Autonomous Oropharyngeal-Swab Robot System for COVID-19 Pandemic

Fuchun Sun, *Fellow, IEEE*, Junjie Ma, *Member, IEEE*, Tianyu Liu, *Member, IEEE*, Huaping Liu, *Senior Member, IEEE*, and Bin Fang, *Member, IEEE*

***Abstract*— The outbreak of COVID-19 has led to the shortage of medical personnel and the increasing need for nucleic acid testing. Manual oropharyngeal sampling is susceptible to inconsistency caused by fatigue and close contact could also cause healthcare personnel exposure and cross infection. The innate deficiency calls for a safer and more consistent way to collect the oropharyngeal samples. Therefore a fully autonomous oropharyngeal-swab robot system is proposed in this paper. The system is installed in a negative pressure chamber and carrying out a standardized sampling process to minimize individual sampling differences. A hierarchical throat detection algorithm is presented and multiple modality sensory information are fused to safely and accurately localize the optimum sampling location. Also, a force/position hybrid control method is adopted to ensure both accurate sampling and subject comfort. The robot system described in this paper can safely and efficiently collect the oropharyngeal sample, providing a scalable solution for large-scale Polymerase Chain Reaction (PCR) Molecular sample collection for various respiratory diseases.**

***Note to Practitioners*—During the COVID-19 pandemic, prediagnostic is essential for both prevention and treatment. Existing approaches, including nasal swab and oropharyngeal-swab, require extensive medical worker training and increase the chance of cross-infection. The robot system introduced in this paper can take oropharyngeal-swab samples from subjects with minimum human intervention, reducing medical worker exposure, alleviating the work pressure of medical staff, and speed up large quantity of sampling plan. The robot will first guide the subject into position with vocal commands, and automatically detect the optimum sampling location with a real-time machine learning algorithm. A dedicated control strategy aiming at minimizing discomfort and uniforming sample quantity is then applied to safely collect nucleic samples from the throat. Eventually, while the swab is being stored in the culture medium, a disinfection process is carried out simultaneously to prepare the robot for the next subject. Preliminary clinical trials show that our robot system can safely and accurately collect samples from subjects.**

***Index Terms*— Oropharyngeal-swab robot system, sample collection, hierarchical throat detection algorithm, force/position hybrid control, self-aligning multi-modal transformer.**

I. Introduction

THE outbreak of coronavirus disease 2019 (COVID-19) has caused a worldwide pandemic which is considered the pandemic of the century [1]. In order to improve our ability to respond to the outbreak and reduce unnecessary exposure of medical staff, some companies and institutes are seeking robotic solutions to cope with a different imminent challenge within the field of diagnosis, disinfection, and care [2], [3]. Pre-diagnosis plays an essential role in both disease treatment and infection prevention. In order to quickly diagnose COVID-19 patients, oropharyngeal-swab (OP-swab) tests are widely adopted for viral nucleic acid detection [4], [5]. However, current sampling solutions have two defects. Firstly, healthcare workers performing OP-swab testing are exposed to viral aerosol from patients during the process. Secondly, different collectors have different sampling habits, which will affect the wiping position and volume of the oral epithelial cell collection, which could lead to misdiagnosis. Yang *et al.* [6] expect that automated or robot-assisted oropharyngeal swabbing helps to speed up the process, reduce the risk of infection, and free up staff for other tasks. Therefore, to minimize the risk of cross-infection, standardize the sampling process, and speed up the testing, a low-cost, fully automated robot system should be developed to collect pharyngeal samples from the throat region.

In the early stage of the pandemic, semi-automatic OP-swab robot developed by Zhong *et al.* [7] adopts teleoperation approach which requires extensive operator training and a large amount of human intervention with the loading and unloading testing materials such as swabs and nucleic acid medium. It is not suitable for large-scale deployment. Ref. [8] also developed a throat swabbing robot for COVID-19 testing. This robot uses its vision system to identify the precise points to swab in the patient's throat. Once the collection process is completed, the robot places the sample in a jar and screws on the lid. The demo shows the effectiveness of the prototype, but there is no reference to the details of the technical implementation and sampling performance.

Manuscript received 1 April 2022; revised 14 July 2022; accepted 13 September 2022. Date of publication 23 September 2022; date of current version 6 October 2023. This article was recommended for publication by Associate Editor K. Harada and Editor J. Yi upon evaluation of the reviewers' comments. This work was supported by the Major Project of the New Generation of Artificial Intelligence, China, under Grant 2018AAA0102900. *(Corresponding author: Bin Fang.)*

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Ethics Committee of Tsinghua University.

The authors are with the State Key Laboratory of Intelligent Technology and Systems, Beijing National Research Center for Information Science and Technology, Department of Computer Science and Technology, Institute for Artificial Intelligence, Tsinghua University, Beijing 100084, China (e-mail: fangbin@tsinghua.edu.cn).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TASE.2022.3207194.

Digital Object Identifier 10.1109/TASE.2022.3207194

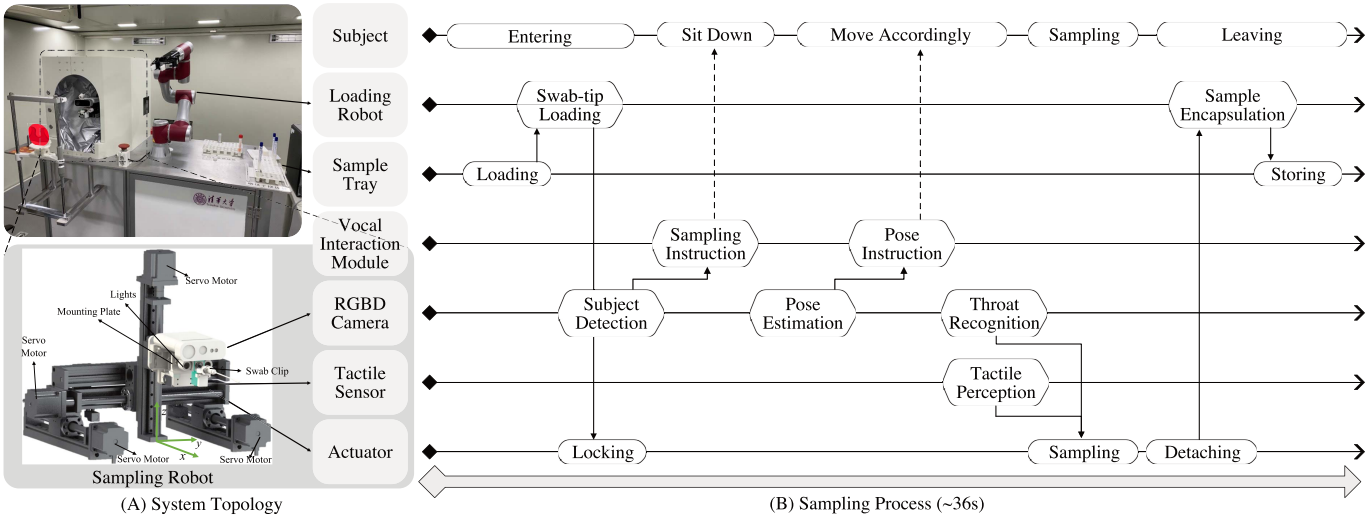


Fig. 1. (a) Design of the sample collection robot. The system consists of a loading robot and a sampling robot. The design of the loading robot is specified in the bottom figure. And each part of our robot system regarding the sampling process is demonstrated by the labels on the right. (b) Standardized OP-swab collection procedure. Different modules of our robot system are shown on the left. The arrows signify the material and information flow, while the dotted arrow states instruction from the robot to the subject.

Also, there have been researched about the novel robotic soft oropharyngeal-swab actuator to mimic the medical doctor's operational movements [9]. To our knowledge, previous robots are either teleoperated or semi-automatically operated. Most of them have shown their sampling ability and evaluated the performance of their robots. However, none has developed a fully automated standardized OP-swab sampling process to avoid medical worker exposure completely.

To design a fully automated robotic system, we formulated the entire testing into a standardized process, which is divided into four parts: swab loading, subject interaction, specimen sampling, specimen storage. Which are shown in Fig. 1-(B). During the **swab loading** phase, the loading arm will grab a swab from the tray and load it onto the sampling robot. During the **subject interaction** phase, the robot will detect the subjects' pose and issue vocal instructions accordingly; after the subject is properly seated and the sampling area is well exposed, the system will go into **specimen sampling** phase, which requires the robot to obtain an accurate fine-scale sampling location in real-time. After acquiring the sampling location, the sample collection robot will guide the OP-swab to reach the posterior pharynx wall where most respiratory viruses reside. Meanwhile, a force/position hybrid control method is introduced to ensure subjects' safety and comfort. The sampling position is recurrently refined after the actual motion of each joint of the sample collection robot. After the sampling process is complete, during the **specimen storage** phase, a loading arm will storage the specimen in a medium tray waiting for further nucleic acid testing.

To the best of our knowledge, our work introduced the first fully autonomous robotic system capable of a complete OP-swab sampling procedure without human control and intervention. On top of that, the robot can also interact with the patient, provide instruction for patients' movements, and encapsulate collected samples. Our robot also introduced several technical advancements and novel solutions:

- First autonomous OP-swab robot system to reach clinical trial.
- Formulated a standardized OP-swab collecting method.
- A pharyngeal swab sampling guidance module integrating visual and tactile modes is adopted to ensure both accurate sampling and subject comfort.
- The effectiveness and safety of the OP-swab system were proved by experiments and clinical trials.

Our work provides a novel solution for the rapid and autonomous OP-swab collecting system for most respiratory infectious diseases, including COVID-19.

The remainder of this paper is organized as follows. Section II introduces the design of the proposed autonomous OP-swab robotic system. Motivated by real OP-swab collection by medical staff, a fully automated, interactive, standardized OP-swab sampling process is also discussed for autonomous OP-swab collection. In section III, a hierarchical throat detection algorithm and a hybrid control method are proposed to ensure effectively and autonomously sampling. In section IV, a series of experiments are carried out to prove the effectiveness and safety of the system. A conclusion is claimed in section V, in which future work is also expected.

II. Robot System Design

The autonomous OP-swab robotic system mainly consists of a sample collection robot and an OP-swab loading robot, as shown in Fig 1-(A). The swab loading and specimen storage phase are completed by a 6-DOF (degree of freedom) *JAKA ZU3* robotic arm with a *DH-Robotics AG-95* gripper, as shown in the right part of the figure. Furthermore, to obtain more accurate sensor information and more stable actuator motions, a proprietary robot is designed for the specimen sampling phase, which is shown in the left part of the figure. The sample collection robot is composed of a three-axis linear platform and end scrubbing mechanism, as shown in Fig.1.

The three-axis platform is composed of four servo motors, including two servo motors in the horizontal x direction, and one servo motor in the direction of axis y and axis z, respectively. This configuration can achieve high stability motion of the platform, which will be shown in the following experiments.

The perception module, including multiple sensors, is located on top of the swab clip, moving with the three-axis linear platform. In our experiment, both Kinetics and RealSense are tested to gather RGB, Depth, and IR information for sampling area detection. In addition to that, a tactile sensor [10] is equipped between the mounting plate and the end scrubbing mechanism to ensure the safety and comfort of the subject during the sample collection process.

The end scrubbing mechanism is fixed on the linear platform to execute the delicate operations of sample collection. The motor installed at the bottom of the OP-swab clip drives the repetitive arc movement mimicking the OP-swab wiping operation of the nurse.

After specimen sampling is completed, the loading robot can remove the swab from sampling robots in accordant with the opening and closing of the actuator and eventually cut the swab after putting it into the medium.

In real-world application scenarios, our swabs are sealed in a plastic cover during transportation. The 6-DOF loading robot arm firstly grabs the OP-swab by its cover and moves it from the loading tray to the clip loading position of the OP-swab clip. After the swab is successfully locked onto the clip, the loading arm pulls the plastic cover off of the swab itself. Then, the subject will put on the mouthpiece to avoid the interference of the tongue and provide a better field of view for the camera. After the subject placed the chin on the bracket and adjusted their head posture according to the voice prompts to exposed the posterior pharynx wall, the sampling process will be triggered. During the sampling process, the end actuator driven by a three-axis platform will gradually approach the desired sampling location with the guidance of our swab point localization algorithm, which will be further introduced in Section III-A3. After the tactile sensor confirmed that the swab tip made contact with the posterior pharynx wall, the rotating motor is activated to start the swiping operation. The whole fine-scale operation is separated into three steps. The rotating motor will first rotate clockwise around the sampling location by 90 degrees as shown in Fig. 2 (b), and rotate counter-clockwise by 180 degrees as shown in Fig. 2 (c), and eventually rotate another 90 degrees clockwise to finish the repetitive arc movement, as shown in Fig. 2 (d). Eventually, the loading arm removes the OP-swab from the clip, stores the swab tip with epithelial cells in the culture media, bend and break the swab rod and discards the remaining part.

III. Methods

To effectively obtain the samples of the throat, the robot needs to confirm the correct state of the subjects and localize the sampling region, which will be introduced in detail in section III-A. Then a control strategy aiming at safely and accurately sampling of the posterior pharynx wall is proposed in section III-B.

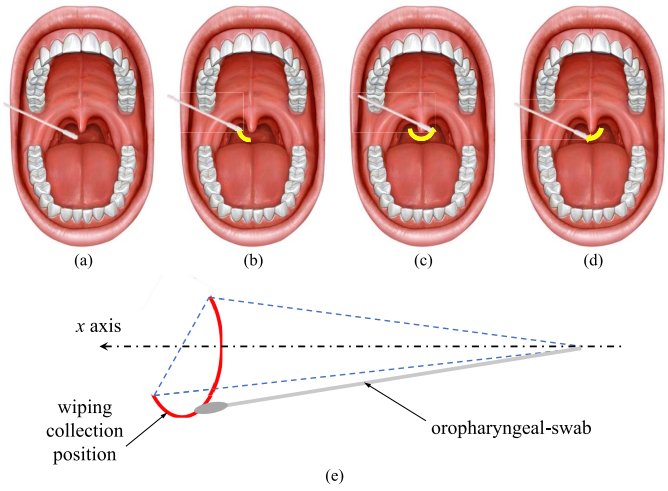


Fig. 2. Distributed wiping actions of the whole wiping operation during OP-swab collection. After the swab reached the optimal sampling position (a), the designed swabbing robot will turn clockwise 90 degrees (b), then counter-clockwise 180 degrees (c), and back to the starting point with another 90-degree clockwise movement (d).

A. Sampling Point Localization

In order to ensure real-time detection and high-accuracy localization at the same time, a hierarchical throat detection algorithm and a hybrid control method are proposed. The inherent conflict between high accuracy and high inference speed compelled us to design a new model that can both exploit high-resolution images' accuracy yet still yield a timely response to any sudden movement of the subject. Moreover, to fully exploit the potential of different sensors in varying configurations and further increase the localization accuracy, a novel self-aligning multi-modal visual transformer is proposed to cope with a severe perspective difference. It automatically attends to multi-modal information without the need for explicit manual registration of different sensors, making it suitable for mass production.

Our model works in a coarse to fine manner. The hierarchical throat detection algorithm we designed is shown in Fig.3. Firstly, a face detection module is trained on a down-sampled small image to decide the situation of the subjects, and the robot will then instruct the subject to move the direction of the head accordingly. On top of that, a landmark detection system specialized in mouth detection is trained and deployed to decide the approximate location and state of the mouth. The robot will instruct the subject to open their mouth according to the mouth state. Also, the original high-resolution image, along with depth and IR data, are clipped with the landmark detection result. By doing so, we can attain a broader spectrum of high-resolution sensory input from a related oral region. Eventually, a key point localization method based on the visual transformer is constructed to emit the fine-scale sampling location based on multi-modal sensory input. All the models are trained/fine-tuned on intra-oral images annotated by professional medical staff.

1) Face Detection: For face detection, high-resolution RGBD images are collected with the camera deployed on the

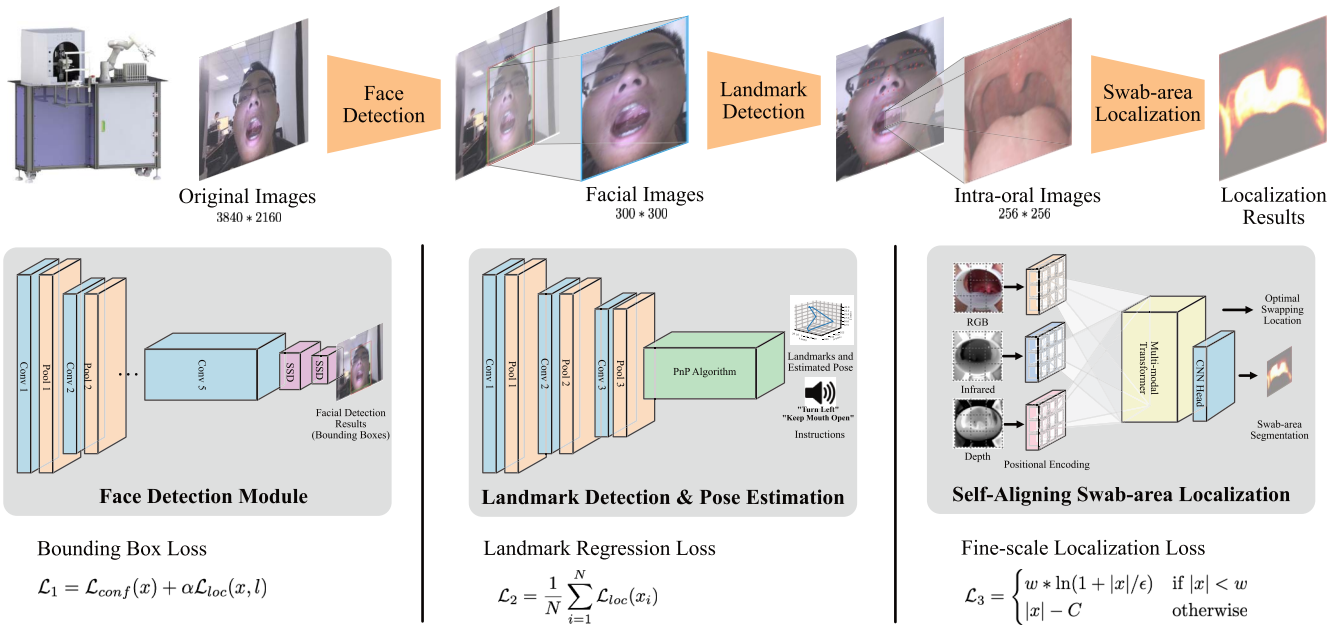


Fig. 3. Overall architecture of the hierarchical throat detection algorithm consists of three parts linked with clipping and Region of Interest (RoI) pooling. The first part is to detect whether a face exists and the subject's status with a high-res image from the RGBD camera. The second part emits the landmarks of the subject from faces in cropped images. And the third part generates viable sampling locations from throat images.

OP-swab robot. In order to extract the landmark, we firstly trained a Single Shot-MultiBox Detector (SSD) [11] to assert the existence of a human face and establish the coarse-grained location of the face. In order to improve the detection accuracy, we built the clinical facial dataset with 2261 images collected by the RGBD camera during manual-controlled collection processes in multiple possible scenarios. The face detector is first trained on publicly available datasets and then fine-tuned to fit our experimental setup better.

The loss function for training the face detector, as shown in equation (1), consists of two parts. The first is the confidence of contents inside the bounding box being a face denoted by $\mathcal{L}_{conf}(x)$. The second part is location regression loss of generated bounding box, denoted by $\alpha\mathcal{L}_{loc}(x,l)$.

$$\mathcal{L}_1 = \mathcal{L}_{conf}(x) + \alpha\mathcal{L}_{loc}(x,l) \tag{1}$$

2) Facial Landmark Detection: After the robot acquired the location of the face with a certain confidence, images are clipped accordingly. And a landmark detection algorithm, based on the regression tree method, is applied to the clipped image to extract a total number of 68 Facial Action Coding Units (FACUs) of the face in the two-dimensional image. The landmark locations of the eyebrow, eye, nose, face contour, and mouth contour, established with PnP algorithm [12], are then converted back to three-dimensional coordinates to decide the status and orientation of subjects, and the convex hull connected by FACUs represents the positioned mouth region.

The loss function of the landmark detection approach is represented by equation (2), which is a distance function averaged over all detected landmarks. The Euclidean distance from x to the nearest landmark in the label set is used as $\mathcal{L}_{loc}(x)$. M is the total number of FACUs on a certain face, which values 68.

$$\mathcal{L}_2 = \frac{1}{M}\sum_{1}^{M}\mathcal{L}_{loc}(x) \tag{2}$$

Here, a ResNet-10 structure is used as the backbone of the Single Shot-MultiBox Detector (SSD) network, and only the coordinates of FACUs of the faces are regressed. The designed mouth landmark detection algorithm was tested on the video data taken by the OP-swab robot.

3) Fine Scale Multi-Modal Swabbing Point Localization: In the OP-swab point localization module, to align different modalities and accurately localize feasible swab points, a novel self-aligning visual transformer is designed. The model design closely follows the original Visual Transformer (ViT), allowing us to implement the model and carry out different comparison experiments efficiently.

In order to deal with relatively small localization errors, wing loss is used during the training process, as shown in equation 3 [13].

$$\mathcal{L}_3 = \begin{cases} w \times \ln(1+|x|/\epsilon) & \text{if } |x| < w \\ |x| - C & \text{otherwise} \end{cases} \tag{3}$$

The final model takes images from the RGBD camera with a resolution of 3840*2160. The image is then clipped according to landmark detection and then resized to 256*256. And with five layers of different feature scales, the overall model contains 3.1M trainable parameters trained with TensorFlow on a single GeForce RTX 2080. And the average online inference time is below 18ms (99.5%) with an average accuracy (Mean squared error) of 0.3726, which is enough to instruct real-time robot movement.

B. Control Strategy

The OP-swab robot system is a semi-invasive human device. The safety of the subject during the sampling process of the OP-swab procedure should be fully considered. It is necessary to design a control strategy to realize the sampling of OP-swabs through effective contact, which can ensure accurate movement and reduce nausea and vomiting reflex during the collection process. Hence, besides throat recognition by visual, force detection is necessary. A tactile sensor is installed at the end of the swab clip. The tactile force feedback mechanism is used to prevent the subject from being hurt, reduce the discomfort, and achieve safe and autonomous collection while ensuring the effectiveness of samples.

The tactile sensor periodically returns the detected force data to the robot system. The medical staff was asked about their OP-swab collection experience. When the sampling force reaches at least 0.1 Newton and no more than 1 Newton, the collection process can be completed well, and the safety of collection can be guaranteed.

In order to eliminate the interference of perturbation noise of tactile sensor, a force control scheme named the Arithmetic Mean Filter algorithm is designed to filter the force data. The collected data waveform becomes smoother through the filtering operation, and the collection process would not be stopped due to sudden abnormal data. The essence of the Arithmetic Mean Filter algorithm is another expression of a low-pass filtering algorithm in the time domain, which has the advantages of low calculation amount and resistance to random interference signals. N samples are used continuously for arithmetic mean operation. When the N value is larger, the signal smoothness is higher, but the sensitivity is reduced. When the N value is small, the signal smoothness is low, but the sensitivity is high. Choose N as ten, and the Arithmetic Mean Filter algorithm is formulated as:

$$\bar{f}_j = \frac{\sum_{i=1}^{N} f_i}{N}, \tag{4}$$

where $\{f_i\}|_{i=1}^{N}$ is the initial force sequence data, and $\{\underbrace{\bar{f}_j, \bar{f}_j, \cdots, \bar{f}_j}_{10}\}|_{j=1}^{floor(N/10)}$ is the filtered tactile force sequence data after Arithmetic Mean Filter operation.

Then a force/position hybrid control method is proposed as shown in Fig.4. Fig.4 (a) represents the dynamic interaction between the subject's heads and the autonomous OP-swab robot system. The coordinate value of OP-swab sampling position in the camera coordinate system P^{vis} is observed by the previously mentioned throat recognition algorithm. And the filtered actual tactile force on the internal tissues of the mouth F^{real} is also perceived by the tactile sensor. The OP-swab sampling position P^{vis} and tactile force F^{real} constitute states space or observation space that describe the subject's environment. The actuator in the sample collection robot performs an action $\hat{q}$ on each linear joint of three axes on the sample collection robot according to the force/position hybrid control approach. Fig.4 (b) reveals the specific force/position hybrid control approach of the actuator in sample collection robot for OP-swab collection. Firstly, the original coordinate difference

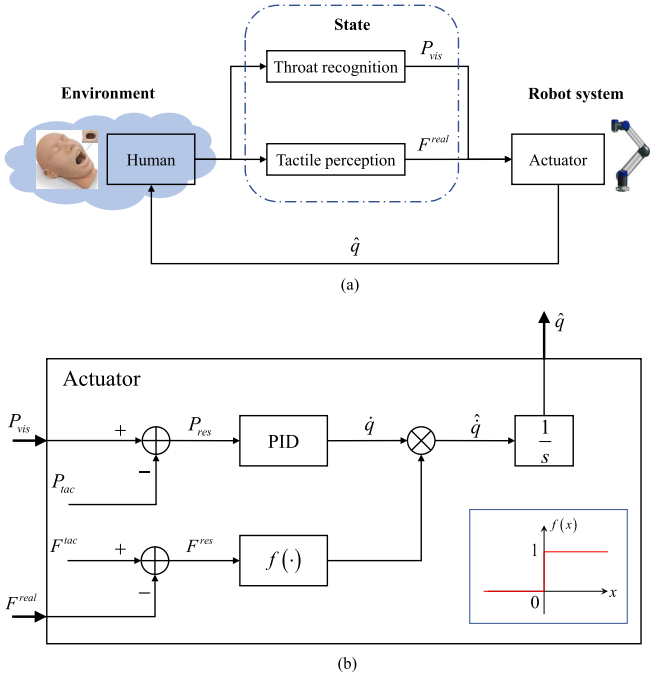


Fig. 4. Force/position hybrid control flowchart. The structure of the force/position hybrid control method is illustrated in (a). It represents the dynamic interaction between the subject's heads and the autonomous OP-swab robot system. (b) reveals the specific hybrid control approach of sample collection robot for OP-swab collection. P^{vis} is the coordinate value of OP-swab sampling position in the camera coordinate system. P_{tac} is the positions of the sampling head of the OP-swab. $\dot{q}$ is the velocity of the joint optimized by a PID controller. $\hat{\dot{q}}$ is the gated velocity of joint by a unit step function $f(\cdot)$. $\hat{q}$ is the actual movement distance of the joint. F^{tac} is the tactile force threshold value. F^{real} is the actual tactile force filtered by the Arithmetic Mean Filter algorithm.

P^{res} between sampling position P^{vis} and the positions of the sampling head of the OP-swab P_{tac} is calculated. Then, the velocity of joint $\dot{q}$ is optimized by the PID controller. Besides, $\dot{q}$ is gated by a unit step function $f(\cdot)$ to ensure safety during the OP-swab collection process. Whence F^{real} is larger than the threshold value F^{tac}, then the actuator is blocked. The sampling position P^{vis} is updated after the actual action $\hat{q}$ of each linear joint of three axes on the sample collection robot.

IV. Experiments and Discussion

To fully evaluate the effectiveness and efficiency of our robot system, several experiments are conducted before and during clinical trials. There are two modules in our sampling pipeline that could cause wrongful sampling location, the localization accuracy is evaluated in Section IV-A and robot locomotion accuracy is evaluated in Section IV-B to demonstrate the overall accuracy of our OP-swab robot system. The clinical setup is introduced in Section IV-C. To ensure sufficient swab contact and subject comfort, numerous force sequences are collected during the clinical trial and analyzed in Section IV-C1. In Section IV-C2, the samples are then sent to a third-party biological lab to evaluate the sampling efficiency. The sampling speed during clinical trials is also analyzed in Section IV-C3.

TABLE I
RESULT ABOUT AVERAGE DISPLACEMENT ERROR OF OP-SWAB POINT GENERATED FROM DIFFERENT MODALITIES AND TRAINED WITH DIFFERENT LOSSES

Modality(s)	L2 Loss	L1 Loss	Wing Loss
V (RGB)	0.7264	0.5615	0.3890
I (Infrared)	1.1323	1.0296	0.9802
D (Depth)	1.0763	0.9710	0.9049
Multi-modal	0.7134	0.5224	**0.3726**

A. Sampling Point Localization Experiment

According to the OP-swab sampling guideline, the optimal sampling region is an approximately 160 mm^2 area located at the back wall of the pharynx. To quantify the accuracy of the localization algorithm, we carried out a series of experiments. According to the medical characteristics of the human body, the average width of the uvula of adults is about 10 mm, and the height is no more than 20 mm. The abstracted sampling location is shown in Fig. 6, which is marked in green. The spatial relationship between the sampling point calculated by the throat recognition algorithm and the real-scale coordinates of the collected location in the camera coordinate system is analyzed, and the sampling positioning results are annotated as follows:

$$S_l = (x_L, y_L), S_r = (x_R, y_R), S_c = (x_C, y_C). \tag{5}$$

As shown in Fig. 6, where P_l and P_r represent the left and right FACUs in the mouth area, and S_l and S_r is the localization result generated by our model, and S_c is the location of the uvula, which are only used for training purpose. By calculating the distance from two sub centred points (O_l, O_r) of abstracted rectangle region $(Area_l \bigcup Area_r)$ to the sampling points S_l and S_r, we can quantitatively evaluate the accuracy (acc) of our proposed robot by mean sampling point deviation:

$$acc = \min(\|S_l, O_l\|, \|S_r, O_r\|). \tag{6}$$

After evaluating 400 samples, the mean sampling point deviation is 3.72mm in the real-world plane, which is more than enough for valid sample collection. Also, as shown in Table I, various losses are used during training, and ablation studies are carried out. The experiment result shows that, although both infrared and depth image contains limited information, adding them to the mix improved the accuracy from 3.890 to 3.726 (+0.164, 4.2%), which indicates the Self-Aligning Multi-Modal Transformer can exploit the strength of different modality and dynamically attend to the significant part relating to the task for better localization accuracy. Also, we found that the Wing Loss is significantly better for fine-scale localization tasks than traditional L1 and L2 losses. The distribution of deviation is plotted as Fig. 5.

The localization experiment is implemented. The facial action unit detection approach is implemented by using the regression tree method in the DLIB library. The OpenCV version is 4.3.0, and the DLIB library version is 19.20 for the configuration of the landmark detection algorithm. The regression tree is finetuned with our collected clinical facial dataset based on a pre-trained model which was well trained on LFPW, HELEN, AFW, and IBUG datasets [14], [15], [16]. Our collected clinical facial dataset is captured by the visual module of the OP-swab robot system by gathering a total number of 2261 frames, including ten individual subjects whose ages range from 21 to 47, with hand-made bounding boxes and facial landmarks, considering possible situations which could happen during the swab collection procedure. The correlated depth information and infrared image are also recorded. This particular setup results in an inference speed on a single NVIDIA 1080 of around 19ms and average accuracy of 3.72mm.

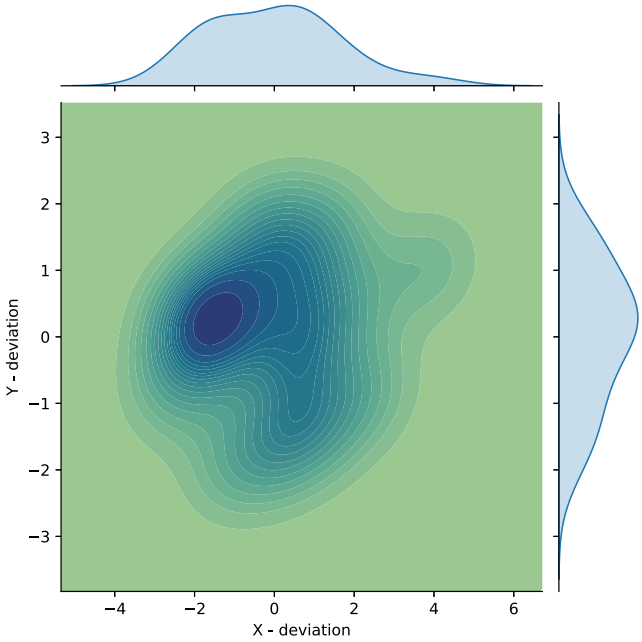


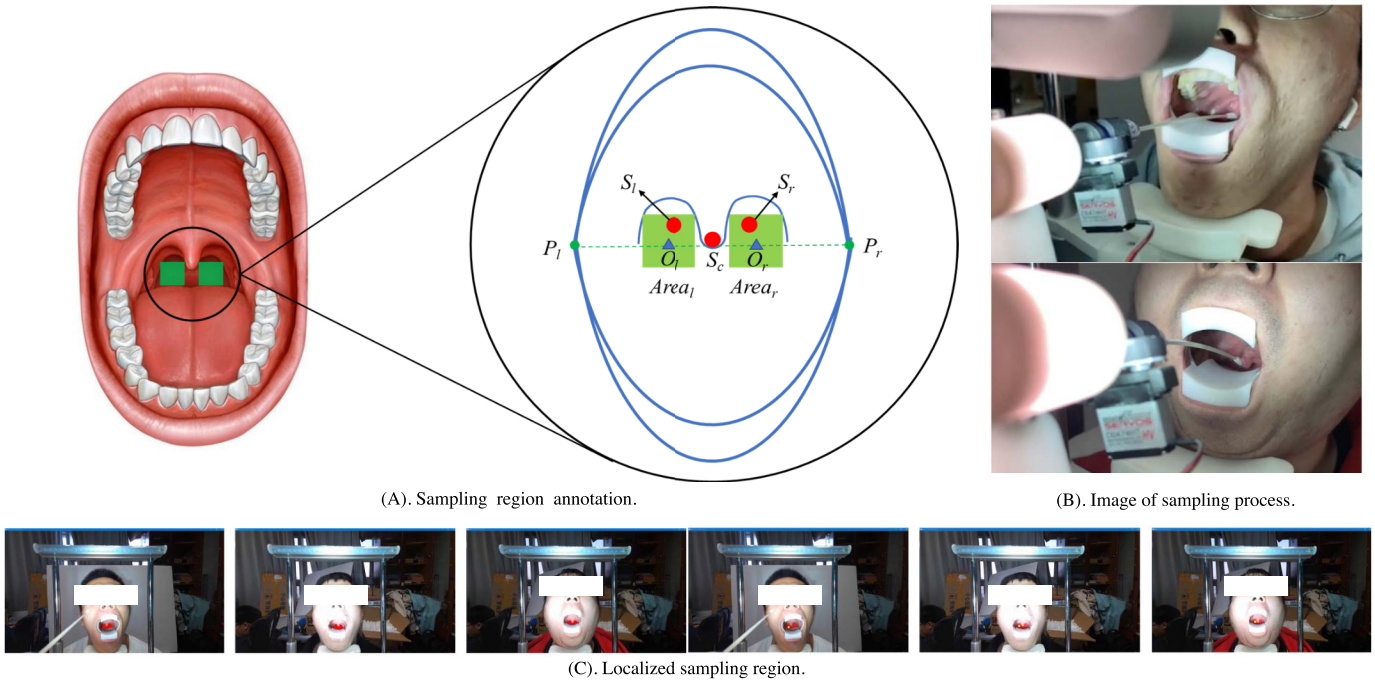
Fig. 5. Kernel density estimation (KDE) of sample point deviations (mm).

B. Robot Locomotion Accuracy

Other than localization error, locomotion deviation is another factor that could cause wrongful sampling. To evaluate the robot locomotion accuracy, experiments are designed where the 3-axis platform conducts a series of actions, and the desired endpoint location of the swab tip is compared with the actual location. The series of actions used here is in similarity with the real-world sampling sequence, and the desired end location is spread through 9 steps in both X-axis and Y-axis (−40.0 mm, −30.0 mm, −20.0 mm, −10.0 mm, 0.0 mm, 10.0 mm, 20.0 mm, 30.0 mm, 40.0 mm). Each target location is repeated ten times, and the actual locations are recorded. After 170 experiments, the maximum recorded deviation is 0.2mm. Taken the 3.72mm localization error into account, the overall system accuracy is still more than enough for a successful OP-swab sampling.

C. Clinical Trials

Our robot system, to our knowledge, is the first autonomous OP-swab sampling robot system to reach clinical trial stages.



(A). Sampling region annotation.

(B). Image of sampling process.

(C). Localized sampling region.

Fig. 6. (a) Sampling region abstraction Sampling Region is abstracted into a rectangle for easier evaluation (areas emphasized in green). The spatial relationship between the pharyngeal swab collection points (S_l) and (S_r), uvula point (S_c), and abstracted rectangle region ($Area_l \bigcup Area_r$) in the pharyngeal area is zoomed in in the middle circle. (b) Image of the real-world sampling process.**(C). Localization result of the real-world sampling process.** The sampling region is shown in red, and the optimal sampling location is shown with a yellow dot. Subject images are redacted due to privacy reasons.

Our clinical trial setup is shown in Fig.1, where the robots are placed in a negative pressure chamber to prevent medical worker infection. The air circulation system is constructed so that the air is filtered and disinfected before being released outside. Also, to prevent cross-infection between patients, ultraviolet disinfection lamps and disinfectant spray are automatically deployed between each sampling process.

In this part, the introduced OP-swab robot system was tested in a clinical environment. We evaluated the subject's safety and comfort during the procedure, the effectiveness of collected samples, and the sampling efficiency of this robot system.

1) Force Sequence: Subject safety is our primary concern when designing this robot system. The tactile sensor installed at the end of the actuator can ensure the subject's safety during the procedure. The data collected during the trial can also be used to analyze subject comfort during the sampling process.

Hence, in this tactile force experiment, fourteen volunteers (six males and eight females) were sampled one time. The relation curve between the tactile force value and the pressure sensor voltage is also calibrated. According to the instrument of the pressure sensor, the relationship between tactile force value and pressure voltage is approximately linear. Therefore, the function is regressed based on 20 groups of measurement:

$$F = 1.7228U. \tag{7}$$

Here, F and U represent tactile forces and pressure voltage. The unit is millinewton and millivolt, respectively. The square of correlation coefficient R^2 is 0.9994.

According to previous research, a successful OP-swab sampling requires a sampling force of at least 0.1 Newton; otherwise, the swab will not collect enough oral epithelial cells for downstream nucleic acid amplification testing. Any force above 0.6 Newton might cause discomfort and force over 1 Newton risk causing physical damage to subjects.

A total of 14 recorded sequences are recorded in Fig. 7, where we can see 12 of which are well within comfort thresh, and all 14 sequences are within a safe range. The tactile sequences also coincide with our control scheme. The OP-swab first approaches the sampling location guided by our localization algorithm, as shown in Step 1. After the swab tip made contact with the pharynx's posterior wall, the tactile force surpasses the preset threshold. Three stages of end swiping are started, and the swab moves in an arc manner in simulation with human sampling. Three peaks in the force sequence correspond to the three wiping movements described previously in Fig. 2. Eventually, the sample collection robot withdraws the OP-swab from a subject's mouth to complete the collection process. The three stages in Step 2 can also be found in other sampling sequences as shown in Fig. 7.

Also, we can observe the enlarged part in Fig. 7, where the system rapidly backs up when the set threshold of 1N is exceeded. The response time of our system is around 20ms, in collaboration with the tactile sensor, our system can quickly respond and make sure that the collection force stays within a safe range.

2) Sampling Effectiveness: The collected sample is packed and handed over to the testing facility. After each sampling process was completed, the swab sample was well packaged and handed over to the testing facility. The effectiveness of the collected samples can be verified by the threshold cycle (C_t)

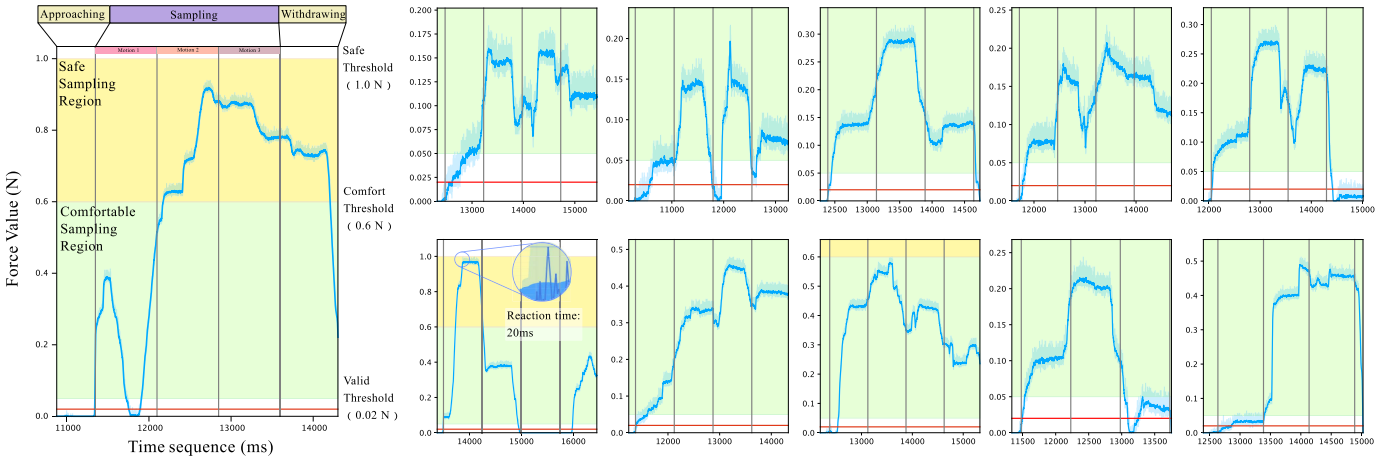


Fig. 7. Tactile force sequence of OP-Swab sampling process. The sequences are shown in blue curves with a moving average smoothing over 50 ms. The horizontal red line in each graph is the activation thresh of the tactile sensor, and the multiple vertical grey lines separate different stages of end actuator swiping motion. The green areas are comfort and successful sampling regions, and the orange areas are the safe sampling regions. The majority of force sequences are within the range of 10-60g proposed by [7], which are sufficient for nucleic acid sampling and not too hard to cause discomfort.

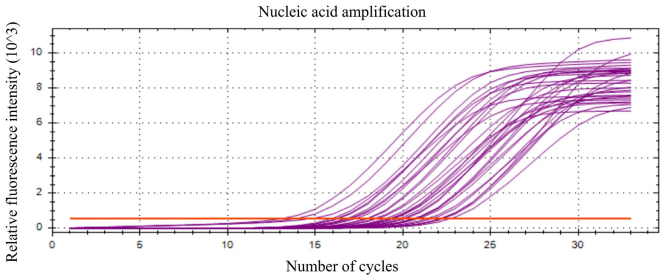


Fig. 8. RT-PCR test results of the OP-swab samples collected by our robot system.

TABLE II

CT DISTRIBUTION OF COLLECTED OP-SWAB SAMPLES BY THE PROPOSED ROBOT SYSTEM

Ct value interval	[13,16)	[16,19)	[19,22)	[22,25)	[25,+∞)
No. of samples	1	5	5	3	0

value of the selected reference gene (RNase P) by RT-PCR tests [17]. 14 of the experiment result are shown in Fig. 8 and results are summarized in Table II. Typically OP-swab samples with Ct value below 28 are considered qualified and suitable for downstream nucleic acid testing. After established normality of Ct value distribution with Shapiro-Wilk testing due to limited sample size, a Gaussian kernel density estimation is carried out showing that our sample has a $C_t < 25.12$ ($p = 0.99$), which proved the effectiveness of samples collected by our OP-swab robot system.

3) Sampling Efficiency: After examining the effectiveness of collecting samples with our OP-swab robot, we carried out experiments to evaluate the swab sampling efficiency. The entire sampling process can be divided into six steps:

i) Place an OP-swab and a culture tube on the designated holder on the platform of the OP-swab robot.
ii) Load the OP-swab onto the swab clip by the robotic arm with a gripper.
iii) Ask the volunteer to adjust the head posture according to the voice prompts by the throat recognition method to prepare for the swab sampling process.
iv) Collect the OP-swab sample.
v) Unload the OP-swab sample from the platform, and put the sample in the culture tube.
vi) Label the culture tube and place it into the designated storage tray.

The total time cost for the subjects from entering the negative pressure chamber to putting the culture tube into the holder for sample storage is measured in Table. IV, the time of Step iv) is shown in Table. III.

TABLE III

TIME CONSUMPTION OF COLLECTED OP-SWAB SAMPLES BY THE PROPOSED ROBOT SYSTEM

Subject	collecting time (seconds)	Subject	collecting time (seconds)
Subject A	20.580	Subject H	16.502
Subject B	15.378	Subject I	14.744
Subject C	15.061	Subject J	13.754
Subject D	15.479	Subject K	15.395
Subject E	14.351	Subject L	15.639
Subject F	14.380	Subject M	13.298
Subject G	14.482	Subject N	13.155

TABLE IV

TIME OF FULL OP-SWAB PROCEDURE

	Test 1	Test 2	Test 3	**Average**
Subject A	2min45s	2min45s	2min43s	2min44.3s
Subject B	2min51s	2min45s	2min45s	2min47.0s
Subject C	2min44s	2min44s	2min45s	2min44.6s
Subject D	2min48s	2min47s	2min45s	2min46.7s
Subject E	2min50s	2min46s	2min46s	2min47.3s
Subject F	2min48s	2min47s	2min47s	2min47.3s
Subject G	2min43s	2min45s	2min43s	2min43.6s
Overall				**2min45.8s**

Through the fully autonomous collection experiment of OP-swabs on seven subjects, each of them was tested three times for a total of 21 experiments, and the experimental process was completed under the guidance of professional physicians. The sampling process of each subject is recorded, and the average total sampling time of seven subjects is analyzed. The experimental result is shown in Table IV. The average collecting time per person is about 166 seconds. Therefore, our autonomous OP-swab robot has a sampling capacity of 20 subjects per hour. From Table III we can see that collection only takes a small portion of the collection procedure; hence the speed can be further improved by stacking the preparing and after-processing procedure.

V. Conclusion

To our knowledge, our OP-swab robot system is the first in the world to reach the clinical trial stage. Unlike previous studies on the OP-swab robot system that used the teleoperation method for swab sampling, which requires extensive professional training, the OP swab robot system we developed can autonomously complete the OP swab sampling process with minimum intervention. Which significantly reduces the workload of medical staff and enables large-scale parallel sampling with limited managing personnel.

In order to improve sampling accuracy and improve subject safety, a hierarchical throat detection algorithm is presented. It can automatically attend to the relevant part of various modalities and provide accurate real-time location of the posterior pharyngeal wall. The extensive experiment also showed improvement in performance brought up by our novel method. Also, the force/position hybrid control method is introduced to improve the comfort of the process and realize safe and autonomous collection while ensuring sampling effectiveness. Furthermore, the standardized OP-swab sampling process we proposed can also be regarded as a reference for future research on autonomous OP-swab collection systems and provide helpful insight into the problem of human-machine integration in the sampling process.

Extensive experiments are carried out around the robot's safety, the effectiveness of collected samples, and the efficiency of the procedure. Our experiment also discovered that the loading/unloading process takes up most of the time during the whole procedure, which significantly reduces the overall sampling efficiency. A new swab loading/unloading device dedicated to loading OP-swab will be designed to improve the overall efficiency in future work. By simultaneously loading new swabs and storing collected samples, the efficiency could be further improved. Overall, the results demonstrate that the robot can safely and efficiently collect OP-swab samples without much need for human intervention, providing a new tool for cheap and accurate OP-swab sampling procedures.

Acknowledgment

The authors would like to thank Xiangdong Mou and Li Li from the Department of Respiratory, Beijing Tsinghua Changgung Hospital; Runqing Li from the Department of Laboratory Medicine, Beijing Tsinghua Changgung Hospital; and Na Guo and Wenjun Zhang from the Department of Computer Science and Technology, Tsinghua University. They made contributions to the clinical trials of the autonomous OP-swab robot system and the testing of sample validity.

References

[1] C. Sohrabi *et al.*, "World health organization declares global emergency: A review of the 2019 novel coronavirus (COVID-19)," *Int. J. Surg.*, vol. 76, pp. 71–76, Apr. 2020.

[2] Y. Shen *et al.*, "Robots under COVID-19 pandemic: A comprehensive survey," *IEEE Access*, vol. 9, pp. 1590–1615, 2021.

[3] P. Vaishnavi *et al.*, "Artificial intelligence and drones to combat COVID-19," Tech. Rep., V 1, 2020, doi: 10.20944/preprints202006.0027.

[4] N. Chen *et al.*, "Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: A descriptive study," *Lancet*, vol. 395, no. 10223, pp. 507–513, May 2020.

[5] C. Ying-Long, S. Fu-Jun, and G. Yong-Jun, "Remote human–robot collaborative impedance control strategy of pharyngeal swab sampling robot," in *Proc. 5th Int. Conf. Autom., Control Robot. Eng. (CACRE)*, Sep. 2020, pp. 341–345.

[6] 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, 2020, Art. no. eabb5589. [Online]. Available: https://robotics.sciencemag.org/content/5/40/eabb5589

[7] S.-Q. Li *et al.*, "Clinical application of an intelligent oropharyngeal swab robot: Implication for the COVID-19 pandemic," *Eur. Respiratory J.*, vol. 56, no. 2, Aug. 2020, Art. no. 2001912.

[8] Accessed: 2020. [Online]. Available: https://www.therobotreport.com/danish-startup-develops-throat-swabbing-robot-for-covid-19-testing/

[9] X. Zhexin *et al.*, "A tapered soft robotic oropharyngeal swab for throat testing: A new way to collect sputa samples," *IEEE Robot. Automat. Mag.*, vol. 28, no. 1, pp. 90–100, Mar. 2021.

[10] F. Bin, S. Fuchun, C. Yang, Z. Chang, X. Ziwei, and Y. Yiyong, "A tendon-driven dexterous hand design with tactile sensor array for grasping and manipulation," in *Proc. IEEE Int. Conf. Robot. Biomimetics (ROBIO)*, Dec. 2019.

[11] W. Liu *et al.*, "SSD: Single shot multibox detector," in *Proc. Eur. Conf. Comput. Vis.* Amsterdam, The Netherlands: Springer, 2016, pp. 21–37.

[12] M. A. Fischler and R. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 24, no. 6, pp. 381–395, 1981.

[13] Z.-H. Feng, J. Kittler, M. Awais, P. Huber, and X.-J. Wu, "Wing loss for robust facial landmark localisation with convolutional neural networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 2235–2245.

[14] C. Sagonas, G. Tzimiropoulos, S. Zafeiriou, and M. Pantic, "300 faces in-the-wild challenge: The first facial landmark localization challenge," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops*, Dec. 2013, pp. 397–403.

[15] C. Sagonas, E. Antonakos, G. Tzimiropoulos, S. Zafeiriou, and M. Pantic, "300 faces in-the-wild challenge: Database and results," *Image Vis. Comput.*, vol. 47, pp. 3–18, Mar. 2016.

[16] C. Sagonas, G. Tzimiropoulos, S. Zafeiriou, and M. Pantic, "A semi-automatic methodology for facial landmark annotation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2013, pp. 896–903.

[17] A. Radonić, S. Thulke, I. M. Mackay, O. Landt, W. Siegert, and A. Nitsche, "Guideline to reference gene selection for quantitative real-time PCR," *Biochem. Biophys. Res. Commun.*, vol. 313, no. 4, pp. 856–862, Jan. 2004.

Fuchun Sun (Fellow, IEEE) is currently a Full Professor with the Department of Computer Science and Technology, Tsinghua University, Beijing, China. His current research interests include robotic perception and cognition. He was a recipient of the National Science Fund for Distinguished Young Scholars. He serves as an Associate Editor for a series of international journals, including the IEEE Transactions on Fuzzy Systems, *Mechatronics*, and *Robotics and Autonomous Systems*.

Junjie Ma (Member, IEEE) received the B.Eng. degree from the Taiyuan University of Technology, Taiyuan, China, in 2012, and the Ph.D. degree from the Beijing Institute of Technology, Beijing, China, in 2020. He was founded by the China Scholarship Council and as an Internship Student at Nanyang Technological University, Singapore, from 2017 to 2019. He is currently an Assistant Researcher with the Department of Computer Science and Technology, Tsinghua University, Beijing. His research interests include computer vision and deep learning.

Tianyu Liu (Member, IEEE) received the B.Eng. degree (Hons.) in computer science from Beijing Forestry University. He is currently pursuing the M.S. degree in computer science with Tsinghua University. His main fields of works include neural network architecture, multi-modal perception and fusion, and computer vision algorithm in the autonomous driving scenario.

Huaping Liu (Senior Member, IEEE) is currently an Associate Professor with the Department of Computer Science and Technology, Tsinghua University, Beijing, China. His current research interests include robotic perception and learning. He serves as an Associate Editor for some journals, including the *Cognitive Computation*, *Neurocomputing*, the IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS, and IEEE TRANSACTIONS ON AUTOMATION SCIENCE AND ENGINEERING, and some conferences, including ICRA and IROS.

Bin Fang (Member, IEEE) received the Ph.D. degree in mechanical engineering from Beihang University in 2014. He is currently a Research Assistant with the Department of Computer Science and Technology, Tsinghua University. His research interests include sensor fusion, wearable device, robotics, and human–robot interaction.