System for Teaching Proper Toothbrushing Techniques using 6DOF Marker Pose Estimation and Machine Learning Methods

Dmytro Fedasyuk[1], Ostap Truba[1] and Tetyana Marusenkova[1]

[1] *Lviv Polytechnic National University, St. Bandery str, 28 a, Lviv, 79013, Ukraine*

Abstract

A novel approach to synthesizing software systems for teaching toothbrushing techniques is proposed. This approach leverages augmented reality technology and machine learning methods to monitor toothbrush movements, recognize cleaning gestures, and analyze and evaluate the user's performance. This paper outlines the proposed approach, determines the optimal marker type, size, and brightness conditions required for precise positioning assessment, and highlights Kalman filtering for suppressing noise introduced by camera imperfections and swift toothbrush movements.

Keywords

Toothbrushing, pose estimation, fiducial markers, augmented reality, convolutional neural networks

1. Introduction

Diseases of the oral cavity represent one of the most significant health challenges for countries and populations worldwide. According to estimates by the World Health Organization (WHO) in 2019, oral cavity diseases affected nearly 3.5 billion people [1].

Among the primary oral cavity diseases are dental caries, gingivitis, oral cavity cancer, HIV infection, cleft lip and palate, and oral cavity and dental traumas. Additionally, research exists showing a correlation between deteriorating oral health and overall health conditions, which may be associated with diseases such as heart disease, endocarditis, and premature births [2].

Inadequate or improper oral hygiene ranks as one of the leading causes of the high prevalence and intensity of dental diseases. While addressing insufficient oral hygiene is relatively widespread and straightforward, identifying and rectifying improper hygiene practices can be challenging. Factors contributing to the development of incorrect oral hygiene habits include low levels of education, habit formation during childhood, and the confusing abundance and variety of tooth-cleaning recommendations [3, 4].

Therefore, the development of systems to educate individuals on proper oral hygiene practices, providing tools to cultivate correct habits in both adults and children, emerges as a prudent approach to prevent the mentioned oral cavity diseases. Furthermore, such systems serve as excellent solutions for promoting nationwide preventive measures.

2. Related work

Recently, the scientific community has increasingly focused on utilizing technological solutions for oral cavity health care. Solutions developed for teaching oral hygiene primarily employ the following approaches: wearable electronics (smartwatches and bracelets), the creation of smart brushes using MEMS sensors, and the use of augmented and virtual reality technologies.

In [5], the authors describe the development of smart toothbrushes for monitoring teeth cleaning effectiveness using a recurrent probabilistic neural network (RPNN). To address the problem, they propose using a modified toothbrush with an inertial measurement unit (IMU) to

CMIS-2024: Seventh International Workshop on Computer Modeling and Intelligent Systems, May 3, 2024, Zaporizhzhia, Ukraine

dmytro.v.fedasyuk@lpnu.ua (D. Fedasyuk); ostap.truba.mnpzm.2022@lpnu.ua (O. Truba); tetiana.a.marusenkova@lpnu.ua (T. Marusenkova)

0000-0003-3552-7454 (D. Fedasyuk); 0009-0004-6177-623X (O. Truba); 0000-0003-4508-5725 (T. Marusenkova)

determine its spatial position. The main part of this work is dedicated to proper processing and recognition of movements occurring during teeth cleaning, utilizing the RPNN model. The authors highlight the advantages of RPNN over convolutional neural networks (CNN) and long short-term memory networks (LSTM), such as low computational resource usage, high recognition accuracy, and efficiency.

In [6], the study aims to present a protocol for developing a serious game to motivate oral hygiene practice in children. Kinect hardware (from Microsoft) is employed to track human movements. The system requires users to perform specific tasks (moves), after which it evaluates their actions and provides recommendations for improving toothbrushing technique. The author bases the work on the Stillman and Fones toothbrushing techniques.

The research [7] addresses toothbrush monitoring using augmented reality technology. The authors use multiple AR markers attached to a dodecahedron base to collect positioning data. The work also discusses the correlation between the number of AR markers and the accuracy of results. It is recommended to use three markers to achieve 95% monitoring accuracy. The authors emphasize the usefulness of this method in developing oral hygiene training systems.

In [8], a smartwatch equipped with an accelerometer is proposed for monitoring movements. A feedforward neural network is used for gesture recognition, with a publicly available UCI repository selected as the dataset. The authors aim to detect human motion primitives using a triaxial accelerometer.

Motion tracking in [9] employs augmented reality technology. The software system attempts to solve two problems: monitoring toothbrush movements and identifying soiled tooth surfaces. OpenCV library is used for toothbrush movement tracking, with a single AR (ArUco) marker attached to the toothbrush for positioning data. The authors state that the soiled tooth surface detection accuracy is 98%.

In [10], a smartwatch equipped with a magnetic sensor and a modified toothbrush with tiny magnets attached is proposed for motion tracking. This setup allows for the transmission of user motion data to the watch. The system comprises two phases: a training phase and a working phase. During the training phase, the user must perform tooth cleaning several times using the Bass technique for calibration. Subsequently, in the working phase, the user continues to use the software, receiving feedback on the correctness of their technique. At the conclusion of the study, the author notes that the system usage significantly improves the tooth cleaning technique in respondents, resulting in more effective plaque removal. The accuracy of gesture monitoring with this approach is reported to be 85.6%.

After analyzing recent publications in this field, one can identify the following shortcomings:

- Use of IMU sensors based on MEMS technology: studies were conducted using sensors of high accuracy, which are impractical for commercial projects. The authors propose using MEMS technology due to its low cost. However, MEMS sensors have low accuracy [11, 12].
- Use of laboratory conditions: studies using computer vision for motion monitoring did not consider external conditions (e.g., lighting).
- Significant computational resource usage for real-time data processing, reducing system data accessibility.
- Some studies utilized no known toothbrushing techniques when evaluating process effectiveness, a crucial factor for developing educational systems.
- Proposed methods that require IMU sensors to work are cost-ineffective, which makes them inaccessible to most people.

The described problems hinder the creation of an effective and accessible training system for oral cavity hygiene. This research aims to address the mentioned problems.

3. System architecture overview

In general, the software system should function as follows. The user selects the toothbrushing technique they aim to learn. Subsequently, the system provides visual instructions for the individual to follow. The software continuously monitors user activity in real time through a webcam. It evaluates the user's actions and provides recommendations for enhancing their performance in the completed exercises. From this standpoint, several issues emerge that the prospective software system needs to tackle:

- Capturing video streams
- Recognizing human faces
- Determining the spatial position of a toothbrush at any given moment
- Identifying brush-cleaning areas
- Evaluating the user's proficiency in executing a specific tooth cleaning technique and offering guidance for improvement.

Hence, to address the mentioned challenges, we propose the system architecture shown in Figure 1. The system comprises five modules (and a GUI), each targeting the mentioned issues.

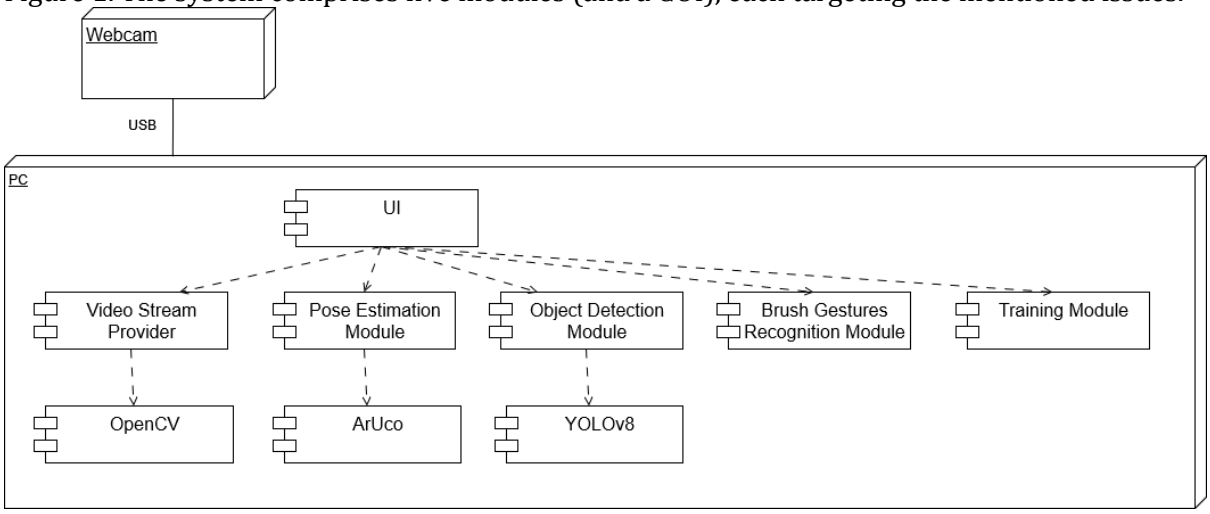

Figure 1: Deployment diagram of the proposed software system

Each module is designed to be interchangeable, facilitating the easy substitution of one implementation with another as needed (for instance, replacing one marker type with another). Furthermore, each module operates concurrently in a separate thread. It ensures that all computations can be executed in parallel, thus enhancing the overall system performance in terms of frames per second (FPS). Inter-module communication is facilitated through signals and slots. This mechanism is specific to QT and similar to the observer pattern. Data acquired from a module are transmitted to the main component (GUI) via signal emission. Subsequently, if necessary, the data is relayed to other components using the same principle.

The proposed architecture exhibits versatility and can be applied universally to similar application types, irrespective of the platform (Mobile, Web, PC). This adaptability underscores its potential utility across diverse technological environments.

4. Face detection

Recognition of the human face is primarily intended to check whether a person is visible and looking directly into the camera frame. It ensures that the system will not start/resume the training with no person present. It also guarantees that brush-cleaning areas are detected accordingly since the person looks directly into the camera.

For the object detection module implementation, we used a CNN specifically employing the YOLO (v8) single-stage detector [13]. We trained the model using an annotated open dataset comprising 1280 images of human faces (900 training and 380 validation). Each image had dimensions of 640x640 pixels. The number of epochs was 25.

5. Toothbrush pose estimation

5.1. Comparison of popular markers libraries

A fiducial marker system comprises planar (2D) markers positioned within a specific environment, intended to be detected by a camera across various applications. These markers enable the estimation of the object pose [7]. However, certain assumptions must be considered regarding the following factors:

- Marker Placement: Strategic positioning of fiducial markers within the environment is essential to maximize their visibility to the camera system while minimizing potential occlusions and obstructions. Also, it directly impacts marker pose estimation accuracy.
- Lighting Control: Environmental lighting conditions must be carefully controlled to minimize fluctuations that could deteriorate marker detection. Consistency in lighting setups is crucial for ensuring reliable performance.
- Distance Consideration: Variations in the distance between the camera and fiducial markers must be accounted for, as they can influence optimal marker size for the project as well as its detection.
- Detection Speed: In real-time application projects, the detection algorithm must work as fast as possible to ensure proper system response.

Given the metrics outlined above, an attempt can be made to evaluate and contrast existing marker libraries to identify the most suitable option for our project requirements. The selected candidates for comparison include ArUco, AprilTag, STag, CCTag, and ARTag (Figure 2).

Pose estimation and motion tracking efficacy highly depend on the marker type, size, tilt, the distance between the marker and camera, and marker occlusions. The smaller the marker size and the greater the mentioned distance, the slower marker detection. Besides, if the marker is partially occluded, it may be misrecognized. Not all the marker types provide the same results, i.e., the efficacy of pose estimation highly depends on the marker type. However, this issue is not covered in the literature. To choose the appropriate marker type, we designed a set of experiments with different marker sizes, distances from the camera, and percentages of occluded areas. The results are presented in Tables 1 – 5 (paragraphs 5.1.1 – 5.1.4).

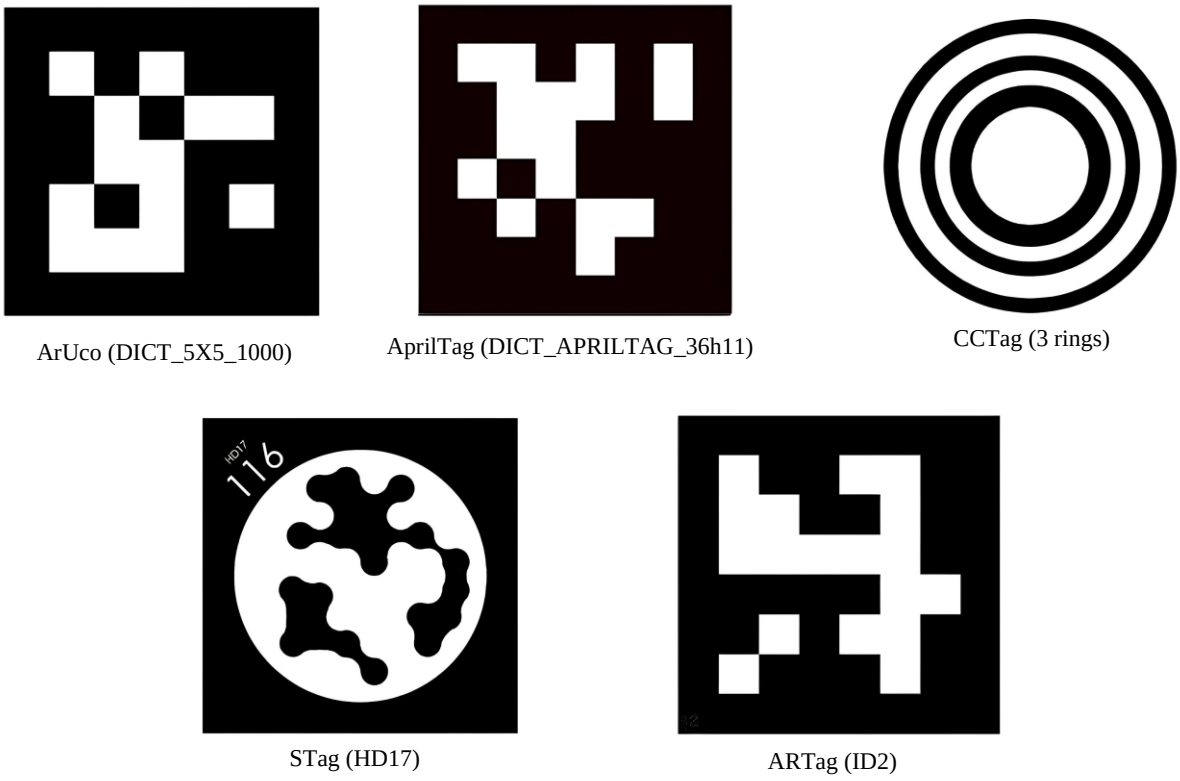

Figure 2: Fiducial markers used in our comparison with specified dictionaries

5.1.1. Marker size

Typically, the monitor should be within the range of distances from the eyes, specifically between the near (52 cm) and middle (73 cm) distances [14]. Therefore, it is imperative to determine the optimal marker size capable of detection within the range from 40 cm to 100 cm. The experimental setup involves placing the camera in a fixed position (XY) while the marker is positioned at various distances (with a 90-degree tilt; pitch rotation), including 40 cm, 60 cm, 80 cm, and 100 cm. Subsequently, the program executes 100 frames trying to detect the marker. We

performed ten experimental runs under the same lighting conditions and averaged the resulting data to derive the final marker detection rate. The results are resented in Tables 1 – 3.

Table 1

Comparing the marker detection rate with MSize = 10 mm

Marker/Distance	40 cm	60 cm	80 cm	100 cm
AprilTag	97%	69%	0	0
ArUco	98%	77%	0	0
STag	95%	75%	0	0
CCTag	0	0	0	0
ARTag	93%	38%	0	0

Table 2

Comparing the marker detection rate with MSize = 15 mm

Marker/Distance	40 cm	60 cm	80 cm	100 cm
AprilTag	100%	89%	62%	2%
ArUco	100%	95%	94%	7%
STag	100%	94%	84%	4%
CCTag	0	0	0	0
ARTag	100%	83%	42%	2%

Table 3

Comparing the marker detection rate with MSize = 20 mm

Marker/Distance	40 cm	60 cm	80 cm	100 cm
AprilTag	100%	100%	96%	93%
ArUco	100%	100%	100%	95%
STag	100%	100%	97%	95%
CCTag	3%	0	0	0
ARTag	100%	100%	92%	91%

5.1.2. Marker tilt angle

Marker tilt angle pertains to the degree of rotation or inclination of a fiducial marker relative to the camera's field of view. A tilted marker departs from its optimal alignment, potentially influencing its detectability and pose estimation accuracy. In the experimental configuration, the camera remains stationary (fixed in the XY plane) while tilted along different axes (pitch and yaw). Our findings indicate that most markers demonstrate satisfactory performance when tilted up to 65 degrees in both counter-clockwise and clockwise directions.

5.1.3. Marker detection speed

Several factors can impact the speed of marker detection, including the marker size, the detection algorithm efficiency, processing power, and lighting conditions. However, including such data in this paper may introduce bias given the variability introduced by these factors. Nonetheless, under stable and fixed conditions, such as consistent lighting, distance, and marker size, it would be beneficial to present speed detection data. In our experiments, we used the marker size of 16mm, positioned at a 40 cm distance. The webcam configuration is given in section 8. Table 4 shows the results.

Table 4

Comparing the marker detection speed

Marker	AprilTag	ArUco	STag	CCTag	ARTag
Time (s)	0.00425	0.00454	0.0143	0.2	0.00913

5.1.4. Marker occlusion

Marker occlusion refers to instances where a portion or the entirety of a fiducial marker is obstructed from the camera's view. Such obstruction can arise due to various factors, including

physical objects blocking the marker or the marker partially concealed behind another object. Some markers are specifically designed to exhibit greater resistance to occlusion.

To assess a marker's resistance to occlusion, we ran an experiment where a piece of paper covered a certain percentage (50%, 25%, 10%, 5%) of the total area of the marker. In this experiment, 20 mm markers were employed, and the camera was fixed along the Z-axis at a distance of 30 cm from the ground. The experiment findings are summarized in Table 5.

Table 5

Comparing the marker resistance to occlusion

Marker/ Occlusion	1% (4 mm)	5% (20 mm)	10% (40 mm)	25% (100 mm)	50% (200 mm)
AprilTag	Partial detection	0	0	0	0
ArUco	1	1	Partial detection	0	0
STag	1	1	1	Partial detection	0
CCTag	1	1	1	1	1
ARTag	Partial detection	0	0	0	0

5.2. Markers placement

To ensure optimal accuracy and precision in pose estimation with markers, it is imperative to arrange them in a configuration where at least two markers are visible simultaneously. This precautionary measure is vital for several reasons:

- Pose estimation encounters challenges when only a single marker is visible within the frame, resulting in ambiguity in solving the Perspective-n-Point (PnP) problem [15]. This limitation arises from the insufficient information provided by a single marker, impacting the accuracy and reliability of the pose estimation.
- Single-marker pose estimation is constrained by limitations in accuracy and the range of viewing angles. Without multiple markers for reference, the system may struggle to determine accurately the pose of the object being tracked.
- When used for brushing the inner side of dental areas, a single marker is susceptible to occlusion, obstructing the marker's view and impeding accurate tracking.

Considering these factors, it is advisable to utilize 3D objects instead of plain markers, such as cubes, tetrahedrons, or octahedrons. However, numerous studies focusing on object position tracking using fiducial markers [7, 16, 17] advocate for the adoption of a dodecahedron-based design, which offers favorable properties for robust tracking and accurate pose estimation.

5.3. Finalizing toothbrush design

After completing all of our experiments, we have finalized the following configuration: an ArUco marker with a size of 15.8 mm coupled with a dodecahedron-based object featuring edges measuring 17 mm in length, produced using an off-the-shelf 3D printer. The ArUco marker was selected due to its superior detection rate and optimal performance compared to other fiducial markers. The marker size was determined through range testing considerations, aiming to strike a balance between detectability and user comfort, particularly regarding the toothbrush handle. Figure 3 depicts our modified toothbrush design. Figure 4 shows its usage in the developed application.

Figure 3: Our proposed modified toothbrush for 6DOF tracking

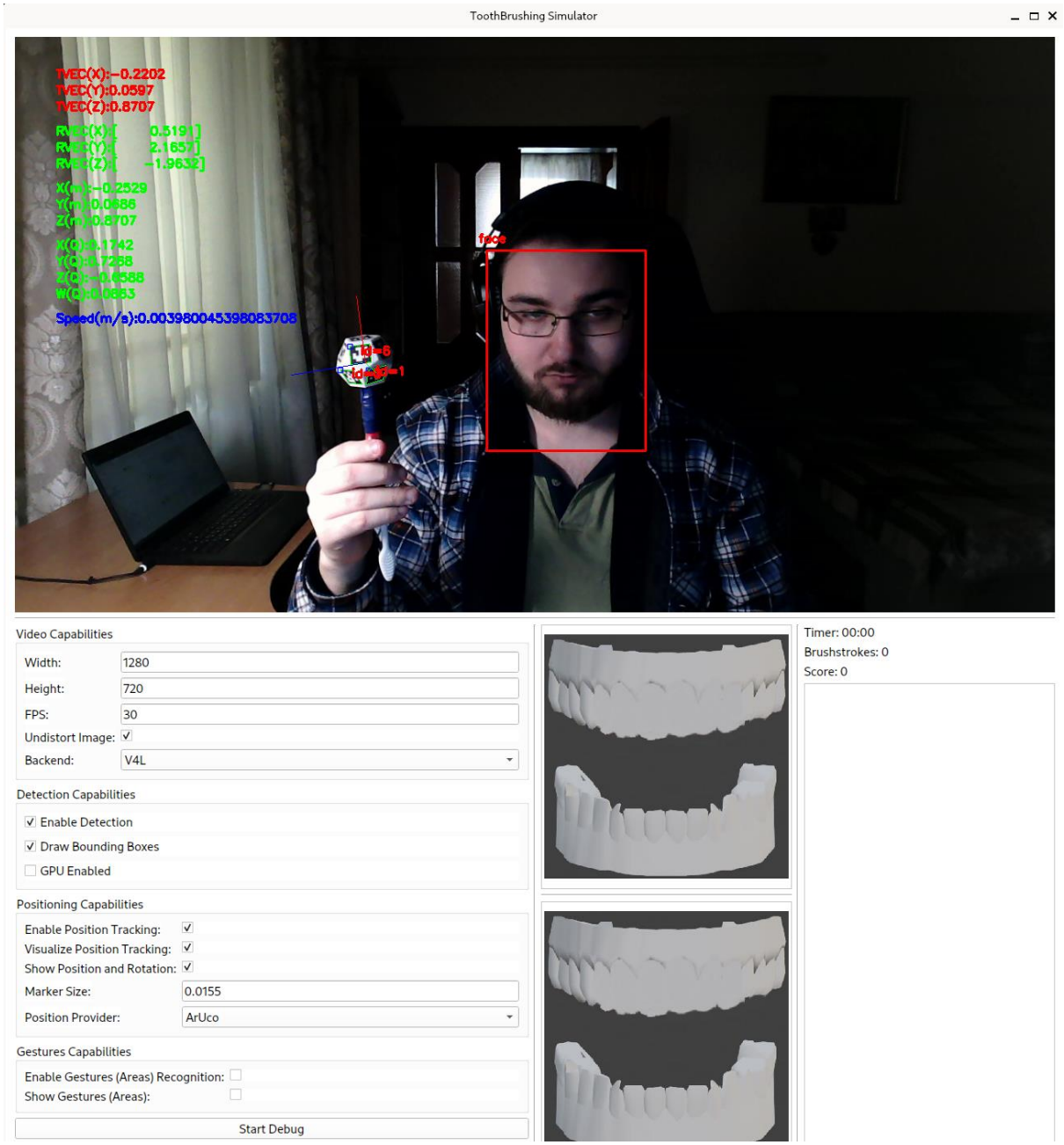

Figure 4: A sample of the marker usage in the developed application

5.4. Data filtration

Given the real-time nature of the system, characterized by swift movements of the marker and variations in camera quality, a considerable amount of noise is introduced within the captured frames. To somewhat mitigate this noise, we implemented a linear Kalman filter. This approach is expected to yield improved accuracy in tracking outcomes [18].

Initially, it is necessary to establish our state vector (1), comprising 18 states. These states encompass positional information (x, y, z) alongside their first and second derivatives (velocity and acceleration). Additionally, rotation is presented as three Euler angles (roll, pitch, yaw), accompanied by their respective first and second derivatives (angular velocity and acceleration).

$$X = (x \quad y \quad z \quad \dot{x} \quad \dot{y} \quad \dot{z} \quad \ddot{x} \quad \ddot{y} \quad \ddot{z} \quad \psi \quad \theta \quad \varphi \quad \dot{\psi} \quad \dot{\theta} \quad \dot{\varphi} \quad \ddot{\psi} \quad \ddot{\theta} \quad \ddot{\varphi})^T \qquad (1)$$

Next, one should determine the number of measurements. It amounts to 6. These measurements are derived from the rotation (R) and translation (t), yielding the positional coordinates (x, y, z) and the Euler angles (ψ, θ, φ). Furthermore, the number of control actions to apply to the system is specified, which, in this scenario, is zero. Lastly, we define the interval between measurements denoted as the differential time. In this instance, it is calculated as 1/T,

where T represents the frame rate of the video. We picked some random values of the process noise, measurement noise, and error covariance matrix.

The matrix (2) represents the state transition model A of the Kalman filter, which is used to predict the evolution of the state vector from one time step to the next in a linear dynamic system.

$$\begin{bmatrix}
1 & 0 & 0 & dt & 0 & 0 & dt2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & dt & 0 & 0 & dt2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & dt & 0 & 0 & dt2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & dt & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & dt & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & dt & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & dt & 0 & 0 & dt2 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & dt & 0 & 0 & dt2 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & dt & 0 & 0 & dt2 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & dt & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & dt & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & dt \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1
\end{bmatrix} \quad (2)$$

The matrix (3) represents the measurement model H of the Kalman filter, which relates the measurements obtained from sensors to the state vector. The first three rows in (3) indicate that the measurements directly correspond to the positional data (x, y, z). The next three rows indicate that the measurements directly correspond to Euler angles (ψ,θ,ϕ) in the state vector.

$$\begin{bmatrix}
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0
\end{bmatrix} \quad (3)$$

6. Detecting brush cleaning area

The Brush Cleaning Areas Recognition module is tasked with identifying specific regions of the brushing area associated with particular brushing techniques. For instance, the Bass technique implies 15 areas, including upper and lower segments on the right and left sides, outer segments, and specific regions within the incisors.

When employing an IMU sensor with 9 Degrees of Freedom (9DOF), contemporary solutions often suggest using deep learning architectures such as CNNs, LSTMs, and RPNNs, among others [5, 10]. However, integrating existing implementations with our 6 Degrees of Freedom (6DOF) tracking system poses considerable challenges due to several factors. The unpredictable positioning of the camera within the user's environment, coupled with the variable distance between the camera and the user (ranging from 40 to 70 cm), renders the utilization of X, Y, and Z positioning impractical. Consequently, the available data is limited to rotations around the X, Y, and Z axes (3DOF), presenting significant obstacles in accurately discerning the brushing region even with a comprehensive dataset.

As a result, the efficient adaptation of current methodologies for identifying brushing areas within the context of our tracking system presents notable difficulties, primarily due to the inherent limitations of rotational data and the unpredictable nature of the user's setup.

7. Training module

The Training module is structured to fulfill the following objectives:

- Monitoring and tallying the number of brushing strokes performed by the user.

- Initiating and terminating the brushing timer to track the duration of brushing sessions.
- Assessing the user's proficiency in tooth cleaning techniques based on available data.
- Managing the training session by pausing or resuming activities when the user is absent or present in the frame.
- Providing personalized guidance and recommendations for improvement.

The counting of brushing strokes can be accomplished through two proposed methods. The first method involves utilizing an acoustic sensor, such as a microphone, to detect the occurrence of brush strokes [19]. However, this approach implies that the user has a noise-free environment and relies on the brush strokes audible enough to be registered on the microphone. Alternatively, the second method entails leveraging positioning data obtained from the Pose Estimation Module. This data encompasses the X, Y, and Z coordinates between two consecutive frames, as well as changes in velocity, enabling the calculation of brush stroke occurrences.

Regarding the assessment of the user's proficiency in executing specific tooth-cleaning techniques, one can conduct a comprehensive analysis utilizing various data collected during the training session. These include but are not limited to the total count of brush strokes performed by the user, the average speed of brush strokes executed during the session, the duration of time allocated to brushing specific areas within the oral cavity, the ratio of correctly executed movements to the total number of movements performed. By incorporating these metrics into the evaluation process, one can attain a holistic understanding of the user's performance and adherence to the prescribed tooth-cleaning techniques.

8. Experimental setup

We evaluate each of the proposed and implemented system modules (specifically Face Detection and Toothbrush Pose Estimation), as well as some other mentioned results (described in section 5.1) on the following hardware/software configuration:

- Desktop: OS: Fedora Linux 39 (6.7.3 kernel); CPU: AMD Ryzen 5 5600X (12) @ 3.700GHz; RAM: 32Gb DDR4.
- SW Dependencies: OpenCV: 4.9.0.80; YOLOv8 (Ultralytics): 8.1.11; PySide6: 6.6.1; Python: 3.12.1; Numpy: 1.26.4.
- Webcam: Model: Asus Webcam C3; Resolution: 1280x720; Sensor Resolution: 2 Mpix; FPS: 30; Codec: MJPEG; Exposure: Auto Exposure is disabled, Exposure Time 1/60s; Auto Focus: Disabled.
- Marker (on dodecahedron): Size: 15.8 mm; Dictionary: DICT_5X5_1000; IDs Range: 0 – 11; Type: ArUco.

Camera calibration is of key importance [20, 21, 22]. Our camera is calibrated using a 10x15 chessboard pattern with a marker size of 18 mm, attached to a wooden board (Figure 5).

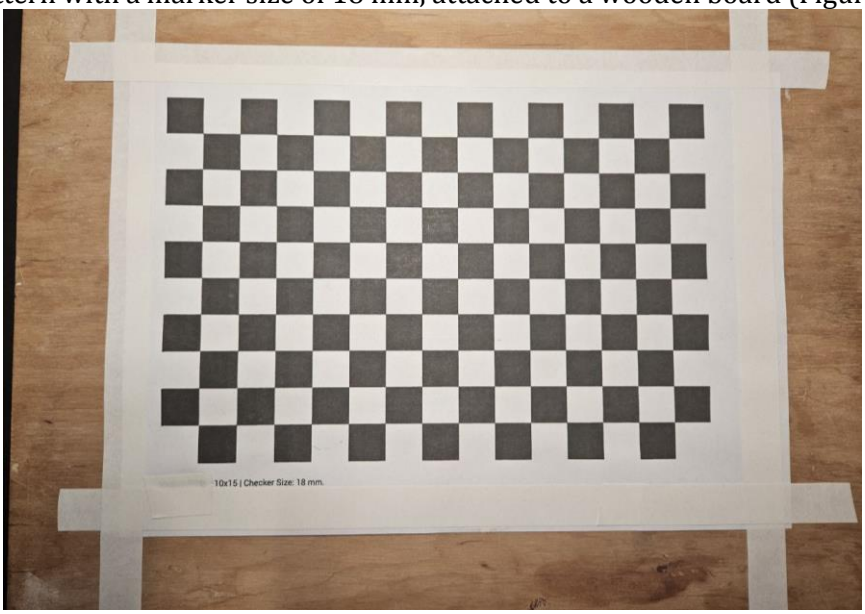

Figure 5: Calibration board (chessboard, 10x15, marker size 18 mm)

Root mean square (RMS) reprojection error – 0.223. The camera intrinsics (camera matrix and distortion coefficients) are the following:

$$\begin{bmatrix} 641.125534 & 0 & 628.572419 \\ 0 & 641.125534 & 351.051382 \\ 0 & 0 & 0 \end{bmatrix} = K$$

$$[0.0519576 \quad -0.0213721 \quad 0 \quad 0 \quad 0.0030762] = D$$

9. Results

9.1. Face detection accuracy and precision

In the assessment of our trained YOLO model, we have selected the following metrics:

- TP (True Positive): Refers to the tally of positively classified samples with accuracy.
- TN (True Negative): Represents the count of negatively classified samples accurately.
- FP (False Positive): Indicates the number of negatively classified samples inaccurately labeled as positive.
- FN (False Negative): Signifies the number of positively labeled samples inaccurately categorized as negative.
- Precision: Quantifies the ratio of TP to the total number of predicted positive instances.
- Recall: Measures the ratio of TP to the total number of actual positive occurrences.
- AP (Average Precision): Provides a measure for assessing the precision-recall curve.
- F1 Score: serving as an overarching performance indicator, reflects the harmonic mean of precision and recall. It is computed by doubling the product of precision and recall, then dividing by their sum.

When using the YOLOv8 model, all essential metrics for evaluating our model's performance are readily available. True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN) can be extracted from the confusion matrix (Figure 6). Specifically, TP = 166, FP = 6, FN = 3, TN = 0. The precision is measured at 0.9651, recall at 0.9822, and the F1 score at 0.9736. The overall Average Precision (AP) stands at 0.993. Inference time – 28.3ms.

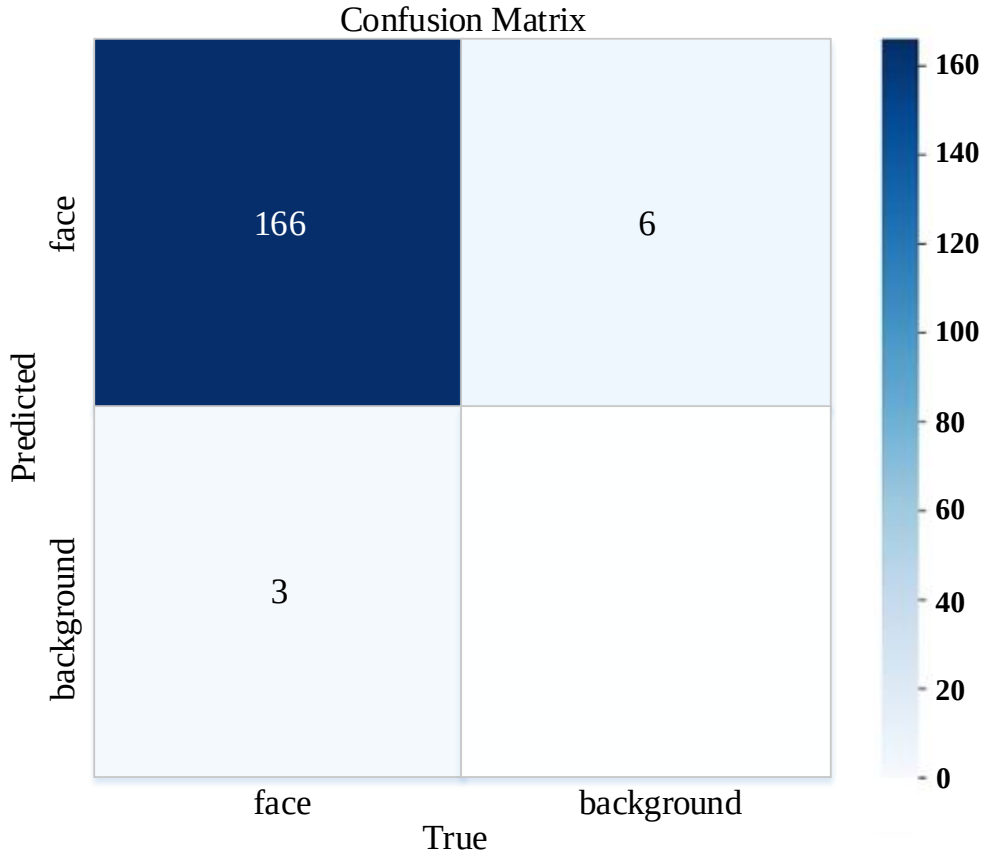

Figure 6: Confusion matrix

9.2. Pose estimation accuracy

To assess the pose estimation accuracy, we established an experimental setup using a grid pattern drawn on an A2 sheet of paper, with each square having a side length of 5 cm. The experiment involved moving a toothbrush incrementally in each plane direction by 5 cm and by 10 cm in the Z-direction while maintaining the camera's position at a fixed Z-coordinate of 1 meter. The experiment outcomes are presented in Table 6.

Table 6

Pose estimation accuracy

	Ground truth	Estimated values	Absolute error	Percentage
X (m) (at Z = 1m)	0.1	0.108	0.008	8%
	0.15	0.159	0.009	6%
	0.2	0.211	0.011	5.5%
Y (m) (at Z = 1m)	0.1	0.1075	0.0075	7.5%
	0.15	0.1611	0.0111	7.4%
	0.2	0.214	0.014	7%
Z (m)	0.8	0.824	0.024	3%
	0.9	0.929	0.029	3.2%
	1	1.031	0.031	3.1%

9.3. Usage of Kalman's filter on bad pose rejection and noise reduction

To evaluate Kalman's Filter efficacy for noise reduction and mitigation of undesired pose fluctuations, we conducted a series of experiments. They were meticulously designed to assess the filter's capability to reject false pose estimations and minimize the ambient noise impact:

1. Steady Movement along the X-axis in one direction: The toothbrush undergoes deliberate, unhurried motion exclusively along the negative direction of the X-axis.
2. Steady Movement along the X-axis in both directions: Similar to the preceding scenario, the toothbrush undergoes deliberate, unhurried motion along the X-axis in both directions.
3. Steady Movement along the X-axis with Marker Occlusion: As in the initial scenario, the toothbrush undergoes consistent motion along the X-axis. However, a distinct feature of this scenario is an intentional occlusion of the marker during the midway point of the trajectory.

In the absence of Kalman filtering (Figures. 7 and 8), the pose exhibits susceptibility to undesirable vibration noise, closely mirroring its fluctuations. Such behavior proves disadvantageous in our application, where stability in pose determination is crucial.

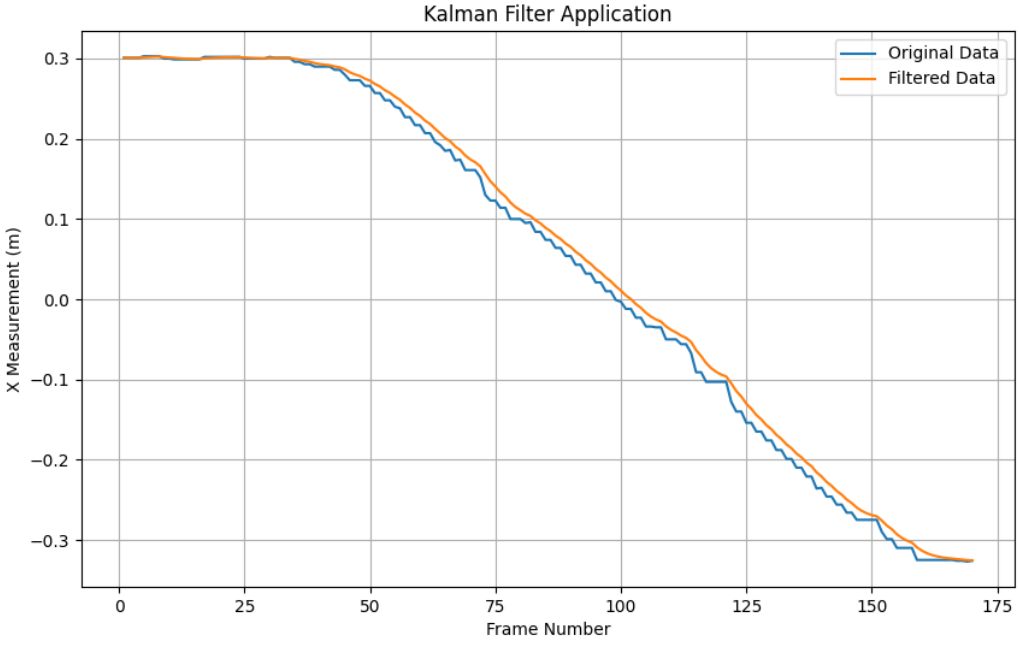

Figure 7: Fluctuations in unfiltered data

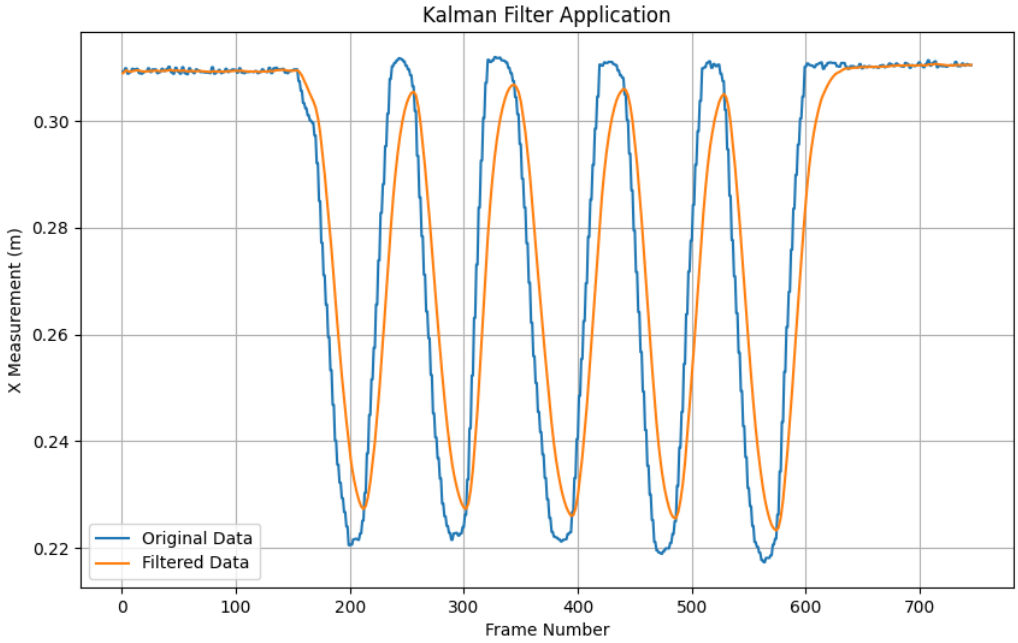

Figure 8: Filtered vs unfiltered data

Conversely, upon the Kalman filter usage, the pose maintains a relatively stable trajectory, mitigating the adverse effects of noise-induced fluctuations.

In scenarios where no pose information is accessible due to occlusion of the ArUco marker (Figure 9), the conventional methods fail to provide reliable estimations. However, with the integration of the Kalman filter into our approach, the system remains capable of inferring the state and anticipating marker locations during these occluded intervals. The outcomes of our analysis demonstrate the algorithm's efficacy in addressing occlusion challenges, yielding optimal estimations despite the absence of direct pose data.

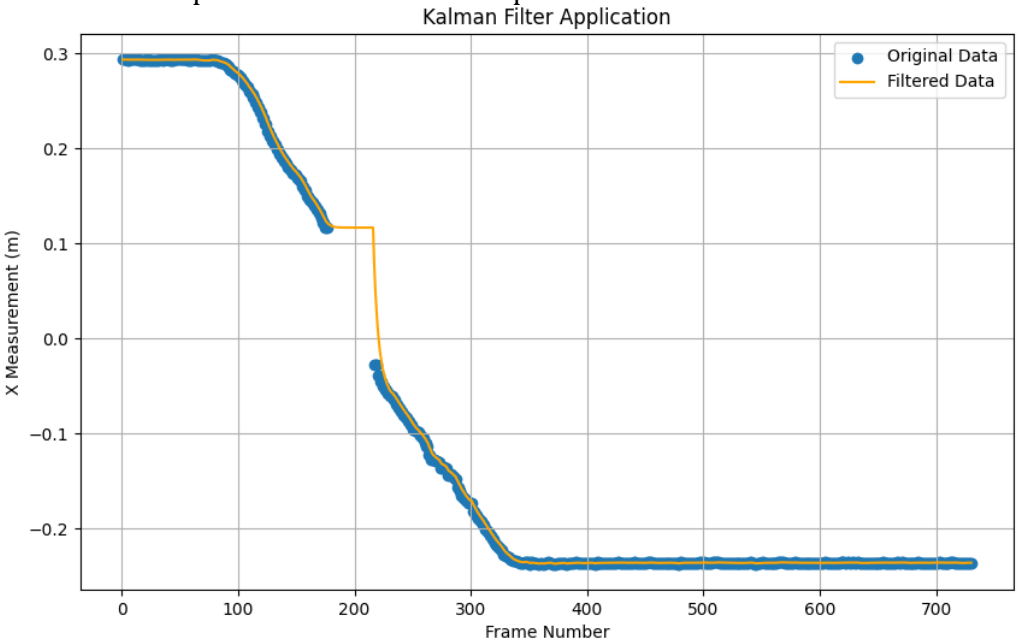

Figure 9: Filtering at ArUco occlusion

10. Discussion

In the proposed system architecture, which holds potential for universal application across similar projects, and exhibits commendable accuracy in pose estimation, several critical impediments to its real-world applicability have been identified.

Initially, the facial detection mechanism within the project predominantly functions as a superficial feature rather than a core component. Its primary use—to ascertain whether a user is

facing the camera, thereby enabling the pause and resume of training sessions and, theoretically, assisting in brush area identification—limits its practical utility. Moreover, the attainment of precise pose estimation is contingent upon fulfilling numerous prerequisites:

- Camera Calibration: Effective pose estimation necessitates user-initiated camera calibration, incorporating an automated procedure within our system. Although explored in existing studies, camera calibration is not a straightforward task. The inevitable variance in distortion coefficients is likely to introduce bias into the data collected, thereby compromising the accuracy of pose estimation.
- Toothbrush Modification: To utilize our system, users have to engage in a modification process for their toothbrush. It involves the creation of a dodecahedron using a 3D printer and affixing markers onto it. However, the accessibility of 3D printers remains limited, and the precise alignment and attachment of markers to the dodecahedron is crucial. Any inaccuracies in this process may compromise the accuracy of marker detection and pose estimation, thus impacting the overall system functionality.
- Rolling-shutter Cameras: Predominantly, contemporary webcams operate on a rolling-shutter mechanism, capturing images not instantaneously but by rapidly scanning the scene. This approach results in predictable distortions of swiftly moving objects or intense light fluctuations, leading to the erroneous detection of markers and inaccurate pose estimation. Solutions to this challenge encompass transitioning to cameras with a global shutter, increasing shutter speed—at the expense of potential exposure issues—and the implementation of advanced Image Processing Algorithms.
- Lighting Conditions: The marker detection is linked to lighting conditions, requiring an environment that is neither excessively bright nor dim. This issue is mitigated when employing an IMU solution but is exacerbated by reduced exposure times.

Lastly, the utilization of the proposed 6DOF marker pose estimation system introduces significant challenges in accurately determining the brushing area, a dilemma that demands a viable resolution as detailed in sections 6 and 7 of our analysis.

This comprehensive evaluation underscores the complexities and limitations inherent in the deployment of the proposed system within practical settings, highlighting the necessity for further refinement and adaptation to overcome these obstacles. The future work implies the data fusion of the obtained results with IMU data and gesture recognition [23].

References

[1] T. A. Ghebreyesus, Global oral health status report: towards universal health coverage for oral health by 2030. regional summary of the African region, 2023. URL: https://www.who.int/publications/i/item/9789240070769.

[2] J. A. Pieren, D. M. Bowen, Darby and Walsh dental hygiene e-book: theory and practice, 5th ed., Elsevier, Amsterdam, 2019.

[3] D. Slot, L. Wiggelinkhuizen, N. Rosema, G. Van Der Weijden, The efficacy of manual toothbrushes following a brushing exercise: a systematic review: how effective are manual toothbrushes? International Journal of Dental Hygiene 10(3) (2012) 187–197. doi: 10.1111/j.1601-5037.2012.00557.x.

[4] A. B. Londero, A. P. Reiniger, R. C. Tavares, C. M. Ferreira, U. M. Wikesjö, K. Z. Kantorski, C. H. Moreira, Efficacy of dental floss in the management of gingival health: a randomized controlled clinical trial, Clinical Oral Investigations 26(8) (2022) 5273–5280. doi: 10.1007/s00784-022-04495-w

[5] C.-H. Chen, C.-C. Wang, Y.-Z. Chen, Intelligent brushing monitoring using a smart toothbrush with recurrent probabilistic neural network, Sensors 21(4) (2021). doi: 10.3390/s21041238

[6] S. N. Amantini, A. A. Montilha, B. C. Antonelli, K. T. Leite, D. Rios, T. Cruvinel, N. Lourenço Neto, T. M. Oliveira, M.A. Machado, Using augmented reality to motivate oral hygiene practice in children: Protocol for the Development of a Serious Game, JMIR research protocols 9(1) (2020). doi: 10.2196/10987.

[7] S. Hayakawa, G. Al-Falouji, G. Schickhuber, R. Mandl, T. Yoshida, S. Hangai, A method of toothbrush position measurement using AR markers, in: 2020 IEEE 2nd Global Conference

on Life Sciences and Technologies, LifeTech, Kyoto Japan, 2020, pp. 91–93. doi: 10.1109/LifeTech48969.2020.1570619103.
[8] M. Fahim, V. Sharma, T. Q. Duong, A wearable-based preventive model to promote oral health through personalized notification, in: 2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society, EMBC'2022, IEEE, Glasgow United Kingdom, 2022, pp. 4282–4285. doi: 10.1109/EMBC48229.2022.9871128.
[9] H. Kondo, K. Funahashi, AR tooth brushing system to promote oral care habits of children, in: 2021 Nicograph International, NicoInt'21, IEEE, Tokyo Japan, 2021, pp. 115–115. doi: 10.1109/NICOINT52941.2021.00033.
[10] Z. Hussain, D. Waterworth, M. Aldeer, W. E. Zhang, Q. Z. Sheng, J. Ortiz, Do you brush your teeth properly? An off-body sensor-based approach for toothbrushing monitoring, in: 2021 IEEE International Conference on Digital Health, ICDH'21, IEEE Chicago IL USA, 2021, pp. 59–69. doi: 10.1109/ICDH52753.2021.00018.
[11] D. Fedasyuk, R. Holyaka, T. Marusenkova, A tester of the MEMS accelerometers operation modes, in: 2019 3rd International Conference on Advanced Information and Communications Technologies, AICT'19, IEEE, Lviv Ukraine, 2019, pp. 227–230. doi: 10.1109/AIACT.2019.8847840.
[12] D. Fedasyuk, R. Holyaka, T. Marusenkova, Method of Analyzing Dynamic Characteristics of MEMS Gyroscopes in Test Measurement Mode, in: 2019 9th International Conference on Advanced Computer Information Technologies, ACIT'19, IEEE, Ceske Budejovice, Czech Republic, 2019, pp. 157–160, doi: 10.1109/ACITT.2019.8780058.
[13] G. Yocher, A. Chaurasia, Ultralytics YOLOv8 Docs, 2023. URL: https://docs.ultralytics.com/.
[14] D. Rempel, K. Willms, J. Anshel, W. Jaschinski, J. Sheedy, The effects of visual display distance on eye accommodation, head posture, and vision and neck symptoms, Human Factors 49(5) (2007) 830-838. doi: 10.1518/001872007X230208.
[15] H. -Y. Tseng, P. -C. Wu, M. -H. Yang, S. -Y. Chien, Direct 3D pose estimation of a planar target, in: 2016 IEEE Winter Conference on Applications of Computer Vision, WACV, Lake Placid, NY, USA, 2016, pp. 1–9. doi: 10.1109/WACV.2016.7477640.
[16] P.-C. Wu, R. Wang, K. Kin, C. Twigg, S. Han, M.-H. Yang, S.-Y. Chien, DodecaPen: Accurate 6DoF tracking of a passive stylus, in: Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology, UIST'17, Association for Computing Machinery, Québec City QC Canada, 2017, pp. 365–374. doi: 10.1145/3126594.3126664.
[17] P. García-Ruiz, F. J. Romero-Ramirez, R. Muñoz-Salinas, M. J. Marín-Jiménez, R. Medina-Carnicer, Fiducial objects: custom design and evaluation, Sensors 23(24) (2023). doi: 10.3390/s23249649.
[18] H. C. Kam, Y. K. Yu, K. H. Wong, An Improvement on ArUco Marker for Pose Tracking Using Kalman Filter, in: 2018 19th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD), IEEE, Busan Korea (South), 2018, pp. 65–69. doi: 10.1109/SNPD.2018.8441049.
[19] H. Huang, S. Lin, Toothbrushing monitoring using wrist watch, in: Proceedings of the 14th ACM Conference on Embedded Network Sensor Systems CD-ROM, SenSys '16, ACM, Stanford CA USA, 2016, pp. 202–215. doi: 10.1145/2994551.2994563.
[20] H. Rezazadegan Tavakoli, H. R. Pourreza, An automated camera calibration framework for desktop vision systems, in: 2009 International Conference on Advances in Computational Tools for Engineering Applications, IEEE, Beirut Lebanon, 2009, pp. 96–100. doi: 10.1109/ACTEA.2009.5227921.
[21] L. Tan, Y. Wang, H. Yu, J. Zhu, Automatic camera calibration using active displays of a virtual pattern, Sensors 17(4) (2017). doi: 10.3390/s17040685.
[22] S. Su, W. Heidrich, Rolling shutter motion deblurring, in: 2015 IEEE Conference on Computer Vision and Pattern Recognition, CVPR, Boston, MA, USA, 2015, pp. 1529–1537. doi: 10.1109/CVPR.2015.7298760.
[23] L. Ivanska, T. Korotyeyeva, Mobile real-time gesture detection application for sign language learning, in: 2022 IEEE 17th International Conference on Computer Sciences and Information Technologies, CSIT'22, IEEE, Lviv Ukraine, 2022, pp. 511–514. doi: 10.1109/CSIT56902.2022.10000440.