第 30 回日本バーチャルリアリティ学会大会論文集（2025 年 9 月）

Visual Feedback and Evaluation System for Analyzing Guide Dog Condition Based on 3D Pose Features

Shuyao HUANG, Ansheng WANG, Yasutoshi MAKINO, and Hiroyuki SHINODA

1) 東京大学 新領域創成科学研究科（〒 277-0882 東京都文京区本郷 7-3-1, vrsj@star.t.u-tokyo.ac.jp）

Abstract: This study proposes a method to help trainers observe guide dogs' condition on a frame-by-frame basis in a visual feedback interface. An evaluation system using spatial features derived from 3D skeleton reconstruction is constructed.
Keyword： Guide Dog Training, Interface

1. Introduction

Currently, the number of guide dogs can hardly meet the demand of people with visual impairments. One of the main reasons for this shortage is the lengthy training process and the low success rate in guide dog training. In particular, inexperienced trainers often struggle due to a lack of practical knowledge in judging the dogs' behavior and condition.Virtual Reality(VR) technology, which is widely used in the field of education and skill transfer, offers promising support in this domain. By simulating realistic training scenarios, VR enables inexperienced trainers to gain hands-on experience without risking the well-being of real dogs. It also allows repeated exposure to rare or difficult situations, helping trainers improve their responsiveness and decision-making.

In order to realize VR-based training assistance, it is necessary to first establish a visual representation of guide dog training within a computational environment. This includes various aspects of the training process, such as the guide dog's condition by integrating visual, behavioral, and analytical information. By presenting the training scenes from a third-person perspective, the system affords trainers a multifaceted view of the dog's condition and behavior. Such visual feedback facilitates a more informed interpretation of training outcomes, supporting data-driven reflection and refinement of instructional strategies.

In this study, we introduce such an interface that integrates various forms of information for each frame of the training session. Specifically, it displays the raw 2D video, four-angle 2D views, a 3D skeleton model, frame-level feature values, a computed training score, and summary charts of key behavioral indicators. Among these, the training score serves as a comprehensive analytical output, integrating three key indicators to evaluate the dog's condition.

2. Related Work

2.1 Guide Dog Training Evaluation

Evaluating guide dog training requires precise behavioral analysis, which 3D skeleton reconstruction can support by capturing detailed motion data. Huang et al. introduced a system that generates 3D skeletal data for both humans and dogs using 2D keypoints and triangulation, enabling outdoor evaluation without fixed cameras [1]. However, due to its limited reconstruction accuracy, it remains insufficient for detecting subtle behavioral cues.

Uchida et al. adopted a fixed camera setup with checkerboard based calibration, allowing a more accurate estimation of both intrinsic and extrinsic parameters [2].

2.2 Training Assistance Interface

Designing an effective training assistance interface requires not only the visualization of raw input data, but also the integration of structured information that helps trainers intuitively assess the subject's condition and performance. Previous work such as PoseCoach has demonstrated that combining 2D video with reconstructed 3D skeletons and synchronized metric displays can support detailed movement analysis and facilitate the understanding of training outcomes [3].

3. Methods

3.1 3D Skeleton Reconstruction

The 3D skeleton reconstruction method mainly follows the previous research from Uchida [2]. Using four-angle videos shot on the same object to extract the 2D skeleton points of guide dogs and humans by the Pytorch-based pose estimation model MMPose [4]. Then, combining the intrinsic and extrinsic parameters obtained from camera calibration, the dog and human skeleton information can be triangulated into the 3D coordinates respectively.

3.2 Task Selection

According to the Japan Guide Dog Association, commonly tested behaviors include HEEL, WAIT, COME, DOWN, SIT. In this study, we focus on HEEL which requires the dog to walk consistently at the trainer's left side, adjusting its pace and direction in response to the trainer's movement and commands.

3.3 Indicators

To assess the dog's engagement during training, we focus on spatial features that are directly related to its position and orientation relative to the trainer. Specifically, we select three indicators: front distance, side distance, and the signed angle between the dog and human.

3.3.1 Front Distance(FD) and Side Distance(SD)

A distance vector from the guide dog's withers to the trainer's left hip is defined. This vector is projected onto the ground plane, and its components are analyzed with respect to the trainer's horizontal motion direction. The component aligned with the anterior-posterior axis is referred to as the FD, representing how far forward or backward the dog is positioned relative to the trainer. The component orthogonal to this direction is defined as the SD, which reflects the lateral deviation of the dog from the trainer's side.

3.3.2 Signed Angle Between Dog and Human(SA)

In addition to positional distances, the angular alignment between the dog and the trainer is also analyzed. Two vectors are defined: one from the dog's withers to its tail base, and another from the trainer's left and right hip midpoint to their torso direction. By projecting both vectors onto the ground plane, we compute the signed angle between them.

3.4 Interface

To assist trainers in intuitively assessing guide dog behavior during training sessions, we designed an interface(Figure 1) that integrates several models: video frames, multi-angle views, 3D skeleton visualization, feature-based evaluations and summary analytics. Each component is designed to provide interprescore and synchronized information at both the frame and session levels.

The interface is developed as a browser-based application using HTML5, JavaScript, and the PapaParse library [5]. Video playback and canvas-based visualizations are implemented with native HTML elements.

3.4.1 Frame Selection(M1)

Users can select specific frames for detailed inspection through two methods. The first is by directly inputting the desired frame number. The second is by interactively scrolling through the video playback.

3.4.2 Four-Angle View(M2)

To offer a more comprehensive visual reference, the interface displays synchronized video frames from four different camera angles, each capturing the dog and trainer from a distinct perspective.

3.4.3 3D Skeleton View(M3)

This module visualizes the 3D skeletons of both the guide dog and the trainer, reconstructed from multi-camera footage as described in Section 3.1.

3.4.4 Feature Evaluation and Training Score(M4)

This section displays the extracted indicator values for the selected frame. The top rows show the raw numerical values of the three indicators introduced in Section 2.3. The lower rows present a qualitative evaluation of each indicator based on threshold values recommended by the Japan Guide Dog Association (see score 1). Values falling outside the normal range are regarded as large deviations.

The last line of M4 presents the training score computed for the current frame. As described earlier, this score integrates the evaluations of the three indicators. A score of 2 is assigned if the indicator falls within the recommended range, 1 if within the normal but not recommended range, and 0 otherwise. The total score is the sum of the three, yielding a range from 0 to 6.

Table 1: Defined criteria ranges for Good Domain (GD) and Normal Domain (ND) for each indicator

	FD(m)	SD(m)	SA(deg)
GD	$[-0.1, 0.1]$	$[0.15, 0.35]$	$[-15, 30]$
ND	$[-0.2, 0.2]$	$[0.1, 0.45]$	$[-25, 50]$

3.4.5 Overall Training Summary(M5)

To provide a session-level overview, this section displays the distribution of frame counts across evaluation categories. The red bars indicate the number of frames falling within the recommended range, while the green bars represent frames within the broader normal range.

4. Experiments and Results

4.1 Effectiveness of the Score Method

To evaluate the effectiveness of the proposed training score, we conducted a comparative experiment using two guide dogs with different training levels. According to the previous research done by Uchida [2], one dog was classified as well-trained(WT), while the other had just started training, here we define it as less-training(LT). Both dogs were instructed to perform the same task, lasting 30 seconds, under identical environmental conditions

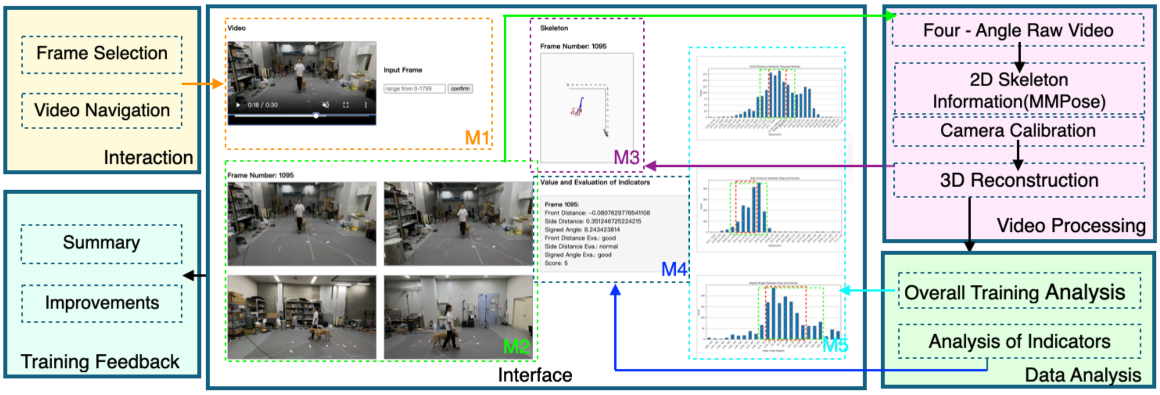

Figure 1: Overall View of the Interface

and with the same trainer.

To facilitate detailed behavioral analysis, the task was segmented into seven stages. These stages are arranged in chronological order, reflecting the actual progression of the training session. Each stage was manually determined based on the moment when the trainer issued a specific instruction.

1. Preparation: The trainer is stationary at the starting location and has not yet issued any verbal commands.
2. After Instruction: Defined as the initial 150 frames (2.5 seconds) immediately following the issuance of the HEEL command.
3. Normal-Speed HEEL: The phase during which the dog is executing the HEEL behavior at the trainer's normal walking pace.
4. Start Turn: A transitional phase encompassing the 100 frames (1.7 seconds) just before the trainer initiates a directional change.
5. After Instruction during HEEL: Defined as the 100 frames (1.7 seconds) immediately after receiving the trainer's instruction of Slow-Speed Heel.
6. Slow-Speed HEEL: The dog performs the HEEL task at a deliberately reduced pace.
7. End: The final 100 frames (1.7 seconds) after the trainer has signaled the end of the HEEL behavior and ceased movement.

The mean scores of both dogs across these stages are shown in Table 2. The WT dog demonstrated consistently high performance throughout the task, with a particularly notable performance during the Slow-Speed HEEL and End stages (mean scores of 4.74 and 4.33, respectively). This suggests that the dog clearly interprets the trainer's intentions and maintains an appropriate behavioral state.

In contrast, the less-trained dog exhibited generally lower scores. Although the score was relatively high immediately after commands were issued (e.g., 3.93 in After Instruction), it declined significantly in later stages, with the lowest values during Slow-Speed HEEL (1.48) and End (1.15). This reflects a common behavioral pattern in less-experienced dogs—initial responsiveness followed by reduced engagement over time.

These results are consistent with practical observations of guide dog behavior. The stage-wise breakdown further illustrates how the score can reveal subtle temporal dynamics in the dog's condition across different phases of task execution. Along with the further analysis based on this measurement, it is desirable to specify the quantified analysis of guide dog attention and capture its dynamics during the training, which is a good way to present the assistance to the training process.

Table 2: Mean scores for each training state (WT dog and LT dog)

State	WT	LT
Preparation	3.945	3.580
After Instruction	3.273	3.927
Normal-Speed HEEL	3.390	2.530
Start Turn	3.560	2.480
After Instruction during HEEL	3.320	3.300
Slow-Speed HEEL	4.743	1.484
End	4.330	1.150

4.2 Interface Output and Interpretation

In addition to the frame-level evaluation scores, the interface was designed to support intuitive interpretation and efficient behavior review. Each module within the interface contributes to a different aspect of information delivery and facilitates understanding of the guide dog's engagement state.

M1: The frame selection module allows users to quickly locate specific frames either by direct input or by scrolling through video playback. This is particularly useful when investigating low-score segments, enabling targeted examination of the dog's behavior in context.

M2: The four-angle view presents synchronized footage from four camera perspectives. By viewing the same moment from different angles, users can visually verify the dog's relative position, gait, and responsiveness, supporting interpretation of score results.

M3: The 3D skeleton view further enhances spatial understanding by visualizing posture and movement direction. This was especially helpful in identifying subtle misalignments during turning phases or slow-paced walking, where 2D video alone might be ambiguous.

M4: The feature value and evaluation panel provides immediate feedback on which indicators the dog's performance. When reviewing frames from the End stage of the less-trained dog, users could observe that multiple indicators fell outside the normal range, making it clear why the score was low.

The training score display aggregates the three indicators into a single, interpretable metric, facilitating rapid comparison across frames and phases.

M5: The overall analysis graph visualizes the frame distribution across score categories throughout the task.

5. Discussion

While the proposed interface shows promise in integrating visual and analytical feedback for guide dog training evaluation, several limitations and future directions remain.

First, the system currently operates only in offline mode using pre-recorded video and pre-processed skeleton data. To support real-time application, a fully automated pipeline is required for 2D keypoint extraction, 3D reconstruction, feature computation, and score visualization. Advancements in lightweight pose estimation models and GPU-accelerated processing could make this technically feasible in future iterations.

Second, the training score relies on rule-based thresholds defined by the Japan Guide Dog Association. Although this ensures initial interpretability, it lacks flexibility across different dogs. Machine learning-based scoring models trained on large-scale annotated data may offer greater adaptability and allow for personalized evaluation of engagement and performance.

Third, with 3D skeletal data already available for both the trainer and the dog, actual motion can be mapped onto 3D models to create immersive VR experiences. By modeling behavior patterns at various training levels and collecting more training data, we aim to reproduce diverse training scenarios in a VR environment, enabling experiential learning for novice trainers.

6. Conclusion

This study presents a visualization interface designed to assist in guide dog training evaluation by integrating multi-view video, 3D skeleton analysis, feature-based interpretation, and a computed training score. Through comparative experiments, we demonstrate that the training score effectively distinguishes between WT and LT dogs, and that the interface supports intuitive understanding of behavioral performance at both the frame and session levels.

The modular design of the interface enables users to explore training behavior from multiple perspectives and provides a foundation for more objective and scalable evaluation of guide dog engagement. Future improvements in real-time processing and data-driven scoring models hold promise for further enhancing the applicability and precision of the system in practical training scenarios and in the VR environment.

Acknowledgments

This work was supported by JSPS KAKENHI 21H05301 and 24K20810. The data used in this study was collected with the generous assistance of Mr. Tanaka of the Japan Guide Dog Association.

Reference

[1] Ansheng Wang, Rongjin Huang, Keisuke Fujii, Shinji Tanaka, Yoshiro Matsunami, Yasutoshi Makino, and Hiroyoki Shinoda. Dog's 3d skeleton reconstruction using a moving trainer for analysis of guide dog training. *2024 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pp. 4594–4599, 10 2024.

[2] Taisei Uchida, Ansheng Wang, Shinji Tanaka, Yasutoshi Makino, and Hiroyoki Shinoda. 三次元骨格解析に基づく盲導犬基礎訓練の評価と活用. ロボティクス・メカトロニクス講演会 2025, pp. 1P1–Q09, Yamagata, Japan, June 2025.

[3] Yuhang Liu, Jinhao Ma, and Yuanchun Shi. Posecoach: A customizable analysis and visualization system for video-based running coaching. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, pp. 1–15, 2022.

[4] OpenMMLab. Mmpose: Openmmlab pose estimation toolbox and benchmark. https://github.com/open-mmlab/mmpose, 2024.

[5] PapaParse. The powerful in-browser csv parser for javascript. https://www.papaparse.com/, 2024.