

Article

Development of a Monocular Camera-Based Sweep Analysis System in Curling

Riku Hara ¹, Shimpei Aihara ^{2,*}  and Takeshi Ito ^{1,*}

¹ School of Information Science and Technology, The University of Electro-Communications, 1-5-1 Chofugaoka, Chofu, Tokyo 182-8585, Japan; r.hara.uec@gmail.com

² Department of Sport Science and Research, Japan Institute of Sports Sciences, 3-15-1 Nishigaoka, Kita-ku, Tokyo 115-0056, Japan

* Correspondence: shimpei.aihara@jpnssport.go.jp (S.A.); ito@cs.uec.ac.jp (T.I.)

Abstract

Curling requires precise sweeping to control both the speed and trajectory of a stone, making quantitative analysis of sweeping behavior an important challenge for performance evaluation and tactical support. However, existing curling studies have primarily focused on stone motion analysis, while automatic estimation of sweep positions and brush orientations from monocular video remains largely unexplored. To address this problem, this study proposes a vision-based system for estimating the position and orientation of curling brushes during sweeping using a single RGB camera. The proposed framework integrates a two-stage YOLOv8-OB detection pipeline for identifying sweeping players and brush regions, ByteTrack-based multi-object tracking for temporal association, and a geometric calibration model that transforms image coordinates into real-world curling-sheet coordinates through nonlinear optimization. The system enables frame-by-frame reconstruction of sweeping trajectories and brush orientations without requiring specialized sensing equipment. Experiments were conducted at a curling facility using synchronized monocular video and a Qualisys Miquis M3 motion-capture system as ground truth. The proposed method achieved position estimation errors of 0.03–0.06 m and orientation estimation errors of 4.80°–5.10°, while maintaining a high detection rate of 0.96–0.99 over a measurement range of 15 m. Additional analyses showed that estimation accuracy was influenced by camera placement, sweep style, and the apparent size of the detected brush region. These results demonstrate that accurate sweep-position and orientation estimation can be achieved using only a monocular camera under realistic curling conditions. The proposed system provides a practical framework for objective sweep analysis and has potential applications in technical training, performance assessment, and tactical support in curling.

Keywords: curling; sweep; computer vision; object detection; multi-object tracking; calibration



Academic Editor: Andrea Prati

Received: 18 June 2026

Revised: 11 July 2026

Accepted: 23 July 2026

Published: 31 July 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and

conditions of the [Creative Commons](https://creativecommons.org/licenses/by/4.0/)

[Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Curling, often described as “chess on ice,” is a sport in which technical precision and tactical reasoning are closely intertwined. The outcome of each shot is shaped by a complex interplay of mechanical control and strategic decision-making. Among the technical elements of the sport, sweeping is particularly important, as it allows players to influence both the trajectory and stopping position of the stone. During sweeping, players rub the ice surface with specialized brush pads—referred to as brushes in this paper—to temporarily reduce friction between the stone and the ice. This reduction in friction

increases the stone's travel distance and reduces its curl, thereby exerting a significant impact on game strategy.

Previous studies have shown that sweeping raises the surface temperature of the ice and reduces its coefficient of friction, providing a quantitative explanation for the increased glide distance of the stone [1]. As competitive standards have advanced, sweeping has shifted from an experience-driven technique to one that is increasingly analyzed and optimized from a scientific and strategic perspective. For athletes and coaches, the ability to record, visualize, and evaluate sweeping motion in an objective and quantitative manner has become valuable for both training and tactical planning. Consequently, there is a growing demand for analytical tools that can accurately capture sweeping behavior under realistic gameplay conditions.

Most prior quantitative analyses of sweeping have relied on physical sensors such as accelerometers and force sensors. These methods enable the measurement of applied pressure, acceleration, and brush vibration, offering insights into the physical characteristics of sweeping motion [2]. However, sweeping effectiveness depends not only on how strongly a player sweeps but also on where and how the sweeping is applied. The brush's position and orientation relative to the stone are especially important because they directly influence heat distribution and ice deformation around the stone's running path, both of which affect stone motion [3]. Therefore, capturing the spatial aspects of sweeping is essential for a more complete understanding of sweeping performance.

Despite their usefulness, sensor-based approaches present practical limitations. Attaching sensors to the player or the brush may interfere with natural gameplay, and obtaining continuous, precise spatial information—such as brush trajectories or brush orientation—over extended sequences remains challenging. For these reasons, non-contact and markerless motion analysis approaches have gained attention. Among them, monocular video-based methods are particularly appealing due to their accessibility, cost-effectiveness, and suitability for both training and competition environments. Recent advances in computer vision and deep learning have made it possible to automatically estimate human and object motion without physical markers. Object detection models have shown strong performance in dynamic sports scenes, enabling automated tracking and quantitative performance assessment. Such approaches have been applied in soccer, tennis, and skiing, among others [4–7]. In curling, Aihara et al. developed an image-based system for detecting and tracking curling stones, highlighting the potential of vision-based analysis for high-precision, markerless measurement [8]. However, research specifically focusing on sweeping analysis—particularly the estimation of brush position and orientation—remains limited.

Estimating brush position and orientation from monocular video involves several technical difficulties that are not present in conventional sports object tracking, such as stone tracking. First, the brush head is a small and slender object that often occupies only a limited number of pixels when captured from a practical viewing distance, making stable detection considerably more difficult than for stones or players. Second, sweeping is a rapid and repetitive motion that introduces motion blur and frequent occlusions by players and stones, causing intermittent detection failures. Third, because the proposed system uses a monocular camera, small image-space localization errors can be amplified when transformed into real-world sheet coordinates.

To address this gap and these technical difficulties, the present study proposes a vision-based analysis system that estimates brush position and orientation from monocular video recordings of sweeping. The system combines three main components: (1) a two-stage detection process based on YOLOv8-OB for identifying sweeping players and brush regions; (2) a tracking framework based on ByteTrack to maintain temporal consistency

and manage occlusion; and (3) a monocular calibration approach that converts image coordinates into real-world curling-sheet coordinates. Using synchronized motion-capture data as ground truth, the system achieved an average positional error of 0.03–0.06 m and an angular error of 4.80°–5.10°, demonstrating accurate performance within a 15 m observation range. By providing objective measurements of sweeping behavior, the proposed system opens new possibilities for performance evaluation, technical training, and tactical analysis in curling. More broadly, this study contributes to sports science by demonstrating how computer vision technologies can link human technique with data-driven decision support.

The main contributions of this study are threefold: (1) a two-stage oriented-bounding-box detection pipeline for detecting small and elongated brush regions during sweeping; (2) a monocular calibration procedure that uses official curling-sheet markings and therefore requires no calibration equipment on the ice; and (3) quantitative validation of monocular brush-position and brush-orientation estimation against a high-precision motion-capture reference under realistic curling conditions.

The remainder of this paper is organized as follows: Section 2 describes the proposed system in detail; Section 3 presents the experimental setup and evaluation procedures; Section 4 discusses the experimental findings; Section 5 outlines potential applications of the system; and Section 6 concludes the paper.

2. Materials and Methods

2.1. System Overview

This study introduces a vision-based analysis system designed to estimate the position and orientation of a curling brush during sweeping using monocular video footage (Figure 1). The system is structured as a three-stage processing pipeline consisting of: (1) sweep detection, (2) sweep tracking, and (3) monocular calibration for coordinate transformation.

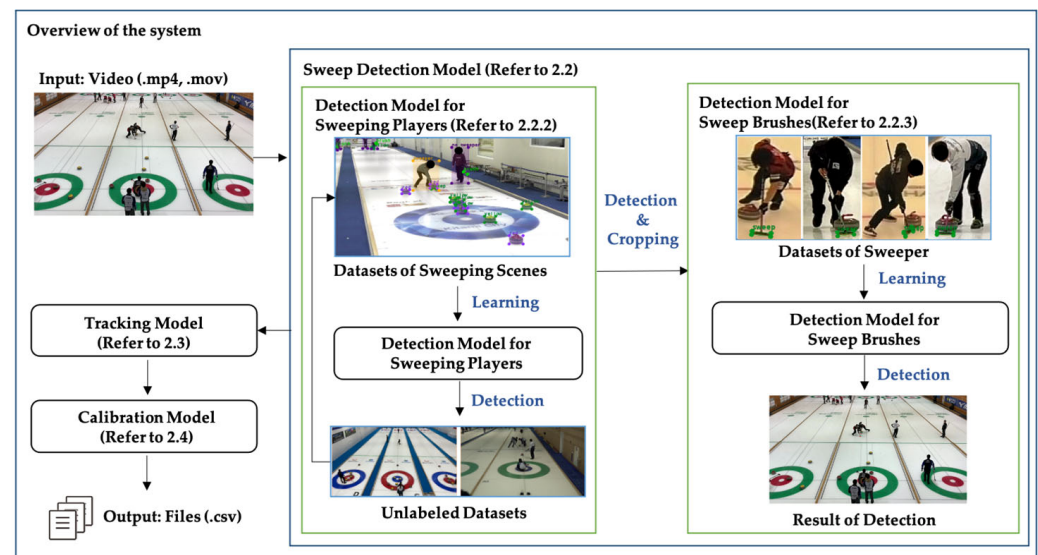


Figure 1. Overview of the sweep analysis system.

In the detection stage, sweeping players and brush regions are identified from each video frame using oriented bounding-box (OBB) object detectors. The tracking module then assigns consistent temporal IDs to brush detections to ensure continuity in the presence of occlusion and motion blur. The calibration module maps the detected brush vertices from image coordinates to real-world sheet coordinates by estimating the camera's intrinsic and extrinsic parameters.

The system was implemented in Python 3.12 using PyTorch 2.11.0 and OpenCV 4.13.0 and was executed on a workstation equipped with an NVIDIA A100 GPU (NVIDIA Corporation, Santa Clara, CA, USA). For each processed video, the system outputs frame-wise brush positions, brush orientations, and reconstructed trajectories as CSV files, enabling quantitative evaluation of sweeping behaviors in both research and coaching applications.

2.2. Sweep Detection Model

2.2.1. Dataset Construction

To develop a reliable sweep detection model, a curling-specific dataset was assembled using video footage obtained from several domestic curling arenas and from publicly available licensed sources. For the footage captured by the authors, consent for use for research purposes was obtained from the individuals appearing in the recordings. The collected footage reflects a variety of match and training environments, including differences in venue layout, illumination, and competitive setting. Recordings were made using a range of consumer and professional devices—such as camcorders, smartphones, and tablets—captured at resolutions from Full HD to 4K. The cameras were positioned at rear-view, side-view, and elevated spectator-stand angles, providing substantial variation in perspective, viewing distance, and background conditions.

Still frames were sampled uniformly from the footage and manually annotated following a standardized labeling protocol. Each frame was labeled across six categories relevant to sweeping analysis: sweeping players, non-sweeping players, brushes in use, brushes not in use, red stones, and yellow stones. For brushes, oriented bounding boxes (OBBs) were used to represent rotation and aspect ratio, which was necessary for later estimation of sweeping direction (Figure 2).



Figure 2. Sample images of the sweeping datasets. Faces and other personally identifiable regions are intentionally blurred to protect privacy.

Annotation was performed by multiple trained annotators who were familiar with the visual characteristics of curling and adhered to a predefined annotation guideline. To reduce inconsistency, all labels were checked through a two-stage quality assurance process, in which a second annotator reviewed and corrected unclear or conflicting annotations. After this procedure, 1847 annotated full-frame images were compiled. The dataset includes diverse sweeping situations, such as changes in brush size due to distance, partial occlusions by other players or stones, and scenes with multiple overlapping objects. This diversity allowed the detection models to be trained under conditions representative of those encountered in actual gameplay.

2.2.2. Detection of Sweeping Players

A multi-class object detection model was developed to recognize the six object categories described earlier. For this purpose, we adopted YOLOv8-OBB [9], an oriented-bounding-box extension of the YOLO (You Only Look Once) family of detectors [10]. The

OBB variant enables the model to localize objects that are rotated or not aligned with the image axes, which is essential for detecting brushes and players under varied viewing conditions [11].

YOLOv8-OBB was adopted because estimating brush orientation is one of the primary objectives of this system. Axis-aligned object detectors can provide approximate localization of the brush region, but they do not directly encode the angular geometry required for brush-orientation estimation. In contrast, OBB detection provides both the position and orientation-related geometry of the detected brush region, making it suitable for the subsequent estimation of sweeping direction.

The detector was initialized with weights pre-trained on the DOTAv1 aerial imagery dataset and subsequently fine-tuned using the curling-specific dataset described in Section 2.2.1. All training images were resized to 1024×1024 pixels. The model was trained for 500 epochs using the Adam optimizer [12] with an initial learning rate of 0.01 and a batch size of 16 on an NVIDIA A100 GPU. Training was conducted using PyTorch 2.11.0 and the Ultralytics YOLOv8 implementation.

To improve robustness in situations where annotated data were limited, a semi-supervised learning procedure was incorporated. Unlabeled images that yielded detections with confidence scores above 0.7 were treated as pseudo-labels and incorporated into the training set. In total, 2347 images—including both manually annotated and pseudo-labeled samples—were used, with 20% reserved for validation. This process helped reduce annotation workload while maintaining consistent detection accuracy for sweeping players. The resulting detector provided reliable full-frame player and brush-region proposals and served as the first stage of the two-stage pipeline introduced in Section 2.2.3.

2.2.3. Detection of Sweeping Brushes

Detecting brushes in curling footage is technically challenging because the brush head often occupies only a very small number of pixels, especially when captured from long distances or oblique viewing angles. Small-object detection is known to be difficult in computer vision due to the limited number of discriminative features and the susceptibility to background noise. Although various methods—such as multi-scale feature extraction and hierarchical detection frameworks—have been proposed to address this issue [13], directly applying a single-stage detector to full-frame curling images resulted in frequent missed detections. In particular, brushes tended to disappear when their apparent size fell below a practical detection threshold.

To address this problem, a two-stage detection pipeline was designed to restrict inference to localized regions of interest. In the first stage, the multi-class detector described in Section 2.2.2 identifies sweeping players in the full-frame image. In the second stage, image patches centered on the detected sweeping players are cropped and passed to a dedicated brush-specific detector. This hierarchical design substantially increases the relative resolution of the brush region and suppresses background clutter, both of which contribute to improved detection sensitivity.

The second-stage detector was implemented using YOLOv8-OBB and was trained on 1398 cropped images collected around the sweeping-player regions. All crops were resized to 640×640 pixels, and the dataset was split in an 8:2 train-validation ratio. The brush detector was trained independently from the player detector but used the same hyperparameters for consistency. Throughout annotation and inference, oriented bounding boxes (OBBs) were used to encode the rotation of the brush head, allowing the detector to output both positional and directional information required for downstream orientation estimation.

2.3. Sweep Tracking Model

In realistic curling matches, brushes frequently become partially or fully occluded by players or stones, and rapid sweeping motion often introduces motion blur that reduces detection reliability. These effects lead to intermittent missed detections, which in turn disrupt the temporal continuity required for accurately reconstructing sweeping trajectories. To mitigate these issues, a multi-object tracking (MOT) framework was integrated into the proposed system to maintain stable brush trajectories across frames.

Among existing MOT algorithms, ByteTrack [14] was selected because of its strong association performance in scenes with fast motion and frequent occlusion. Unlike trackers that rely solely on high-confidence detections, ByteTrack incorporates both high- and low-confidence bounding boxes during data association, enabling trajectories to remain connected even when the brush briefly becomes blurred or partially hidden. After each brush detection is assigned a unique temporal ID, associations across adjacent frames are computed using the intersection-over-union (IoU) of their oriented bounding boxes (OBBs), along with motion similarity and detection confidence. This procedure provides consistent ID tracking even during short-term detection failures.

Appearance-based trackers such as DeepSORT [15] are effective when reliable appearance features are available; however, in the present task, the brush head often occupies only a small number of pixels, and its appearance can change substantially depending on viewpoint, blur, and occlusion. Therefore, a detection-confidence- and motion-association-based tracking strategy was considered more appropriate for maintaining brush trajectories in the proposed system.

The output of this module consists of temporally coherent sequences of brush positions and orientations, each linked to a stable track ID. These continuous trajectories are then used as input for the calibration and coordinate transformation described in the next section. By incorporating ByteTrack into the pipeline, the system achieved reliable tracking performance under challenging on-ice conditions, providing high-quality temporal data essential for quantitative sweeping analysis.

A systematic comparison with alternative object detectors and multi-object tracking algorithms was beyond the scope of the present study. Future work should evaluate whether other OBB detectors, segmentation-based methods, or appearance-assisted trackers can further improve robustness under severe occlusion and multi-player sweeping conditions.

2.4. Calibration Model

To convert detected brush positions from image coordinates to physical locations on the curling sheet, a dedicated monocular camera calibration procedure was developed. In real match environments, placing calibration boards or checker patterns on the ice is not feasible due to gameplay constraints and broadcasting regulations. Instead, the proposed method uses permanent sheet markings—such as house rings, hog lines, and tee lines—as natural calibration targets for establishing 2D–3D point correspondences (Figure 3). Because these markings are defined with fixed dimensions in official curling sheet specifications, they provide reliable geometric references without requiring any additional equipment.

A world coordinate system was defined with its origin at the center of the house, the x-axis aligned with the tee line, and the y-axis aligned with the center line. For calibration, several landmark points were manually selected on the sheet markings appearing in the image, and their corresponding real-world coordinates were assigned according to the official dimensions. At least four or more such 2D–3D correspondences were required to obtain a valid solution for the camera pose. Using these point pairs, the camera's extrinsic parameters (rotation and translation) were estimated through nonlinear least-squares optimization using the Levenberg–Marquardt algorithm [16], which minimizes the reprojection

error between observed and projected points. Because broadcast and publicly recorded videos often do not provide intrinsic metadata, the focal length was estimated separately using a one-dimensional optimization procedure based on hill-climbing search [17]. The algorithm iteratively explored candidate focal-length values and selected the one that minimized the same reprojection-error objective. This one-dimensional search strategy was effective because only a single intrinsic parameter required estimation; the principal point was fixed at the image center, and lens-distortion parameters were taken from camera metadata when available.

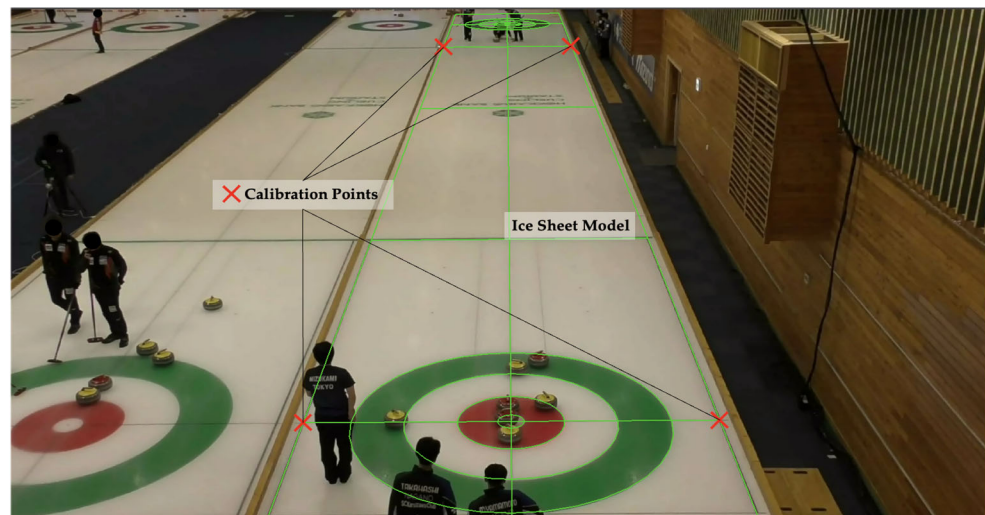


Figure 3. Calibration on ice sheet.

This calibration approach is applicable to various recording settings, including fixed broadcast cameras and elevated spectator-stand viewpoints. Once calibration was completed, the intrinsic and extrinsic parameters were used to convert the selected OBB endpoints of each brush detection into real-world sheet coordinates. This enabled frame-by-frame reconstruction of sweeping trajectories for subsequent quantitative analysis.

2.5. Sweep Orientation Estimation

To quantify the orientation of the sweeping motion, the proposed system uses the geometric properties of the oriented bounding boxes (OBBs) predicted for each brush. After the camera calibration step, each brush OBB is represented in the calibrated image space, allowing its long edges to serve as reliable indicators of brush direction. In this study, the long edge of the OBB was used to compute the sweep orientation. Since an OBB has two symmetric long edges, either theoretically provides the same directional information. To ensure consistency across frames, the system selects the long edge whose endpoints yield the most stable and continuous direction vector (Figure 4). This helps reduce the effects of minor detection noise and prevents angle discontinuities. The sweep orientation θ is defined as the angle between this direction vector and the horizontal axis of the curling sheet. Given two endpoints of the selected long edge, (x_1, y_1) and (x_2, y_2) , the angle is computed using the two-argument arctangent function:

$$\theta = \text{atan2}(y_2 - y_1, x_2 - x_1).$$

The result is converted to degrees and normalized to the range $[0^\circ, 180^\circ]$ to account for the symmetry of brush orientation. This method provides a rotation-invariant and physically meaningful measure of sweep direction that is robust to viewpoint differences and detection variability, enabling reliable quantitative analysis of sweeping motion.

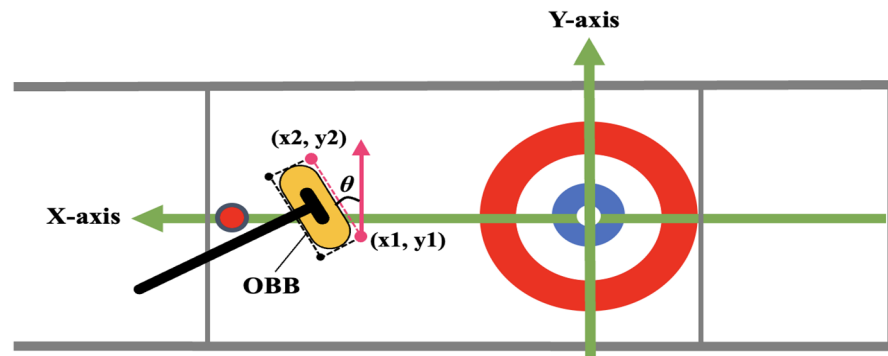


Figure 4. Orientation of the sweeping brush.

3. Evaluations

3.1. Evaluation of Position

3.1.1. Data Acquisition Experiment for Position Evaluation

The evaluation of sweep position accuracy was conducted at the Allographics Kitami Curling Hall. The experiment combined the sweep analysis system developed in this study with a high-precision motion capture system capable of three-dimensional motion measurement. Ground-truth measurements were obtained using the Miquis M3 motion capture system (Qualisys AB, Gothenburg, Sweden). The Miquis M3 system provides a frame rate of 100 Hz and sub-centimeter measurement accuracy (position error < 0.01 m) and is widely used in sports biomechanics research.

In this experiment, as illustrated in Figure 5, the motion capture system was permanently installed on the ceiling of the curling facility so as to cover the entire house area of the curling sheet. Multiple Miquis M3 cameras were used to record the three-dimensional position of the brush. Eight reflective markers were attached to the brush, enabling the center position of the brush pad to be computed for each frame. These coordinates were treated as the ground-truth brush positions throughout the evaluation.



Figure 5. Scene of the experiment (left). Motion capture camera and reflective marker settings (right).

For the video-based estimation, a monocular RGB camera capable of 4K recording (iPhone 15 Pro, Apple Inc., Cupertino, CA, USA) was used. The camera was mounted on a tripod on the center line at a distance of 15.0 m from the end of the sheet and at a height of 2.0 m. The video was recorded at a resolution of 3840×2160 pixels, a frame rate of 60 fps, and a shutter speed of $1/500$ s, allowing clear capture of rapid sweeping motions with minimal motion blur.

Three camera placement conditions were considered: (1) on the center line (lateral offset 0 m), (2) laterally shifted by 1.5 m from the center line, and (3) laterally shifted by 3.0 m from the center line. In all cases, the camera height and the distance from the sheet

edge were kept identical. The camera orientation was adjusted so that the entire sheet—particularly the house area—was contained within the field of view. This design allowed us to examine how lateral viewpoint displacement influences sweep position estimation accuracy. None of the participants, brushes, recording conditions, or camera angles used in this evaluation were part of the training dataset of the detection models.

During the experiment, one participant performed three representative sweeping patterns shown in Figure 6: longitudinal sweeping (aligned with the stone's traveling direction), lateral sweeping, and diagonal sweeping. Each sweeping pattern was repeated multiple times. The participant was instructed to perform sweeping motions within a predefined region around the house so that the brush repeatedly moved back and forth within the measurement area.

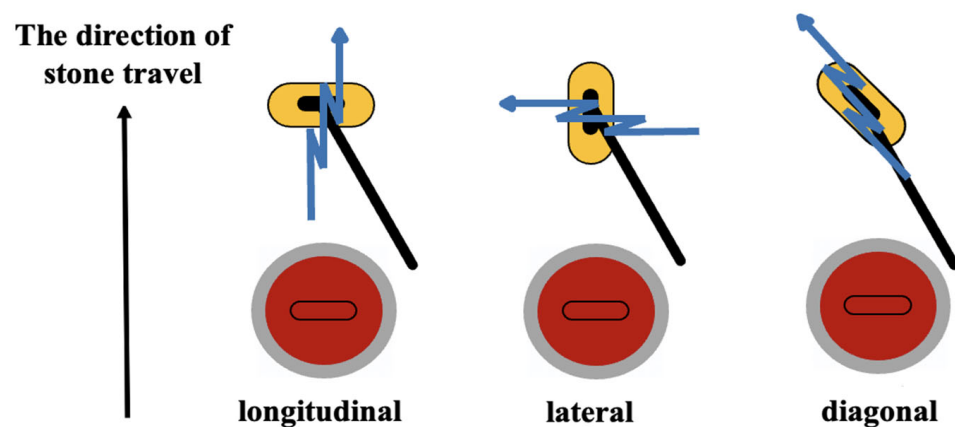


Figure 6. Example of three sweep styles.

The motion capture data and the monocular video frames were synchronized using an initial synchronization motion—where the brush position was aligned in both recordings. Frame-based alignment was then performed to establish correspondence between the two systems. For each frame, the positional error was computed as the difference between the brush coordinates obtained from the motion capture system (ground truth) and those estimated by the proposed system. The estimation accuracy was evaluated using the mean absolute error (MAE), standard deviation (SD), and root mean square error (RMSE). In addition, the detection rate (Rate) was defined as the proportion of frames in which the proposed system successfully produced valid detection and tracking results among all frames where the brush was observed by the motion capture system.

This study was approved by the Ethics Review Committee of The University of Electro-Communications (Approval No. H25010). Informed consent was obtained from the participant before data collection.

3.1.2. Position Estimation Results

To establish a basis for evaluating the proposed method, Tables 1–3 first present the sweep position estimation errors of the conventional single-stage method, which detects brushes directly from the full-frame image without the player-region cropping described in Section 2.2.3. Tables 4–6 then report the results of the proposed two-stage method, and the comparison between the two clarifies how the two-stage design contributes to detection stability and estimation accuracy.

Table 1. Sweep Position Estimation Errors Obtained Using the Conventional Single-Stage Method (Lateral Offset: 0 m from the Center Line).

Sweep Styles	MAE [m]	SD [m]	RMSE [m]	Rate
Longitudinal	0.03	0.02	0.04	0.84
Lateral	0.04	0.02	0.04	0.88
Diagonal	0.03	0.02	0.04	0.83
Overall	0.03	0.02	0.04	0.85

Table 2. Sweep Position Estimation Errors Obtained Using the Conventional Single-Stage Method (Lateral Offset: 1.5 m from the Center Line).

Sweep Styles	MAE [m]	SD [m]	RMSE [m]	Rate
Longitudinal	0.04	0.03	0.05	0.93
Lateral	0.04	0.02	0.04	0.91
Diagonal	0.07	0.03	0.08	0.86
Overall	0.05	0.03	0.06	0.90

Table 3. Sweep Position Estimation Errors Obtained Using the Conventional Single-Stage Method (Lateral Offset: 3.0 m from the Center Line).

Sweep Styles	MAE [m]	SD [m]	RMSE [m]	Rate
Longitudinal	0.03	0.02	0.04	0.93
Lateral	0.04	0.02	0.04	0.89
Diagonal	0.08	0.05	0.09	0.86
Overall	0.05	0.03	0.06	0.89

Table 4. Sweep Position Estimation Errors Obtained Using the Proposed Two-Stage Method (Lateral Offset: 0 m from the Center Line).

Sweep Styles	MAE [m]	SD [m]	RMSE [m]	Rate
Longitudinal	0.03	0.02	0.04	0.99
Lateral	0.03	0.02	0.04	0.97
Diagonal	0.03	0.02	0.04	0.99
Overall	0.03	0.02	0.04	0.98

Table 5. Sweep Position Estimation Errors Obtained Using the Proposed Two-Stage Method (Lateral Offset: 1.5 m from the Center Line).

Sweep Styles	MAE [m]	SD [m]	RMSE [m]	Rate
Longitudinal	0.03	0.02	0.04	0.98
Lateral	0.04	0.02	0.04	0.97
Diagonal	0.08	0.05	0.09	0.96
Overall	0.05	0.03	0.06	0.97

Table 6. Sweep Position Estimation Errors Obtained Using the Proposed Two-Stage Method (Lateral Offset: 3.0 m from the Center Line).

Sweep Styles	MAE [m]	SD [m]	RMSE [m]	Rate
Longitudinal	0.03	0.02	0.04	0.97
Lateral	0.04	0.02	0.04	0.96
Diagonal	0.11	0.06	0.13	0.98
Overall	0.06	0.03	0.07	0.97

Tables 4–6 present the results of the proposed two-stage method. Under the center-line configuration (Table 4), the three sweep styles exhibited almost identical accuracy. The mean absolute error (MAE), standard deviation (SD), and root mean square error (RMSE) were 0.03 m, 0.02 m, and 0.04 m, respectively, for all sweep styles. The overall average was $MAE = 0.03$ m and $RMSE = 0.04$ m. The detection rate was also high, reaching 0.98 across 432 frames.

When the camera was laterally shifted by 1.5 m (Table 5), the estimation accuracy slightly deteriorated. The average MAE increased to 0.05 m and the RMSE to 0.06 m. For longitudinal and lateral sweeping, the MAEs remained relatively small at 0.03 m and 0.04 m, respectively. In contrast, diagonal sweeping produced larger errors, with an MAE of 0.08 m and an RMSE of 0.09 m. Nevertheless, the detection rate remained high for all sweep styles, exceeding 0.96 across 463 frames.

When the lateral offset was further increased to 3.0 m (Table 6), the overall MAE and RMSE increased to 0.06 m and 0.07 m, respectively. While the errors for longitudinal and lateral sweeping remained nearly unchanged, the error for diagonal sweeping increased substantially, reaching an MAE of 0.11 m and an RMSE of 0.13 m. This trend indicates that the proposed method becomes increasingly sensitive to camera offset when estimating diagonal sweeping motions. Despite the degradation in estimation accuracy, the detection rate remained high, with an overall average of 0.97 across 483 frames.

Compared with the conventional single-stage method (Tables 1–3), the proposed two-stage method achieved comparable positional accuracy but a consistently higher detection rate across all camera offsets (overall 0.85–0.90 $\rightarrow$ 0.97–0.98), indicating that the main benefit of the two-stage design lies in stably capturing small brush regions rather than in reducing the positional error itself.

3.2. Evaluation of Sweep Orientation

3.2.1. Data Acquisition Experiment for Orientation Evaluation

The experimental environment used for the orientation evaluation was identical to that described in Section 3.1.1. The same monocular video recordings and motion capture measurements were used to ensure consistent experimental conditions.

Ground-truth brush orientation was derived from the motion capture data using the configuration of the reflective markers attached to each brush. The orientation of the brush pad was calculated for each frame and used as the reference value.

For each frame, the orientation estimated by the proposed system was compared with the corresponding ground-truth orientation obtained from the motion capture system. The angular error was defined as the absolute difference between the estimated orientation and the ground-truth orientation.

The estimation accuracy was evaluated using the mean absolute error (MAE), standard deviation (SD), and root mean square error (RMSE). Similar to the position evaluation, the analysis was conducted separately for three representative sweep styles: longitudinal, lateral, and diagonal.

3.2.2. Orientation Estimation Results

Similarly, Tables 7–9 present the sweep orientation estimation errors of the conventional single-stage method, with the corresponding results of the proposed two-stage method reported in Tables 10–12 for comparison.

Table 7. Sweep Orientation Estimation Errors Obtained Using the Conventional Single-Stage Method (Lateral Offset: 0 m from the Center Line).

Sweep Styles	MAE [°]	SD [°]	RMSE [°]	Rate
Longitudinal	2.99	2.49	3.89	0.84
Lateral	7.14	7.62	10.44	0.88
Diagonal	11.28	9.11	14.50	0.83
Overall	7.23	6.41	9.66	0.85

Table 8. Sweep Orientation Estimation Errors Obtained Using the Conventional Single-Stage Method (Lateral Offset: 1.5 m from the Center Line).

Sweep Styles	MAE [°]	SD [°]	RMSE [°]	Rate
Longitudinal	2.54	2.23	3.38	0.93
Lateral	6.03	4.71	7.65	0.91
Diagonal	9.02	6.43	11.08	0.86
Overall	5.86	4.46	7.36	0.90

Table 9. Sweep Orientation Estimation Errors Obtained Using the Conventional Single-Stage Method (Lateral Offset: 3.0 m from the Center Line).

Sweep Styles	MAE [°]	SD [°]	RMSE [°]	Rate
Longitudinal	3.34	2.87	4.40	0.93
Lateral	5.27	3.97	6.60	0.89
Diagonal	8.26	7.46	11.13	0.86
Overall	5.62	4.77	7.37	0.89

Table 10. Sweep Orientation Estimation Errors Obtained Using the Proposed Two-Stage Method (Lateral Offset: 0 m from the Center Line).

Sweep Styles	MAE [°]	SD [°]	RMSE [°]	Rate
Longitudinal	3.50	2.75	4.45	0.99
Lateral	4.52	3.61	5.78	0.97
Diagonal	7.27	5.27	8.98	0.99
Overall	5.10	3.88	6.41	0.98

Table 11. Sweep Orientation Estimation Errors Obtained Using the Proposed Two-Stage Method (Lateral Offset: 1.5 m from the Center Line).

Sweep Styles	MAE [°]	SD [°]	RMSE [°]	Rate
Longitudinal	2.48	2.33	3.40	0.98
Lateral	5.70	3.85	6.89	0.97
Diagonal	6.23	5.54	8.34	0.96
Overall	4.80	3.91	6.19	0.97

Table 12. Sweep Orientation Estimation Errors Obtained Using the Proposed Two-Stage Method (Lateral Offset: 3.0 m from the Center Line).

Sweep Styles	MAE [°]	SD [°]	RMSE [°]	Rate
Longitudinal	2.96	2.66	3.98	0.97
Lateral	4.78	3.56	5.96	0.96
Diagonal	6.90	7.05	9.29	0.98
Overall	4.88	3.91	6.58	0.97

Tables 10–12 present the orientation estimation errors of the sweeping brush when the camera was positioned on the center line (lateral offset: 0 m) and when it was shifted laterally by 1.5 m and 3.0 m. Under the center-line configuration (Table 10), the overall average angular error was $MAE = 5.10^\circ$, $SD = 3.88^\circ$, and $RMSE = 6.41^\circ$, with a detection rate of 0.98. Among the sweep styles, longitudinal sweeping produced the smallest error ($MAE = 3.50^\circ$), followed by lateral sweeping ($MAE = 4.52^\circ$), while diagonal sweeping showed the largest error ($MAE = 7.27^\circ$).

When the camera was laterally shifted by 1.5 m (Table 11), the overall accuracy remained comparable to the center-line condition, with $MAE = 4.80^\circ$ and $RMSE = 6.19^\circ$. Longitudinal sweeping exhibited an even smaller error ($MAE = 2.48^\circ$), whereas lateral and diagonal sweeping resulted in MAE values of 5.70° and 6.23° , respectively.

When the lateral offset was further increased to 3.0 m (Table 12), the overall accuracy remained stable, with $MAE = 4.88^\circ$ and $RMSE = 6.58^\circ$, indicating no substantial degradation in orientation estimation performance. The MAE values for longitudinal and lateral sweeping were 2.96° and 4.78° , respectively, while diagonal sweeping produced $MAE = 6.90^\circ$. Across all three camera configurations, a consistent trend was observed: longitudinal sweeping produced the smallest errors, whereas diagonal sweeping resulted in the largest errors. In all cases, the detection rate remained high, exceeding 0.96.

Compared with the conventional single-stage method (Tables 7–9), the proposed two-stage method reduced the overall orientation error across all camera offsets (overall MAE $7.23^\circ/5.86^\circ/5.62^\circ \rightarrow 5.10^\circ/4.80^\circ/4.88^\circ$), in addition to the improvement in detection rate. The reduction was most pronounced for diagonal sweeping, where the single-stage method produced large errors due to frequent missed detections of small brush regions.

3.3. Computational Performance

To assess the computational feasibility of the proposed system, we measured the overall processing speed of the current implementation using the existing evaluation videos. The measurements were conducted in the same workstation environment as that used in the accuracy evaluation. The workstation was equipped with an NVIDIA A100 GPU, an Intel(R) Xeon(R) CPU @ 2.20 GHz (Intel Corporation, Santa Clara, CA, USA), and 160 GB of RAM. The implementation was based on Python 3.12, PyTorch 2.11.0, OpenCV 4.13.0, and the Ultralytics YOLOv8 implementation. The input videos were recorded at a resolution of 3840×2160 pixels and 60 fps.

The average processing time per frame was calculated by dividing the total elapsed processing time by the number of analyzed frames, and the corresponding throughput was calculated as frames per second. Initialization overhead, including model loading and GPU warm-up, was excluded from the timing calculation. The current implementation processed 4K footage at approximately 16.7 fps, corresponding to an average processing time of 59.9 ms per frame.

Although this throughput is below the original recording frame rate of 60 fps, the practical requirement depends on the intended use case. For post-shot review, training feedback, and tactical discussion between shots, the system does not necessarily need to process all frames at the original video frame rate. Instead, it must provide analysis results before the next shot or the next trial. Based on the measured throughput, processing a shot segment of approximately 1200 frames requires approximately 72 s, suggesting that the current implementation is suitable for offline analysis and practical training feedback.

In contrast, applications that require continuous visualization during a shot or immediate real-time coaching would necessitate further optimization. Possible strategies include adopting lighter detector variants, reducing the input resolution, restricting inference to regions of interest, applying selective frame processing, and using inference-optimization

frameworks such as TensorRT. In applications where frame-by-frame high-precision sweep estimation is not required, selective frame processing may reduce the computational load while preserving the overall trajectory and temporal pattern of sweeping. Similarly, lowering the input resolution may enable real-time or near-real-time tracking, although the resulting decrease in apparent brush size may reduce detection accuracy. Therefore, the trade-off among estimation accuracy, temporal resolution, and processing speed should be evaluated in future work toward full real-time deployment.

4. Discussion

This section interprets the quantitative results reported in Section 3, focusing on why the observed error characteristics arose and how camera placement, sweep style, and the apparent size of the brush influenced estimation accuracy.

4.1. Interpretation of Position Accuracy

Tables 4–6 indicate that when the camera was positioned on the center line (lateral offset: 0 m), the estimation errors were nearly identical across the three sweep styles. All three sweep styles were reconstructed with the same centimeter-level accuracy (Table 4), indicating stable estimation performance. The detection rate (Rate) also remained high, ranging from 0.97 to 0.99, which confirms that the proposed system can reliably reconstruct sweep positions with minimal tracking loss under this camera configuration.

When the lateral camera offset increased, the overall estimation error also increased slightly. Specifically, the average MAE increased to 0.05 m (RMSE = 0.06 m) at an offset of 1.5 m and to 0.06 m (RMSE = 0.07 m) at an offset of 3.0 m. Compared with the center-line configuration, this corresponds to a moderate degradation of approximately 2–3 cm. However, this increase was not uniform across all sweep styles. The error increase was particularly pronounced for diagonal sweeping. For example, the MAE increased to 0.08 m (RMSE = 0.09 m) at 1.5 m offset and further to 0.11 m (RMSE = 0.13 m) at 3.0 m offset, whereas longitudinal and lateral sweeping remained relatively stable (MAE $\approx$ 0.03–0.04 m, RMSE $\approx$ 0.04 m). These results indicate that the proposed system exhibits not only a dependence on camera offset but also a style-dependent error characteristic in which diagonal sweeping is more sensitive to viewpoint displacement.

Despite this increase in positional error, the detection rate remained consistently high (0.96–0.99) regardless of camera offset or sweep style. This suggests that the primary source of error is not missed detections but rather inaccuracies introduced during projection-based coordinate reconstruction. In monocular perspective projection, position estimation becomes more sensitive to geometric distortion as objects move farther from the image center. Consequently, even small pixel-level detection errors in the OBB vertices or center positions can be amplified into larger spatial errors in real-world coordinates.

The larger error observed in diagonal sweeping is particularly noteworthy. At least two factors may contribute to this phenomenon. First, the brush has a slender geometry and undergoes rapid motion, which can cause instability in the orientation estimation of the OBB during diagonal sweeping. Variations in the estimated angle can shift the estimated vertex positions and center location, thereby increasing the reconstructed position error. Second, diagonal sweeping often includes a depth component relative to the camera view. Similar to the effect of lateral camera offset, this depth motion can increase the influence of perspective nonlinearity, leading to amplified spatial reconstruction errors.

In curling, players intentionally vary the area of the ice surface swept in front of a moving stone to control its behavior. Sweeping across the entire width of the stone's running path reduces friction more uniformly and primarily increases the travel distance of the stone. In contrast, sweeping only the left or right side of the running path selectively

reduces friction on that side, thereby influencing the magnitude and direction of the curl. Consequently, distinguishing whether sweeping is performed over the left half, right half, or entire width of the stone's running path is tactically important. Since the diameter of a curling stone is approximately 30 cm, half of its width corresponds to about 15 cm. This value represents the spatial scale at which the sweeping location becomes tactically meaningful. The positional accuracy achieved by the proposed method (MAE = 0.03–0.06 m) is substantially smaller than this 15 cm threshold, indicating that the system is sufficiently precise to determine which region of the stone's running path was swept during play.

For comparison, Aihara et al. developed a curling-stone tracking system and reported an average positional error of 0.02 m [8]. The sweep position estimation proposed in this study achieved an overall MAE of 0.03–0.06 m, which is comparable to that result under favorable camera conditions. It should be noted, however, that Aihara et al. focused on curling stones, which have stable appearances and relatively smooth motion. In contrast, the present study targets brushes, which are slender, small, and move rapidly. Considering these additional challenges, achieving centimeter-level accuracy from monocular video in real curling environments demonstrates the practical feasibility of the proposed method.

As a reference from another sport, Zhang and Li proposed a tennis analysis system using YOLOv5 and TensorRT and reported a combined detection error of 0.39 m for tennis-ball detection [5]. Although direct comparison is difficult due to differences in task definition, camera configuration, and motion characteristics, it is noteworthy that the present method achieves centimeter-level accuracy under a single-camera setup. This level of precision is sufficient for quantitative analysis of tactical sweeping strategies and their influence on stone trajectories.

Finally, the measurement range in this experiment was limited to 15 m from the camera. This range corresponds to the section of the curling sheet where sweeping is most actively performed during actual gameplay. After being delivered, a stone travels toward the house while gradually decelerating after crossing the hog line. Sweeping is performed alongside the moving stone to control its speed and trajectory. In particular, the section from approximately the middle of the sheet to the center of the house is strategically the most important sweeping zone, as it is where the final adjustments to the stone's distance and line are made and can ultimately determine the success or failure of the shot. This section corresponds to approximately 15 m. Therefore, the measurement range used in this study was designed to cover this critical region.

In contrast, during the early phase immediately after delivery, when the stone is moving at a relatively high speed, the influence of sweeping is comparatively limited. Therefore, achieving high positional accuracy within the 15 m measurement range (MAE = 0.03 m and RMSE = 0.04 m) demonstrates that the proposed system can effectively cover the most practically important sweeping region in curling.

To investigate the effect of the apparent brush size on position estimation accuracy, Figure 7 shows the relationship between the pixel area of each detected brush and the corresponding position error. The pixel area serves as an indicator of the apparent brush size in the image and varies depending on factors such as the distance from the camera and the viewing angle. As shown in Figure 7, the position error remained small and stable when the pixel area was large, whereas the error increased as the pixel area decreased.

To quantitatively characterize this relationship, inverse square-root, inverse, power-law, logarithmic, and linear models were compared. Among these candidates, the power-law model $E = 2.30 \times 10^7 A^{-2.20} + 15.9$ provided the best fit to the data, with a coefficient of determination of ($R^2 = 0.722$). Here, E denotes the position error [mm], and A denotes the brush pixel area [pixel²]. This result indicates that the estimation error increases nonlinearly as the apparent brush size decreases. A plausible explanation is

that, when the brush occupies only a small number of pixels in the image, even a slight localization error of a few pixels can be amplified into a larger error after transformation to real-world coordinates. This tendency is also consistent with the results presented in the previous section, where larger errors were observed for diagonal sweeping motions and camera-offset conditions.

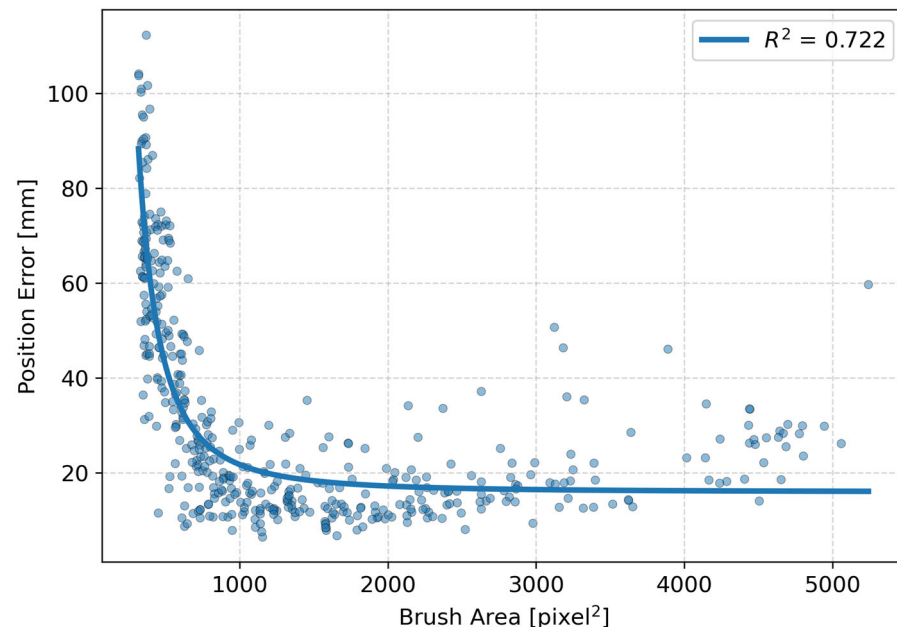


Figure 7. Relationship Between Brush Pixel Area and Position Error.

4.2. Interpretation of Orientation Accuracy

Tables 10–12 show that the sweep orientation estimation achieved an overall accuracy of $\text{MAE} = 4.80^\circ\text{--}5.10^\circ$, $\text{RMSE} = 6.19^\circ\text{--}6.58^\circ$, with a detection rate of 0.97–0.98. Similar to the position estimation results, the proposed system maintained a high detection rate while achieving orientation reconstruction within a few degrees. Unlike positional accuracy, however, the orientation error did not exhibit a clear monotonic increase as the lateral camera offset increased. The overall averages across the three camera configurations differed by less than 0.4° , indicating that orientation estimation was relatively insensitive to camera offset.

Instead, the estimation error appears to depend more strongly on the sweep pattern and the stability of brush-shape estimation. Longitudinal sweeping consistently produced the most accurate results, with $\text{MAE} = 2.48^\circ\text{--}3.50^\circ$ and $\text{RMSE} = 3.40^\circ\text{--}4.45^\circ$. In contrast, lateral sweeping resulted in larger errors ($\text{MAE} = 4.52^\circ\text{--}5.70^\circ$), while diagonal sweeping produced the largest errors ($\text{MAE} = 6.23^\circ\text{--}7.27^\circ$, $\text{RMSE} = 8.34^\circ\text{--}9.29^\circ$). This tendency is consistent with the characteristics of the proposed method, where orientation estimation relies on the angular prediction and vertex configuration of the oriented bounding box (OBB). Small fluctuations in the estimated OBB angle can cause instability in the inferred endpoints and orientation vector, which may lead to larger angular errors.

These observations suggest that further improvements in orientation estimation may be achieved not by correcting camera offset effects but by improving the robustness of angle estimation. Potential approaches include temporal smoothing, spike removal in the orientation signal, and geometric constraints that better reflect the physical shape of the brush when estimating endpoints.

From a practical perspective, the achieved orientation accuracy ($\text{MAE} \approx 4.8^\circ\text{--}5.1^\circ$) is sufficient to characterize sweeping direction in tactical and technical analyses. In curling, the direction of the brush determines the direction in which friction is applied to the

ice surface. Players typically adjust the brushing direction depending on the tactical situation, switching between longitudinal, lateral, or diagonal sweeping relative to the stone's travel direction. Therefore, in many analytical contexts, the absolute angle value itself is less important than identifying directional categories or detecting changes in brushing direction during a play. An error of approximately five degrees is small relative to the angular differences between these categories, which typically span several tens of degrees. Consequently, the proposed method allows meaningful identification of sweeping patterns and directional transitions during gameplay.

As a related example of orientation estimation, Roos et al. reported an RMSE of 9.77° when estimating vehicle orientation from high-resolution radar images [18]. Although the sensing modality (radar vs. optical), target object (vehicles vs. brushes), and experimental conditions differ substantially, the overall RMSE obtained in this study ($\approx 6.2^\circ$ – 6.6° , Tables 10–12) demonstrates that the proposed method achieves orientation estimation with several-degree-level accuracy using image-based observations. This is particularly noteworthy because the target objects in this study are small, slender brushes undergoing rapid motion and motion blur in a real sports environment. Maintaining a detection rate of ≈ 0.97 – 0.98 under such conditions highlights the practical robustness of the proposed system.

Because the system relies only on a monocular RGB camera, it offers significantly greater flexibility and ease of deployment compared with systems requiring permanently installed multi-camera setups. Furthermore, the proposed two-stage detection strategy—where brush detection is restricted to regions around detected players—helps improve detection accuracy for small brush objects while suppressing false positives. As a result, sweeping actions can be reliably extracted even in complex game scenes with multiple players and stones.

Traditionally, sweeping locations and directions have often been assessed qualitatively based on player intuition and experience. The proposed system enables automatic recording and quantitative analysis of sweeping behavior using ordinary match footage. Such a non-contact and low-cost framework has the potential to support training analysis, technical feedback, and performance visualization. In addition, the resulting dataset of sweep positions and orientations could serve as a foundation for future studies investigating the causal relationship between sweeping actions and stone trajectories.

To investigate the effect of apparent brush size on orientation estimation accuracy, Figure 8 shows the relationship between the pixel area of each detected brush and the corresponding orientation error. The orientation angle was estimated from the direction of the major axis of the oriented bounding box (OBB). When the brush appears smaller in the image, the number of pixels representing the major axis decreases, making the estimated angle more sensitive to slight localization errors at its endpoints.

As shown in Figure 8, the orientation error remained relatively small and stable when the pixel area was large, whereas larger errors tended to occur when the pixel area was small. Although a statistically significant negative correlation was observed between pixel area and orientation error ($r = -0.259$, $p < 0.001$), the correlation was weak. These results suggest that the apparent brush size has some influence on orientation estimation accuracy. However, orientation estimation error cannot be explained solely by brush size and is also affected by other factors, such as brush posture and variations in the detected object boundaries.

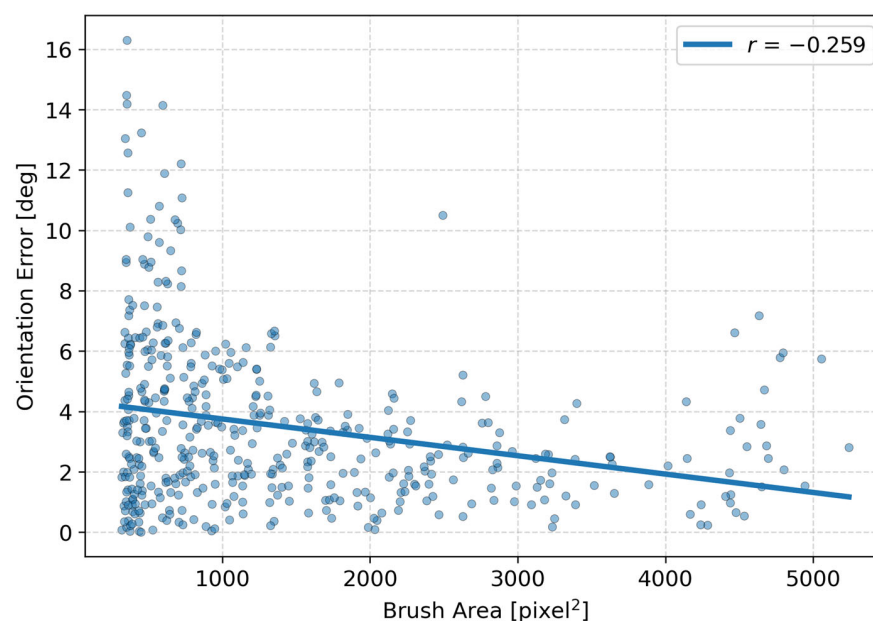


Figure 8. Relationship Between Brush Pixel Area and Orientation Error.

4.3. Comparison with Sensor-Based Sweeping-Analysis Approaches

Sensor-based systems and the proposed vision-based system measure different aspects of sweeping behavior. Force and pressure sensors can directly quantify how strongly a player sweeps, and IMU-based systems can capture acceleration, vibration, and stroke frequency with high temporal resolution [2]. However, these systems generally require instrumented equipment or attached sensors and do not directly provide continuous brush position and orientation in the curling-sheet coordinate system.

In addition, sensor-based approaches may have practical limitations in real curling environments. Attaching sensors to the brush or the player may affect the weight, stiffness, balance, grip feeling, and naturalness of the sweeping motion. Such effects are particularly important in high-performance sport, where even small changes in equipment or tactile feedback may influence the athlete's technique and performance. Moreover, the use of additional sensors or instrumented brushes may be restricted in official competitions depending on competition rules and event regulations.

In contrast, the proposed monocular-camera-based system estimates where and in which direction the brush moves on the sheet without attaching sensors or markers to the player or brush. Because the method analyzes video frame sequences, it can also describe temporal aspects of sweeping, such as trajectory, duration, and movement patterns, although its temporal resolution is limited by the video frame rate. A direct numerical comparison of accuracy is therefore not straightforward because the measured quantities differ fundamentally. Instead, the two approaches should be regarded as complementary: sensor-based systems quantify the intensity and high-frequency temporal dynamics of sweeping, whereas the proposed system quantifies the spatiotemporal distribution and direction of sweeping on the sheet (Table 13).

Table 13. Concise comparison between sensor-based and vision-based sweeping analysis.

Aspect	Sensor-Based Approaches	Proposed Vision-Based Approach
Main measured quantities	Force, pressure, acceleration, vibration	Brush position, orientation, trajectory, duration, and movement pattern

Table 13. *Cont.*

Aspect	Sensor-Based Approaches	Proposed Vision-Based Approach
Type of information	Mechanical and high-frequency temporal information	Spatiotemporal and directional information in sheet coordinates
Contact/equipment modification	Instrumented brush or attached sensors required	Non-contact and markerless; no modification of brush or player equipment required
Practical constraints	May affect natural sweeping motion and may be restricted in official competitions	Sensitive to camera placement, apparent brush size, occlusion, and image quality
Accuracy considerations	Accuracy is defined for force, pressure, or acceleration and cannot be directly compared with spatial error	Position MAE = 0.03–0.06 m; orientation MAE = 4.80°–5.10°

4.4. Alternative Imaging and Camera Configurations

Monochrome cameras are a promising alternative for future sweep-analysis systems. Since the final orientation estimation in the proposed method is computed from the geometric direction of the detected OBB, this stage does not directly require color information. Moreover, monochrome sensors can offer higher light sensitivity, lower noise, and higher frame-rate acquisition, which may be beneficial in curling halls and may help reduce motion blur during rapid sweeping.

However, the current system should not be regarded as completely independent of color information. The brush detector is based on a CNN model trained using RGB images, and the detection stage may exploit not only geometric shape but also color, brightness, texture, and contrast cues, such as the color of the brush pad, its contrast against the ice surface, and boundaries between the brush and the player's clothing. Therefore, applying the current RGB-trained model directly to monochrome images may not yield the same performance. In addition, if color information is used to distinguish red and yellow stones in the pipeline, a fully monochrome configuration would require an alternative stone-identification strategy. For these reasons, the use of monochrome cameras should be evaluated through dedicated data collection, retraining, and quantitative validation.

Cameras operating in other spectral ranges also merit investigation as future extensions of the proposed system. Near-infrared imaging may improve robustness to variations in visible-light illumination and reflections from the ice surface, which can differ substantially across curling venues. Thermal infrared imaging is also particularly interesting in the context of sweeping analysis, because sweeping changes the thermal state of the ice surface and may leave a measurable thermal footprint. Such observations could potentially link the spatial sweeping information obtained by the proposed system with the physical effect of sweeping on the ice.

However, these imaging modalities are outside the main scope of the present study, whose objective was to develop a simple and easily deployable sweep-analysis system using a single RGB camera. Infrared and thermal infrared cameras may increase equipment cost, require additional installation and calibration procedures, and necessitate new annotated datasets and model retraining because the appearance of brushes, players, stones, and the ice surface differs from that in RGB images. Thermal infrared imaging may also be affected by environmental temperature, air conditioning, surface reflections, and observation distance. Therefore, although infrared-based imaging is a promising direction for future

high-functionality analysis, it should be regarded as a complementary extension rather than a replacement for the simple RGB-camera-based framework proposed in this study.

Multiple synchronized cameras positioned at different angles would likely reduce perspective-related errors and improve the stability of spatial reconstruction, particularly when the brush is far from the camera or viewed obliquely. Multi-view reconstruction could therefore be beneficial for applications that require higher positional accuracy. However, the primary objective of this study was to develop a practical single-camera framework that can be deployed with minimal equipment and operational burden. Multi-camera systems require additional installation space, camera synchronization, multi-view calibration, and higher operational cost, which may limit their applicability in ordinary training or competition environments. We therefore regard multi-camera reconstruction as a complementary extension for future high-accuracy analysis rather than as the main configuration of the present system.

4.5. Limitations

The limitations of this study can be broadly categorized into two groups: those inherent to the proposed method and those related to the scope of the evaluation.

First, regarding the limitations of the method itself, the estimation accuracy depends on the apparent size of the brush in the image. As shown in Figures 7 and 8, both positional and orientation errors increase as the brush appears smaller. This is a fundamental limitation arising from the reduced number of pixels available for detecting small objects and from the monocular camera setup, in which a small image-space error can be amplified into a large error in the real-world coordinate system. Consequently, this limitation cannot be completely eliminated through improvements to the detection algorithm alone. Therefore, the accuracy reported in this study may not be maintained under conditions where the brush appears extremely small, such as long-distance or low-resolution imaging, or when the camera view is highly oblique relative to the sheet surface. Potential solutions include applying super-resolution techniques to improve the effective image resolution, incorporating temporal smoothing to stabilize detections, and using higher-resolution cameras or telephoto lenses.

Second, several limitations arise from the scope of the evaluation. First, the experiments were conducted using data from only one participant. Since sweeping motions vary depending on factors such as body size, skill level, and handedness, the results obtained from a single participant may not generalize to all players. Second, the evaluation considered only single-player sweeping scenarios and did not investigate simultaneous sweeping by multiple players, which frequently occurs in actual games. In such situations, increased occlusions among brushes and players may make brush detection, tracking, and ID association more challenging. Third, only one type of brush was used in the evaluation, and the influence of variations in brush shape and color on the estimation accuracy was not examined. To address these limitations, datasets containing multiple players, simultaneous sweeping scenarios, and multiple brush types have already been collected, and further evaluations are currently underway.

It should also be noted that the participant, brush, imaging conditions, and camera viewpoints used in the evaluation were not included in the training data of the detection model. Therefore, although the evaluation scope remains limited, the results provide evidence of the proposed method's ability to generalize to previously unseen conditions.

5. Application of the Sweep Analysis System

Building on the estimation accuracy and error characteristics discussed in Section 4, this section describes how the proposed system can be applied in practical training and competition settings.

In competitive curling, ice conditions evolve dynamically throughout the match as repeated sweeping alters the microtexture of the ice surface. This gradual abrasion reduces local friction and can accumulate over the course of multiple ends, leading to changes in stone curl, running distance, and the effectiveness of sweeping interventions. Coaches and players routinely adjust their strategies based on perceived ice conditions, yet these assessments have remained subjective due to the absence of objective tools for monitoring how sweeping is distributed across the sheet. The proposed system provides a quantitative foundation for these decisions by reconstructing sweep movements directly from video-based brush tracking. The visualization interface (Figure 9) enables simultaneous inspection of the original video, the reconstructed brush trajectory, and a magnified view of brush orientation. This design allows users to correlate the player's sweeping technique with the spatial relationship between the brush and the stone. For example, the reconstructed trajectories reveal whether sweeping is concentrated during the stone's early travel or whether the player continues sweeping closer to the house—information that is often decisive in managing stone deceleration and curl. Beyond simple trajectory display, the system extracts time-series indicators such as stroke frequency, instantaneous sweeping speed, and variability in sweep direction. These metrics provide a quantitative basis for evaluating technique consistency and the physical workload imposed on the athlete. In practical settings, coaches can use these indicators to assess temporal changes within a single end. For instance, a noticeable decline in sweeping speed in later ends may indicate fatigue or insufficient conditioning, prompting adjustments in player rotation or endurance training.

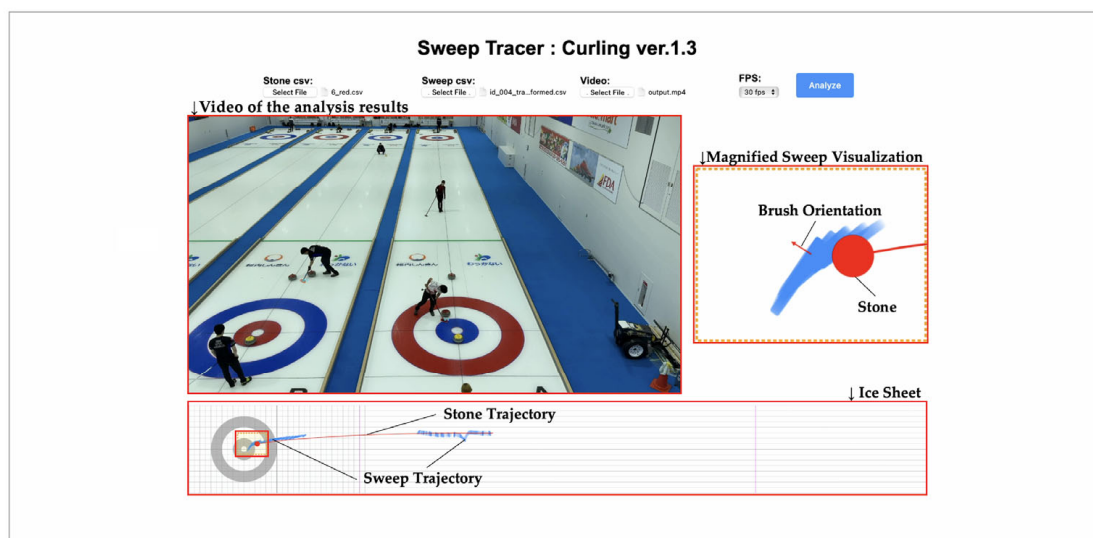


Figure 9. Visualization of the sweep estimation.

Aggregating sweep data across the entire match further enables high-level spatial analysis. The accumulated sweep coordinates can be transformed into sweeping-density maps. Such heatmaps could highlight areas of disproportionately heavy sweeping across ends, revealing how teams adapt to evolving ice conditions. For example, a shift from wide-sheet sweeping in early ends to house-proximal sweeping in later ends may indicate increased frost or pebble wear, offering a quantitative explanation for observed changes in

curl behavior. This approach has the potential to support post-game strategic review and, if combined with the faster processing discussed in Section 3.3, tactical decision-making between shots in the future. The sweep visualization can also be integrated with our existing stone-tracking framework [8] to perform combined analyses of brush movement and stone response. This enables finer-grained evaluation of tactical decisions. For example, the system can reveal whether sweeping on one side consistently suppresses curl under specific ice conditions or whether prolonged sweeping significantly alters the stone's deceleration profile. Such combined analysis could provide insight into the relationship between sweeping technique and stone trajectory, although establishing causal relationships will require further controlled studies—information that has traditionally relied on subjective interpretation. Taken together, the visualization and analytical capabilities of the system offer a comprehensive tool for understanding sweeping behavior within real match environments. By linking spatial, temporal, and directional information, the proposed approach provides practical value for athlete feedback, coaching decisions, and tactical planning while also offering a quantitative basis for understanding how sweeping shapes the physics of curling play.

For deployment in actual training or competition environments, the camera should be installed outside the playing area so that it does not interfere with players, stones, officials, or competition operations. Possible installation locations include spectator stands, facility-mounted elevated positions, or other fixed positions outside the sheet area. It should be noted that the camera configuration used in the present evaluation—a camera positioned 15.0 m behind the end of the sheet and 2.0 m above the ice surface—was selected as a controlled experimental setup for validating estimation accuracy and is not intended to represent a recommended deployment layout for actual competitions.

In practical use, the camera placement should satisfy two requirements: the official sheet markings must be visible for calibration, and the apparent size of the brush must be sufficient for reliable detection. Higher-resolution sensors or longer-focal-length optics may help preserve the brush pixel area when the camera must be placed farther from the sheet. However, long-focus optics also narrow the field of view, which may make it more difficult to include sufficient sheet markings for calibration. Therefore, practical camera placement requires balancing brush visibility, calibration-marker visibility, safety, and non-interference with competition operations.

Pan-tilt or auto-tracking cameras synchronized with player movement are also promising because they could keep the sweeping region large and near the image center throughout the stone's travel. However, such dynamic camera operation invalidates the assumption of fixed extrinsic parameters and would require continuous or frame-wise camera-pose estimation, for example, by re-estimating the camera pose from sheet markings. We therefore regard dynamic-camera deployment as a future extension beyond the scope of the present study.

6. Conclusions

This study proposed a vision-based analysis system for estimating the position and orientation of curling brushes during sweeping using monocular video footage. The system integrates three main components: sweep detection using a two-stage YOLOv8-OB framework, multi-object tracking based on ByteTrack, and geometric calibration that converts image coordinates into real-world curling-sheet coordinates. By combining these modules, the system enables frame-by-frame reconstruction of sweeping trajectories and brush orientation without requiring specialized sensing equipment.

To evaluate the accuracy of the proposed method, experiments were conducted at the Allographics Kitami Curling Hall using a Qualisys Miquis M3 motion capture system

as ground truth. The results demonstrated that the proposed system achieved sweep-position estimation accuracy of $MAE = 0.03\text{--}0.06$ m and orientation estimation accuracy of $MAE = 4.80^\circ\text{--}5.10^\circ$, while maintaining a high detection rate of 0.96–0.99. These results indicate that centimeter-level positional accuracy and several-degree orientation accuracy can be achieved using a single RGB camera under realistic sports conditions.

The analysis also revealed that the estimation accuracy depends on both camera placement and sweep style. While positional error increased slightly with larger lateral camera offsets, orientation accuracy remained relatively stable. In addition, diagonal sweeping produced larger estimation errors than longitudinal or lateral sweeping, suggesting that brush geometry and rapid motion influence the stability of OBB-based orientation estimation.

Because the proposed system requires only a monocular RGB camera, it offers a practical and easily deployable framework for quantitative analysis of sweeping behavior in curling. The system enables objective recording and visualization of sweep positions and directions, which may support post-game analysis, training feedback, and tactical studies.

Future work will focus on expanding the dataset to include multiple players, simultaneous sweeping situations, and a wider variety of brush types in order to further evaluate and improve the robustness of the proposed method. In addition, future studies will investigate the performance of the proposed system under a broader range of recording conditions and camera placements. Efforts will also be made to optimize the computational pipeline for faster processing, with the goal of enabling near real-time sweep analysis during training sessions and competitive matches.

Author Contributions: Conceptualization, S.A.; methodology, R.H. and S.A.; software, R.H.; validation, R.H.; formal analysis, R.H.; investigation, R.H., S.A. and T.I.; resources, R.H. and S.A.; data curation, R.H.; writing—original draft preparation, R.H.; writing—review and editing, S.A. and T.I.; visualization, R.H.; supervision, S.A. and T.I.; project administration, S.A. and T.I.; funding acquisition, S.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: This study was approved by the Ethics Review Committee of The University of Electro-Communications (Approval No. H25010).

Informed Consent Statement: Informed consent was obtained from the participant in the evaluation experiment and from the individuals appearing in the footage recorded by the authors for the dataset.

Data Availability Statement: The data presented in this study are not publicly available due to privacy and ethical considerations but are available from the corresponding author upon reasonable request.

Acknowledgments: This work was supported by “The Enhancement of HPSC Infrastructure through Technology Innovation Project” of the Japan Sports Agency. During the preparation of this manuscript, the authors used ChatGPT GPT-5 (OpenAI) to improve the readability of the text. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Conflicts of Interest: The authors have no conflicts of interest to declare.

References

1. Marmo, B.A.; Farrow, I.S.; Buckingham, M.-P.; Blackford, J.R. Frictional heat generated by sweeping in curling and its effect on ice friction. *Proc. Inst. Mech. Eng. Part L J. Mater. Des. Appl.* **2006**, *220*, 189–197. [[CrossRef](#)]
2. Yanagi, H.; Maeno, N.; Maruyama, T. Development of a curling brush to measure sweeping force. In Proceedings of the 30th International Conference on Biomechanics in Sports, Melbourne, Australia, 2–6 July 2012; pp. 354–356.
3. Sonobe, H.; Aoki, Y.; Miya, O.; Murata, K.; Ogihara, E.; Okawara, Y.; Ozaki, S.; Tomizawa, N.; Murata, J. Sweeping effects on curling and friction estimation. *Sports Eng.* **2024**, *27*, 31. [[CrossRef](#)]

4. Naik, B.T.; Hashmi, M.F.; Bokde, N.D. A Comprehensive Review of Computer Vision in Sports: Open Issues, Future Trends and Research Directions. *Appl. Sci.* **2022**, *12*, 4429. [\[CrossRef\]](#)
5. Zhang, X.; Li, B. Tennis ball detection based on YOLOv5 with tensorrt. *Sci. Rep.* **2025**, *15*, 21011. [\[CrossRef\]](#) [\[PubMed\]](#)
6. Qi, J.; Li, D.; He, J.; Wang, Y. Optically Non-Contact Cross-Country Skiing Action Recognition Based on Key-Point Collaborative Estimation and Motion Feature Extraction. *Sensors* **2023**, *23*, 3639. [\[CrossRef\]](#) [\[PubMed\]](#)
7. Elfmark, O.; Ettema, G.; Groos, D.; Ihlen, E.A.F.; Velta, R.; Haugen, P.; Braaten, S.; Gilgien, M. Performance Analysis in Ski Jumping with a Differential Global Navigation Satellite System and Video-Based Pose Estimation. *Sensors* **2021**, *21*, 5318. [\[CrossRef\]](#) [\[PubMed\]](#)
8. Aihara, S.; Chen, G.; Hara, R.; Nakagawa, M.; Ogasawara, A.; Yanagi, H.; Takegawa, Y.; Ito, T.; Masui, F. Development of a simple tracking system to monitor curling stone dynamics. In *Proceedings of the 12th International Conference on Sport Sciences Research and Technology Support*; SciTePress: Setúbal, Portugal, 2024; pp. 22–33. [\[CrossRef\]](#)
9. Jocher, G.; Chaurasia, A.; Qiu, J. Ultralytics YOLOv8. 2023. Available online: <https://github.com/ultralytics/ultralytics> (accessed on 10 November 2025).
10. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788. [\[CrossRef\]](#)
11. Feng, S.; Huang, Y.; Zhang, N. An Improved YOLOv8 OBB Model for Ship Detection through Stable Diffusion Data Augmentation. *Sensors* **2024**, *24*, 5850. [\[CrossRef\]](#) [\[PubMed\]](#)
12. Kingma, D.P.; Ba, J.L. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980. [\[CrossRef\]](#)
13. Zhao, Q.; Sheng, T.; Wang, Y.; Tang, Z.; Chen, Y.; Cai, L.; Ling, H. M2Det: A single-shot object detector based on multi-level feature pyramid network. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Honolulu, HI, USA, 27 January–1 February 2019; pp. 9259–9266. [\[CrossRef\]](#)
14. Zhang, Y.; Sun, P.; Jiang, Y.; Yu, D.; Weng, F.; Yuan, Z.; Luo, P.; Liu, W.; Wang, X. ByteTrack: Multi-object tracking by associating every detection box. In *Proceedings of the European Conference on Computer Vision (ECCV)*, Tel Aviv, Israel, 23–27 October 2022; pp. 1–21. [\[CrossRef\]](#)
15. Wojke, N.; Bewley, A.; Paulus, D. Simple online and realtime tracking with a deep association metric. In *Proceedings of the IEEE International Conference on Image Processing (ICIP)*, Beijing, China, 17–20 September 2017; pp. 3645–3649. [\[CrossRef\]](#)
16. Moré, J.J. The Levenberg–Marquardt algorithm: Implementation and theory. In *Numerical Analysis*; Watson, G.A., Ed.; Springer: Berlin, Germany, 1977; Volume 630, pp. 105–116. [\[CrossRef\]](#)
17. Goldfeld, S.M.; Quandt, R.E.; Trotter, H.F. Maximization by quadratic hill-climbing. *Econometrica* **1966**, *34*, 541–551. [\[CrossRef\]](#)
18. Roos, F.; Kellner, D.; Dickmann, J.; Waldschmidt, C. Reliable orientation estimation of vehicles in high-resolution radar images. *IEEE Trans. Microw. Theory Tech.* **2016**, *64*, 2986–2993. [\[CrossRef\]](#)

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.