${}^{vs}\boldsymbol{P}_{tip}^{i+1}$, respectively. Consequently, the final goal position ${}^{goal}\boldsymbol{P}_{tip}^{i+1}$ is defined by enforcing axial movement as in (10).

$${}^{goal}\boldsymbol{P}_{tip}^{i+1} = {}^{vs}\boldsymbol{P}_{tip}^{i+1} + \boldsymbol{v}_{tip}^{i+1} \cdot ({}^{pre}\boldsymbol{P}_{tip}^{i+1} - {}^{pre}\boldsymbol{P}_{tip}^{i}) \boldsymbol{v}_{tip}^{i+1} , \quad (10)$$

where $\boldsymbol{v}_{tip}^{i+1} = (\boldsymbol{P}_{nostril} - {}^{vs}\boldsymbol{P}_{tip}^{i+1}) / \left| \boldsymbol{P}_{nostril} - {}^{vs}\boldsymbol{P}_{tip}^{i+1} \right|$ and ${}^{pre}\boldsymbol{P}_{tip}^{i}$ are the pre-planned tip positions at the ith control-loop tick. The

TABLE I.   AUTOMATED SWAB PLACEMENT RESULT

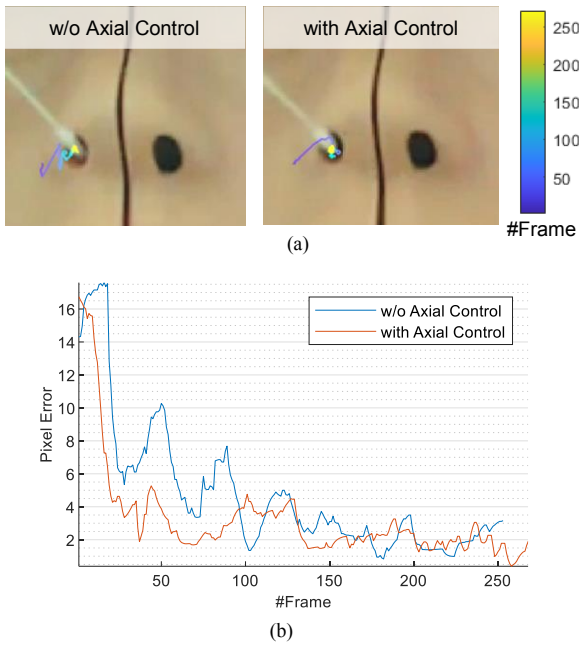| Error	Position Control	Visual Servo Control
Avg.	16.54 ± 3.53[a] px (10.58 ± 2.26 mm)	1.74 ± 0.72 px (1.11 ± 0.46 mm)
Max	25.38 px (16.24 mm)	3.19 px (2.04 mm)

a. Standard Deviation.

Figure 8. (a) Swab tip trajectory for different axial control schemes and (b) positioing error over time during the visual servoing.

Figure 9. Visual servoing demonstration on a human subject: : (a) still-cut images during the test and (b) the positioning error over time.

corresponding joint angles for $^{goal}\boldsymbol{P}_{tip}^{i+1}$ are then calculated using the inverse kinematics of the robot arm and used as control inputs. The corresponding notations and control procedures are illustrated in Fig. 6.

III. EXPERIMENTS AND RESULTS

To validate the performance of the proposed visual servoing in the automated NP swab, the swab placement at the nostril was tested in a head phantom model attached to the head fixture with a pre-determined insertion angle (approximately parallel to the palate [15]). The procedures were as follows: 1) identifying a nostril target location and 2) positioning the swab at the nostril using visual servo control. First, the 3D nostril target position was identified by deep-learning-based nostril detection in 2D and subsequent mapping from its image coordinates to the 3D LiDAR coordinates. The 3D target location in the LiDAR coordinates was then transformed to robot coordinates by camera-robot calibration and continuously updated every 33 ms. Considering the 3D target location of the nostril in robot coordinates, the 6-DOF robotic arm was controlled to accurately locate the sampling swab at the nostril via visual servoing of the swab detected in 2D. To consider initial swab deflection, the test was repeated for 30 intact NP swabs, (Noble Biosciences, Inc., Korea). The results of the visual servo control were compared with those of the position control. The position control simply follows a pre-planned trajectory, given the detection and localization of a nostril target in the 3D robot coordinates without visual feedback involved. The execution time for the procedure was set to 10 s, and the nostril and swab locations were logged using videos during the test for further analysis.

Fig. 7(a) shows the still-cut images of the videos collected during the test for the visual servo control versus the position control. Note that the actual swab shown in the camera image
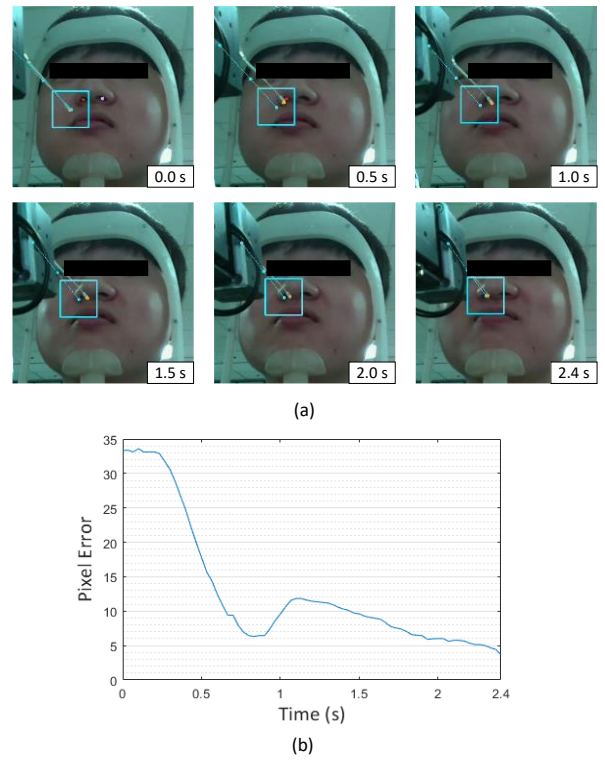
is different from the projection of its ideal kinematic model, as shown by the cyan line in the images. Nevertheless, the visual servoing allowed the swab tip to converge to the nostril target within 1.0 s in the 2D image (see the red (nostril target) and yellow (tip) markers in the images). However, for the position control, such an initial discrepancy yielded a final positioning error in the swab tip. The average positioning error was found to be 1.74 pixel (corresponding to 1.11 mm error) for the visual servo control, whereas it drastically increased to 16.54 pixel (10.58 mm) for the pure position control without visual feedback; to convert the error in pixel to mm, we adopted a scale factor 0.64 mm/pixel defined at a plane tangent to the phantom's nostril.

From a safety perspective, the maximum error can be an important metric because any failure of an operation may damage patients. The maximum error of the visual servo control was approximately 2 mm, thus all 30 trials were successful. In addition, the final swab location by the visual servo control is well clustered around the nostril target as resulting in a standard deviation of 0.46 mm. However those by the open-loop control are scattered on the left of the target, while yielding a larger standard deviation of 2.26 mm as shown in Fig. 7(b). A consistent offset from the target may be led primarily by camera-robot calibration, while such a scattered pattern can occur owing to variations in swab deflection.

We also investigated the performance of the visual servo control incorporating the axial control scheme compared to the 2-DOF angular control, as shown in Fig. 8. The error between the swab and nostril locations was tracked during the visual servoing procedure. As expected, without the axial adjustment in control, the trajectory of the 2-DOF control

fluctuated with time. However, the axial control allows for the swab tip to rapidly converge to the nostril target without significant fluctuation in its trajectory. Hence, it is found that the augmentation of the axial control can contribute to speeding up the approaching procedure as mitigating oscillation and setting higher gains for control.

Finally, the proposed visual servoing was demonstrated on a human subject using an identical protocol performed in the phantom test, except approaching speed; the execution time for the swab placement was set to 3.0 s as applying the higher control gains. Accordingly, the positioning test was conducted until the swab reached the nostril, excluding further steps for actual NP swab sampling. Unlike the immobilized phantom model, any head motion can likely induce the movement of the nostril target. Nevertheless, we obtained a similar outcome in the phantom test, as shown in Fig. 9. The swab could reach the nostril target within 2.5 s

IV. Conclusion and Future Work

We presented a visual servo control framework that can address the issues raised potentially in placing the swab at the nostril—the first and critical procedure of fully automated NP swab robots. Deep learning-based nostril detection with the cascade detection approach allows the accurate and reliable detection of nostrils in real time for a variety of shapes and sizes. The partitioned visual servo control scheme that combines image-based visual servoing with axial control can successfully compensate for errors caused by the robot-camera calibration and unavoidable swab deflection. Consequently, the proposed framework enables highly accurate swab placement at the nostril in real-time during fully automated NP swab sampling.

Future work will include the development of robust swab detection algorithms using deep learning-based methods. Potential issues that may arise in practical applications but not tackled in this study are the instantaneous head movement and the occlusion of the nostril. These will be addressed in further development by tracking the patient's head. In addition, the patient-specific insertion angle for NP swabs will also be studied. To accomplish a fully automated NP swab for human subjects, sample collection with force control, automatic vial capping mechanism, and sterilizing interfaces for the robot will also be involved.

References

[1] G. Z. Yang *et al.*, "Combating COVID-19-The role of robotics in managing public health and infectious diseases," *Sci. Robot.*, vol. 5, no. 40, pp. 1–3, 2020.

[2] A. Gao *et al.*, "Progress in robotics for combating infectious diseases," *Sci. Robot.*, vol. 6, no. 52, 2021.

[3] Y. Shen *et al.*, "Robots under COVID-19 Pandemic: A Comprehensive Survey," *IEEE Access*, vol. 9, pp. 1590–1615, 2021.

[4] "SwabBot™." https://biobotsurgical.com/swab-robot/ (accessed Sep. 14, 2021).

[5] J. Seo, S. Shim, H. Park, J. Baek, J. H. Cho, and N. H. Kim, "Development of robot-assisted untact swab sampling system for upper respiratory disease," *Appl. Sci.*, vol. 10, no. 21, pp. 1–15, 2020.

[6] S. Q. Li *et al.*, "Clinical application of an intelligent oropharyngeal swab robot: Implication for the COVID-19 pandemic," *Eur. Respir. J.*, vol. 56, no. 2, 2020.

[7] "Lifeline Robotics." https://www.lifelinerobotics.com/ (accessed Sep. 14, 2021).

[8] Y. Hu *et al.*, "Design and Control of a Highly Redundant Rigid-Flexible Coupling Robot to Assist the COVID-19 Oropharyngeal-Swab Sampling," *IEEE Robot. Autom. Lett.*, earily access.

[9] X. Wang *et al.*, "Comparison of nasopharyngeal and oropharyngeal swabs for SARS-CoV-2 detection in 353 patients received tests with both specimens simultaneously," *Int. J. Infect. Dis.*, vol. 94, pp. 107–109, 2020.

[10] "Nasal Swab Robot – Brain Navi." https://brainnavi.com/product/nasalswabrobot/ (accessed Sep. 14, 2021).

[11] V. A. Schriever, T. Hummel, J. N. Lundström, and J. Freiherr, "Size of nostril opening as a measure of intranasal volume," *Physiol. Behav.*, vol. 110–111, pp. 3–5, Feb. 2013.

[12] K. Zawawi, E. Al-Badawi, S. Lobo, M. Melis and N. Mehta, "An index for the measurement of normal maximum mouth opening", *J. Can. Dent. Assoc.*, vol. 69, no. 11, pp. 737-741, 2003.

[13] M. Quigley, K. Conley, B. Gerkey, J. Faust, T. Foote, J. Leibs, R. Wheeler, and A.Y. Ng, "ROS: An Open-source Robot Operating System," in *Proc. IEEE Int. Conf. Robot. Autom. Workshop on Open Source Robotics*, 2009.

[14] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "YOLOv4: Optimal speed and accuracy of object detection," 2020. [Online]. Available: arXiv:2004.10934.

[15] "Interim Guidelines for Collecting and Handling of Clinical Specimens for COVID-19 Testing", *Centers for Disease Control and Prevention*, 2022. [Online]. Available: https://www.cdc.gov/coronavirus/2019-ncov/lab/guidelines-clinical-specimens.html.

[16] S. Yang, R. A. MacLachlan, J. N. Martel, L. A. Lobes, and C. N. Riviere, "Comparative evaluation of handheld robot-aided intraocular laser surgery," *IEEE Trans. Robot.*, vol. 32, no. 1, pp. 246–251, 2016.