Research Note

Jaw Motion Tracking in Unity using a monocular camera and ArUco markers

INOUE Takanobu
Digital Hollywood University, Graduate School, Graduate Student, Kobayashi Dental Clinic

SHIRASHIMA Naoto
Digital Hollywood University, Graduate School, Graduate Student

HOSHINO Hiroyuki
Digital Hollywood University, Graduate School

HASHIMOTO Masatsugu
Digital Hollywood University, Graduate School

This study developed a low-cost jaw movement tracking system using a monocular camera and planar markers (ArUco) with promote the widespread adoption of jaw movement analysis in dental clinical practice. The system was built on Unity 6 and OpenCV for Unity platforms, implementing algorithms for Perspective-n-Point (PnP) problem solving, three-dimentional rigid body fusion using the Kabsch method, and terminal hinge axis estimation through weighted least squares methods. Compared to existing expensive jaw movement measurement devices, this system can be configured using only general-purpose cameras and markers, enabling easy chairside utilization. This paper reports on the accuracy verification for clinical applications and discusses future prospects.

1. Introduction

1.1 Clinical Significance of Jaw Movement Measurement

Accurate recording and analyses of jaw movement are essential for successful prosthetic treatment. Particularly in full-mouth occlusal reconstruction and the diagnosis of temporomandibular disorders (TMD), understanding mandibular movement trajectories and centers of rotation serves as a crucial indicator for treatment planning decisions. However, in current clinical practice, jaw movement measurement using dedicated equipment is not positioned as a standard examination in the Diagnostic Criteria for Temporomandibular Disorders (DC/TMD)[1] or in the clinical guidelines of the Japanese Society for the Temporomandibular Joint[2]. The primary reasons for this are the high implementation costs of the existing systems, complex operational procedures, and difficulties in interpreting measurement results.

1.2 Historical Development of Jaw Movement Kinematics

Jaw movement fundamentally consists of two components: rotational and translational (sliding) movements. While anteroposterior movements (protrusive and retrusive movements) primarily involve only translational motion, opening and closing movements represent a complex combination of rotational and translational motions. Understanding these complex movement patterns has gradually deepened with the development of jaw movement kinematics.

In the early 20th century, the gnathology school founded by Dr. Beverly B. McCollum and Charles E. Stuart proposed the existence of a fixed rotational axis (terminal hinge axis) in mandibular movement. Posselt theorized that during the initial phase of the opening movement (approximately 20-25mm), the mandibular condyle performed purely rotational motion, with anterior translational movement being added only during subsequent opening phases[3]. This "initial rotation-later translation" two-stage theory became the theoretical foundation for the development of fully adjustable articulators and pantographs based on the hinge axis has served as a guideline for prosthetic treatment for many years.

Subsequently, Dr. Shoji Kohno from Japan proposed the concept of the "kinematic axis," which captured jaw movement more dynamically[4]. Dr. Kohno mathematically analyzed the trajectories of anterior-posterior and opening-closing movements, and deveroped methods to determine instantaneous centers of rotation. In this theory, the rotational axis of jaw movement was understood not as fixed, but as a dynamic axis that continuously changes according to movement.

Aprecise three-dimensional analysis using modern digital jaw movement measurement devices (CADIAX, JMA-Optic, etc.) has led to significant modifications to conventional theories. These high-precision measurements revealed that even during the initial phase of the opening movement, translational motion is present alongside rotational motion, albeit minimally, rather than pure rotational movement[5]. This demonstrated that the hypothesis of "initial purely hinge rotation" proposed by the gnathology school differs from actual biological movement.

Furthermore, detailed research on the instantaneous rotational axis of jaw movement has advanced in recent years. In 2023, Bescond et al. reported complete three-dimensional kinematic parameters of the temporomandibular joint using a novel method that fused MRI data with optical motion capture[6]. This study demonstrated through an instantaneous helical axis (IHA) analysis that the rotational axis is oriented obliquely during opening-closing movements and its position continuously changes during motion. Particularly interesting was the observation of asymmetric movement patterns between the left and right temporomandibular joints, with different movement trajectories on the working and balancing sides. These findings indicate the reality of more complex jaw movements that cannot be explained by conventional fixed-axis theories or simple dynamic axis models, suggesting the need for fundamental reconstruction of the jaw movement theory.

Based on this historical background and the latest findings, The present study aimed to simple and accurate a method for obtaining the terminal hinge axis from the perspective of clinical reproducibility. Specifically, the goal was to establish a practical approach applicable to clinical use while leveraging

the advantages of both conventional fixed-axis and dynamic-axis theories.

The current diagnosis of temporomandibular disorders is based on interviews, visual examination, palpation, and simple measurements of mouth opening range and lateral movement as indicated by DC/TMD and domestic guidelines, with jaw movement measurement using dedicated equipment not positioned as a standard examination. This is because interpretation of device measurements is difficult, and the diagnostic added value relative to cost and time is limited. However, if "simple and reproducible movement indicators" that can be used rapidly in clinical settings and intuitively communicate results to patients could be established, they would contribute to disease type stratification and objective evaluation of treatment efficacy, potentially becoming a powerful tool to reinforce the chairside diagnosis emphasized by the guidelines.

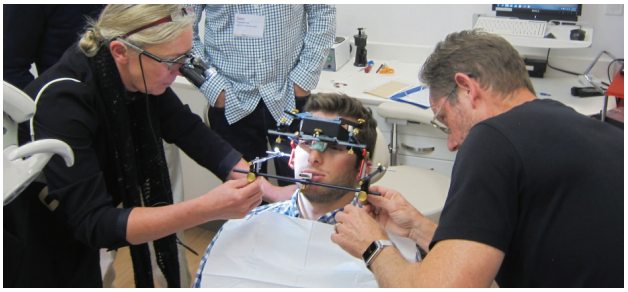

Figure1: CADIAX4: A Digital Jaw Movement Measurement Device as the Modern Version of the Pantograph

This system records mandibular movement trajectories with high precision using a digital recording plate mounted on the temporal region. The condylar movement pathway is back-calculated from stylus movements to identify the position of the terminal hinge axis. While this device can record motion in 6 degrees of freedom, the implementation cost exceeds 5 million yen.

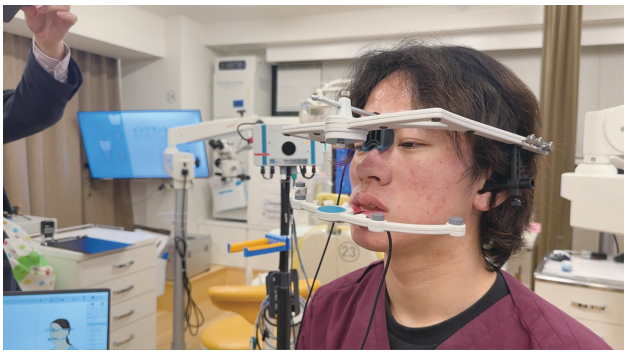

Figure 2: WinJMA

This system measures jaw position and orientation in a contactless manner using ultrasonic sensors. Eight sensors record motion in 6 degrees of freedom, enabling real-time three-dimensional display. Compared to magnetic systems, it has the advantage of being less susceptible to interference from metallic prostheses. The latest JMA-Optic model is not approved in Japan.

2. Related Work and Positioning

2.1 Classification of Existing Jaw Movement Measurement Systems

Table1 shows the major commercially available jaw movement measurement systems.While these systems enable high-precision measurements, they typically exceed 3 million yen in implementation cost, making introduction difficult for small-scale dental clinics.

Table 1: Comparison of jaw tracking systems

No.	Device Name	Manufacturer	Recording Method	Sensor Details	DOF	Contact Type
1	JMA-Optic	Zebris Medical GmbH	Optical	Digital camera ×2	6 DOF	Non-contact
2	ARCUSdigma 3	KaVo Dental	Ultrasonic	ultrasonic sensor x8	6 DOF	Non-contact
3	MM-J2	Shofu	Optical	LED + PSD sensor ×2	6 DOF	Non-contact
4	K7x	Myotronics	Magnetic	Magnetic sensor ×8	6 DOF	Non-contact
5	Modjaw	Modjaw	Optical	Digital camera ×2	4D tracking	Non-contact
6	e-Motion	SHINING 3D	Optical	Digital camera ×4	6 DOF	Non-contact
7	Cadiax 4	GAMMA Dental	Electronic	sensor flags ×2	6 DOF	Contact
8	**Present work**	-	Optical	ArUco ×6, solvePnP	6 DOF	Non-contact

2.2 Novelty and Clinical Significance of This Study

The system proposed in this study is differentiated from existing systems in the following aspects:

(1)**Low-cost implementation**: Composed only of a digital camera ZVE-10M2 (Sony Corporation, Tokyo, Japan) and ArUco markers (printing cost)

(2)**Open-source foundation**: Uses Unity 6 (Unity Technologies, San Francisco, USA) and OpenCV, enabling easy customization

(3)**Immediate chairside utilization**: Completion from calibration to measurement within 5 minutes

(4)**Visual feedback**: Enables use in patient education through 3D real-time display

3. Methods

3.1 System Configuration and Operating Environment

The basic configuration of this system is shown below:

(a)Camera: Sony ZV-E10M2 (Sony Corporation, Tokyo, Japan) (4K@30fps, via NVIDIA Broadcast (NVIDIA Corporation, Santa Clara, USA))

(b)Markers: ArUco dictionary DICT_4X4_50, edge length 10mm × 6 pieces

(c)Validation model: Dental cast mounted on Handy 2A average value articulator (Handy Dental Supply, Tokyo, Japan)

Software Environment

(a)Unity 6 (Unity Technologies, San Francisco, USA) (3D environment construction)

(b)OpenCV for Unity 3.0.0 (OpenCV.org, open-source project) (image processing and pose estimation)

(c)Modules used: Aruco, Calib3d, Imgproc

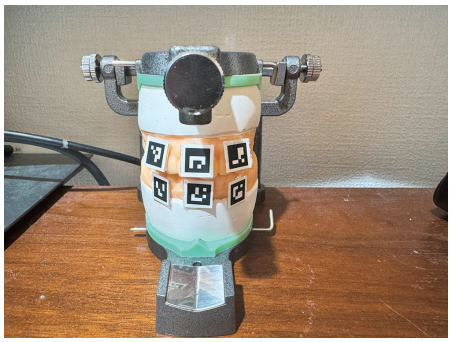

Figure 3: Marker Placement on Experimental Articulator

Three ArUco markers (ID 0-2) were placed on the maxillary section and three markers (ID 3-5) on the mandibular section of the Handy 2A average value articulator (Shofu Inc., Kyoto, Japan). The markers were fixed with strong adhesive to maintain a consistent positional relationship relative to the dental arch.

3.2 Camera Calibration and Coordinate System Transformation

3.2.1 Estimation of Intrinsic Parameters

Camera calibration is an important process for establishing the correspondence between 2-D image coordinates and 3-D world coordinates. In this system, using OpenCV's checkerboard detection function, the procedure is as follows:

Checkerboard specifications:

- Number of inner corner points: 10 × 7 (grid points)
- Square size: 25 mm
- Minimum required number of frames: 15

Figure4: Checker Board

During calibration, the Calib3d.findChessboardCorners function detects the corner points in each frame, and pairs of corresponding 3-D coordinates (given by the known checkerboard dimensions) and 2-D image coordinates are collected. Once 15 or more multi-view frames have been acquired, the camera intrinsic parameter matrix is estimated with Calib3d.calibrateCamera.

3.2.2 Implementation of Coordinate System Transformation

Because OpenCV and Unity define their coordinate systems differently, an appropriate transformation is required:

Table2: Axis Orientation in OpenCV and Unity Coordinate Systems

Axis	OpenCV (image coordinate convention)	Unity
X	Rightward	Rightward
Y	Downward (image coordinates)	Upward
Z	Forward from the camera	Forward

In the code implementation (ConvertToMatrix4x4 method), this coordinate system transformation is achieved by inverting the signs of the Y and Z components of the rotation matrix obtained via the Rodrigues formula: Y-axis inversion (OpenCV downward → Unity upward)

3.3 ArUco Marker Detection and Pose Estimation

3.3.1 Image Preprocessing Pipeline

To improve marker detection accuracy, the following preprocessing steps are applied:

(1)Grayscale conversion: convert RGBA color images to luminance images

(2)Bilateral filter: remove noise while preserving edges (d = 9, σ_color = 75, σ_space = 75)

(3)CLAHE (Contrast Limited Adaptive Histogram Equalization): local contrast enhancement (clipLimit = 2.0, tileGridSize = 8 × 8)

These preprocessing steps enable stable marker detection even under non-uniform illumination conditions.

3.3.2 6DoF Pose Estimation via the PnP Problem

For ArUco marker detection, the ArucoDetector class in OpenCV for Unity 3.0.0 is used. The detection process is as follows:

Marker detection:

Obtain the predefined ArUco dictionary DICT_4X4_50, combine it with the default DetectorParameters, and construct an ArucoDetector. Then call detectMarkers(grayMat, corners, ids) to detect ArUco markers in the input grayscale image grayMat, storing the four-corner coordinates of each marker in corners and the corresponding marker IDs in ids. Because these objects are created inside using blocks, they are automatically disposed when the block ends, releasing their native resources.

Pose estimation via a PnP solver:

For each detected marker, the rotation vector (rvec) and translation vector (tvec) are estimated by the solvePnP function from the 2D–3D correspondences of the four corner points. The 3D coordinate model of the marker is defined as a square with an edge length of 10 mm:

Outlier handling and smoothing:

If the difference in pose from the previous frame exceeds 90 degrees, it is discarded as an outlier, and the pose is smoothed by an exponential moving average (α = 0.2).

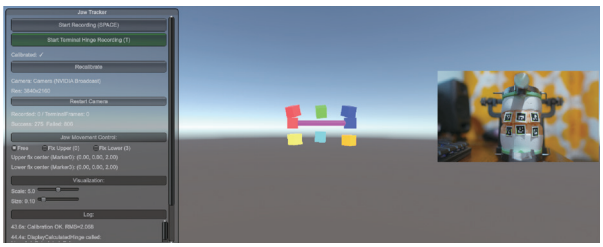

Figure 5: Unity runtime screenshot after camera calibration.

With the camera directed at the marker board, all six ArUco markers (IDs 0–5) are detected. For each marker, a color-coded sphere is rigidly attached to the estimated 6-DoF marker frame, so its position and orientation update in real time, demonstrating stable per-frame tracking. The left panel reports calibration status and runtime diagnostics, and the right inset shows the live camera feed.

3.4 Rigid Fusion via the Kabsch Method

3.4.1 Formulation of the Weighted Kabsch Method

The core technique of this system, the weighted Kabsch method, optimally aligns up to 12 corner 3D points obtained from multiple markers with a single rigid transformation. The minimization problem is formulated as follows:

$$\min_{R,t} \sum_{j=1}^{N} w_j \, \|Y_j - (RX_j + t)\|^2, \quad R^\top R = I, \ \det(R) = 1 \tag{1}$$

X_j: Template point in the jaw-local coordinate system (fixed in the initial frame)
Y_j: Observed point in the current frame (camera coordinates; metric units [m])
j >= 0: Confidence weight for point j (larger = more trusted)
N: Number of points used (up to 12 = 3 markers x 4 corners)
R: 3x3 rotation matrix (orthonormal; right-handed; det(R) = +1)
t: 3x1 translation vector

In short: Find R and t that move X_j to best match Y_j, minimizing the weighted sum of squared errors and using w_j as confidence weights.

3.4.2 Confidence Weight Computation

For each corner point, the weight w_j is computed dynamically based on the reprojection error:

$$w_j = \exp\left(-\frac{e_j^2}{2\kappa^2}\right) \tag{2}$$

where:
- e_j : reprojection error of corner j (in pixels)
- κ : scale parameter (default: 3 pixels)

This weighting function automatically down-weights corners with large reprojection errors (e.g., low detection accuracy or partial occlusion).

3.4.3 Algorithm for Solving the Kabsch Method

In implementation, the optimal rigid transformation is obtained by the following steps:

(a) Compute the weighted centroids:

$$\bar{\mathbf{X}} = \frac{\sum_{j=1}^{N} w_j \mathbf{X}_j}{\sum_{j=1}^{N} w_j}, \qquad \bar{\mathbf{Y}} = \frac{\sum_{j=1}^{N} w_j \mathbf{Y}_j}{\sum_{j=1}^{N} w_j} \tag{3}$$

Compute the weighted centroids of the template and observed points to remove global translation and reflect each point's confidence.

(b) Build the cross-covariance matrix:

$$\mathbf{H} = \sum_{j=1}^{N} w_j \, (\mathbf{X}_j - \bar{\mathbf{X}}) \, (\mathbf{Y}_j - \bar{\mathbf{Y}})^\top \tag{4}$$

Center both point sets and build the weighted cross-covariance matrix, which captures how their coordinates vary together under the optimal rotation.

(c) Perform singular value decomposition (SVD):

$$\mathbf{H} = \mathbf{U} \mathbf{\Sigma} \mathbf{V}^\top \tag{5}$$

Perform SVD on the cross-covariance to extract orthonormal bases whose alignment yields the least-squares rotation.

(d) Compute the optimal rotation matrix:

$$\mathbf{R} = \mathbf{V} \, \mathrm{diag}\big(1, \ 1, \ \mathrm{sign}(\det(\mathbf{V}\mathbf{U}^\top))\big) \, \mathbf{U}^\top \tag{6}$$

Construct the rotation, enforcing a proper rotation (detRRR=+1) and rejecting reflections.The determinant term ensures a proper rotation (i.e., no reflection).

(e) Compute the optimal translation vector:

$$\mathbf{t} = \bar{\mathbf{Y}} - \mathbf{R} \, \bar{\mathbf{X}} \tag{7}$$

Recover the translation by aligning centroids—set ttt so the rotated template centroid coincides with the observed centroid.

3.5 Recording Mode and Quality Control

3.5.1 Template Fixation and Monotonic Motion

At the start of recording (key **T**), in the first **complete frame** (all six markers detected), the corner 3D template points Xj for the maxilla and mandible are fixed in each jaw's local coordinate system. This design enables:

(a)Consistent rigid alignment against the same reference in all subsequent frames

(b)Robustness to partial marker dropouts (operates with at least three markers)

(c)Improved measurement repeatability

Motion is restricted to **monotonic opening or closing**. This ensures:

(d)Consistent screw-axis direction

(e)Avoidance of numerical instabilities

(f)Short acquisition time (10–15 seconds)

3.5.2 Quantifying Frame Quality

For each frame, define a quality score

$$\omega_i \in [0, 1] \tag{8}$$

computed from the mean reprojection error:

$$\omega_i = \exp\left(-\frac{\bar{e}_i^2}{2\kappa^2}\right) \tag{9}$$

where $\bar{e}_i$ is the mean reprojection error of all corners in frame i.

3.6 Screw-Axis Estimation Algorithm

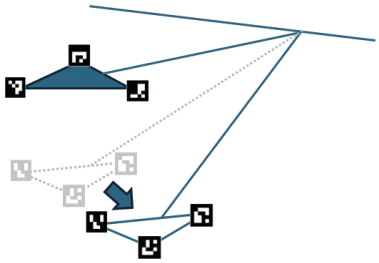

Figure6: Schematic of screw-axis estimation from fused upper/ lower corner sets (12 + 12 points)

Corner points (up to 12 per jaw) are rigidly fused (Kabsch) into maxillary and mandibular frames. The frames' relative motion yields per-frame axis directions, which are sign-aligned and weighted by rotation magnitude and frame quality to obtain a single global axis. A representative point on the axis is then estimated, and left/right hinge points are placed using the known intercondylar distance. Monotonic opening/closing preserves a consistent axis orientation.

3.6.1 Relative Transform and Axis–Angle Representation

For each frame iii, the maxilla–mandible relative transform is

$$\mathbf{T}_{\mathrm{rel},\,i} = \mathbf{U}_i^{-1}\,\mathbf{L}_i \tag{10}$$

Where U_i is the maxillary frame and L_i is the mandibular frame (both are outputs of the Kabsch step).

Decompose each relative transform into an axis–angle representation:

(a)Axis direction: w_i (unit vector)

(b)Rotation angle: θ_i (radians)

3.6.2 Weighted Axis Direction Estimation

The axis direction is estimated by a weighted average using the rotation angle and frame quality:

Figure 6. Weighted estimation of the axis direction

$$\hat{w} = \frac{\left|\sum_{i=1}^{M} s_i\,\theta_i\,\omega_i\,w_i\right|}{\sum_{i=1}^{M} s_i\,\theta_i\,\omega_i\,w_i} \tag{11}$$

where

(a) $s_i \in \{+1, -1\}$: sign-consistency factor (aligned with the frame having the maximum rotation angle),

(b)θ_i: rotation angle of frame iii,

(c)ω_i: frame quality,

(d)M: total number of frames.

This weighting scheme:

(e)increases the contribution of frames with larger rotation angles,

(f)automatically suppresses low-quality frames,

(g)improves robustness against outliers.

3.6.3 Least-Squares Estimation of a Representative Point on the Axis

Linear approximation of the screw motion:

$$\mathbf{t}_{\mathrm{rel},\,i} = \left(\mathbf{I} - \mathbf{R}_{\mathrm{rel},\,i}\right)\mathbf{q} \tag{12}$$

where q is a representative point on the axis.

Stack all frames with quality weights:

$$\begin{bmatrix} \sqrt{\omega_1}\,(\mathbf{I} - \mathbf{R}_{\mathrm{rel},\,1}) \\ \sqrt{\omega_2}\,(\mathbf{I} - \mathbf{R}_{\mathrm{rel},\,2}) \\ \vdots \\ \sqrt{\omega_M}\,(\mathbf{I} - \mathbf{R}_{\mathrm{rel},\,M}) \end{bmatrix} \mathbf{q} = \begin{bmatrix} \sqrt{\omega_1}\,\mathbf{t}_{\mathrm{rel},\,1} \\ \sqrt{\omega_2}\,\mathbf{t}_{\mathrm{rel},\,2} \\ \vdots \\ \sqrt{\omega_M}\,\mathbf{t}_{\mathrm{rel},\,M} \end{bmatrix} \tag{13}$$

Solve this overdetermined system via SVD to obtain the least-squares solution $\hat{q}$.

3.7 Placement and Validation of Hinge Points

3.7.1 Determination of Left and Right Hinge Points

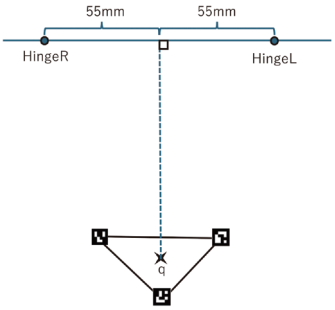

Figure 7: Hinge-center selection and placement of left/right hinge points

After the global screw axis is estimated, the mandibular centroid is projected perpendicularly onto the axis. The foot of this perpendicular defines the hinge center. The left (L) and right (R) hinge points are then placed by translating from the hinge center along the axis in opposite directions by half of the known intercondylar distance. The mandibular markers illustrate how the centroid is obtained; the axis orientation is used to assign L/R consistently in the mandibular local coordinate system.

The representative point $\hat{q}$

on the screw axis is determined by least squares as the point on the axis that minimizes the perpendicular distance from the mandibular centroid. This point corresponds to the midpoint between the condyles and provides the most stable estimate of the mandibular center of rotation.

From $\hat{q}$, place the left and right hinge points using the known intercondylar distance W (default: 110 mm):

$$\mathbf{q}_R = \hat{\mathbf{q}} - \frac{w}{2}\,\hat{\mathbf{w}}, \qquad \mathbf{q}_L = \hat{\mathbf{q}} + \frac{w}{2}\,\hat{\mathbf{w}} \tag{14}$$

where q_R is the right hinge point (patient's right condyle), q_L is the left hinge point (patient's left condyle), and $\hat{w}$ is the normalized screw-axis direction vector.

The left/right assignment is decided by the sign of $\hat{w} \cdot x$ in the mandibular local coordinate system: a positive value indicates the axis points toward the patient's left; a negative value indicates it points toward the patient's right.

3.7.2 Saving to the Local Coordinate System

Store the estimated screw axis in the mandibular local coordinate system as follows:

(a)Convert the hinge-axis center point from world/camera space into mandibular local space by applying the inverse of the current mandibular pose (transform).

(b)Convert the hinge-axis direction vector in the same way, then normalize it to maintain unit length.

Persist the screw axis by saving these two local quantities—the local center and the local direction—as the canonical local representation used for subsequent reconstruction.

With the axis saved in mandibular local coordinates, you can reconstruct its world/camera-space position and direction at any later frame by applying the current mandibular pose to the stored local center and direction. This allows robust recovery of the axis even when some markers are temporarily not detected, since reconstruction depends only on the current mandibular pose and the saved local representation.

3.8 Diagnostic Check of Arc Validity

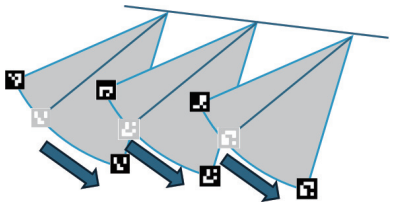

Figure 8: Pairwise arc-based validation of the screw axis.

During the validation step, each maxilla–mandible marker pair is examined under a pure hinge assumption. Their trajectories are projected onto the plane orthogonal to the candidate axis, where they should form circular arcs with a common center aligned with the axis. The figure illustrates successive frames for multiple pairs (arrows indicate motion). Consistency of arc centers and radii across pairs supports the estimated screw axis, whereas systematic offsets reveal axis misalignment. These discrepancies are summarized by the validity metrics reported in Section 3.8.2.

3.8.1 Circle Fitting by the Kåsa Method

Project the trajectories of the mandibular markers (IDs 3, 4, 5) onto the plane orthogonal to the estimated axis, and perform **algebraic circle fitting**:

$$\min_{a,b,c} \sum_{i=1}^{n} \left(x_i^2 + y_i^2 + a\,x_i + b\,y_i + c\right)^2 \tag{15}$$

Circle parameters:

(a)Center: $(-a/2, -b/2)$

(b)Radius:

$$\sqrt{\frac{a^2}{4} + \frac{b^2}{4} - c} \tag{16}$$

(Also known as the Kasa method.)

3.8.2 Validity Metrics

Quantify the validity of the estimated axis using the following metrics:

(a)**RMS residual**: root-mean-square error from the fitted circle.

(b)**Coefficient of variation (CV) of radius**: temporal consistency of the radius.

(c)**Axis distance**: distance from the circle center to the estimated axis (ideally 0).

(d)**Plane thickness**: standard deviation of axial offsets before projection.

Additionally, fit a **total least squares (TLS)** line to **the three circle centers** (from markers 3, 4, 5) and verify that its direction agrees with $\hat{\mathcal{W}}$ to further validate the estimate.

3.9 Separation of Numerical Computation and Visualization

This system strictly separates **numerical computation (Math)** from **visualization (Render)**:

Numerics layer (Math):

(a)All geometric computations use metric units [m][m][m].

(b)Processing is unified in the **camera coordinate system**.

(c)Floating-point precision is utilized to the fullest.

Visualization layer (Render):

(d)Apply a **display scale** (VIS_SCALE).

(e)Add a **Z-axis offset** for visibility.

(f)Place visual elements as **children of the camera**.

The unidirectional data flow (Math → Render) ensures that:

(g)Numerical consistency and reproducibility are ensured.

(h)Changes to visualization parameters do **not** affect computed results.

(i)Debugging and validation are simplified.

4. Results

4.1 System Functionality Check

We implemented the system in a Unity 6 environment and measured jaw motion using a dental arch model mounted on a Handy 2A average-value articulator. With a Sony ZV-E10M2 camera (via NVIDIA Broadcast at 4K@30fps), we confirmed real-time detection and tracking of six ArUco markers (DICT_4X4_50; side length 10 mm).

4.2 Calibration and Pose-Estimation Stability

Camera calibration using a checkerboard (10x7 inner corners, 25 mm squares) was performed by automatically collecting 15 frames. In per-marker solvePnP pose estimation, outliers were rejected when the inter-frame orientation change exceeded 90 degrees, and smoothing with an exponential moving average (alpha = 0.2) was confirmed to be effective.

4.3 Jaw-Frame Estimation via Kabsch Fusion

Using 3D–3D rigid least squares (the Kabsch method), we implemented generation of a unified jaw frame (maxilla and mandible) from up to 12 corner points across multiple markers. Reprojection-error weighting (kappa = 3 px) was incorporated, and jaw-frame estimation was confirmed to execute in frames where all six markers were detected.

4.4 Terminal Hinge-Axis Estimation

From recordings of monotonic opening/closing motion, we implemented the following pipeline:

(a)Compute the axis direction ($\hat{\mathcal{W}}$) from relative transforms using a rotation-angle × quality weighted average.
(b)Estimate a representative point q on the axis via a linearized screw-motion model using least squares.
(c)Place left/right hinge points using the known intercondylar distance (110 mm).
These processes run continuously, and the estimated hinge axis is visualized in 3D within the Unity scene.

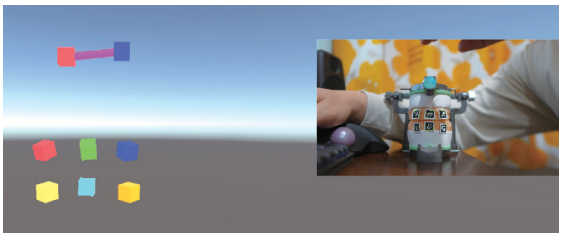

Figure 9: Real-time visualization of the estimated screw axis and left/right hinge points

Unity runtime view showing the screw axis and the right/left hinge points rendered together with the marker-attached reference cubes. As the ArUco marker poses change, the estimated axis and hinge points update frame-by-frame, demonstrating stable closed-loop tracking. The inset shows the live camera feed; the colored cubes correspond to the individual marker frames used for fusion.

4.5 Arc-Validity Diagnostics

We implemented a diagnostic that projects the trajectories of mandibular markers (IDs 3, 4, 5) onto the plane orthogonal to the axis and performs circle fitting using the Kasa method. The following metrics were computed and displayed:
(a)RMS residual of the circle fit
(b)Coefficient of variation (CV) of the radius
(c)Distance from the fitted circle center to the estimated axis
(d)Variance of axial offsets (thickness) prior to projection

5. Novelty of This Study

5.1 Technical Novelty

Realization with a single monocular camera.
Unlike existing commercial systems (e.g., Zebris, Modjaw) that require multiple cameras or dedicated hardware, our method achieves jaw-motion measurement and hinge-axis estimation using only a general-purpose monocular camera and six ArUco markers. This can potentially reduce equipment cost to less than one-tenth of existing solutions.
Applying Kabsch-based rigid fusion to jaw-motion estimation.
We estimate a jaw frame by simultaneously aligning up to 12 marker corner points with a single rigid transform using the Kabsch method. This differs from simple averaging of per-marker poses or three-point methods and provides a statistically optimal rigid pose. Reprojection-error-based weighting improves robustness to noise and outliers.
Weighted least-squares screw-axis estimation.
By considering both frame quality (ωi) and rotation angle (Θi), the proposed weighted estimation naturally down-weights low-quality or small-angle frames. Under the monotonic-motion constraint, this enables axis estimation with a limited number of frames.

5.2 Implementation Novelty

Integrated implementation in Unity.
Combining Unity 6 with OpenCV for Unity integrates real-time processing and 3D visualization. A one-way separation of numerical computation (Math) and visualization (Render) achieves both reproducibility of calculations and flexibility in display.
Use of an open-source technology stack.
By combining accessible components—ArUco (open-source marker system), OpenCV (image processing), and Unity (3D environment)—the system can be customized and extended by researchers and clinicians.

5.3 Clinical Novelty

Chairside readiness.
The setup is simple, and calibration-to-measurement can be completed quickly. Results can potentially be presented intuitively to patients via 3D visualization.
Design emphasizing reproducibility.
Choices such as monotonic motion, template fixation, and using only the Kabsch output improve measurement reproducibility, enabling longitudinal evaluations and inter-operator comparisons.
Path for incremental accuracy improvements.
The base system can be introduced at low cost, with accuracy improved stepwise by upgrading the camera, increasing the number of markers, or refining algorithms—contrasting with the all-at-once investment required by dedicated systems.

6. Discussion and Future Extensions

Weight design can be extended from a simple reprojection-error function to a hybrid scheme incorporating edge contrast and detector confidence. Introducing explicit outlier removal (thresholding angle/position residuals to exclude bad corners, followed by 1–2 refits) would increase robustness. When environmental changes occur (camera motion or fixture shift), allowing template re-anchoring enables stable long-duration recording. Accurate camera intrinsics (K, D) and marker size are essential; utilities for recalibration and size verification are desirable. The approach remains valid in monocular settings as long as marker dimensions are known and calibration is sufficient. By combining batch rigid alignment of corner points via Kabsch and angle × quality weighting for screw-axis estimation, stable recovery of jaw frames and the principal rotation axis is possible even with a small number of markers. Next steps include comparison against optical references and clinical datasets to quantify error characteristics, improve robustness, and refine UI/UX, as well as exploring integration with intraoral scan data and CBCT scans for enhanced visualization.

References

[1] Schiffman E, Ohrbach R, Truelove E, et al. Diagnostic Criteria for Temporomandibular Disorders (DC/TMD) for Clinical and Research Applications: recommendations of the International RDC/TMD Consortium Network and Orofacial Pain Special Interest Group. *J Oral Facial Pain Headache*. 2014;28(1):6-27.
[2] The Japanese Society for the Temporomandibular Joint. Treatment Guidelines for Patients with Temporomandibular Disorders. Ishiyaku Publishers; 2020 (in Japanese).
[3] Posselt U. Terminal hinge movement of the mandible. J Prosthet Dent. 1957;7(6):787-97.

[4] Kono S. Study of condylar movement corresponding to sagittal mandibular movement: Part 2. Analysis of sagittal movement axis using multi-flash device. J Jpn Prosthodont Soc 1968;12:350–80 (in Japanese).
[5] Mapelli A, Galante D, Lovecchio N, Sforza C, Ferrario VF. Translation and rotation movements of the mandible during mouth opening and closing. Clin Anat. 2009;22(2):203-10.
[6]Rupin M, Jager M, Peyrodie L, Thiong-Ly C, Ingrand P, Le-Dain M, et al. Complete 3D Kinematics Parameters of the Temporo-Mandibular Joints Using in Vivo Data Fusion. IRBM. 2023;44(6):100804.