

A Dataset and Evaluation for Complex 4D Markerless Human Motion Capture

Supplementary Material

Abstract

Continuing with our main paper, this supplementary material provides additional details on the motion activities present in HUM4D and how the dataset is organized. First, we describe motion activities present in the dataset and how it is captured, with visual examples. We then present the folder structure of the proposed dataset with a flow chart for easier understanding.

6. Motion and Activity Type

In this supplementary, we provide a more detailed description of the motion categories in HUM4D and explain how the dataset is organized. HUM4D is designed to capture challenging motion patterns that are not sufficiently represented in existing markerless motion-capture benchmarks, including rapid local motion, heavy interaction occlusion, identity ambiguity, and depth variation. The dataset groups activities into four motion types, namely **Jittering**, **Occlusion**, **Near Far Camera**, and **ID Swap**. Representative examples for each motion type are provided in Fig. 8.

(i) **Jittering** refers to motion sequences with rapid or highly dynamic body movements that are difficult to estimate consistently across time. This category is intended to stress-test temporal stability and robustness under fast articulation changes, sudden pose transitions, and rapid appearance changes caused by motion.

- **Single Spin:** A single subject continuously rotates the body, producing fast orientation changes.
- **Single Jump:** A single subject performs repeated jump motion with strong vertical displacement and fast pose transitions.
- **Single Run in Place:** A single subject performs running motion in place, often followed by sudden stopping, which introduces rapid temporal changes in limb dynamics.
- **Group Spin:** Multiple participants rotate simultaneously, increasing temporal ambiguity and making consistent tracking more difficult.
- **Group Jump:** Multiple subjects perform fast jump motions together.
- **Group Cross Path:** Multiple subjects repeatedly walk in crossing directions.

Overall, the jittering category is designed to evaluate how well a method handles fast body motion and temporal inconsistency.

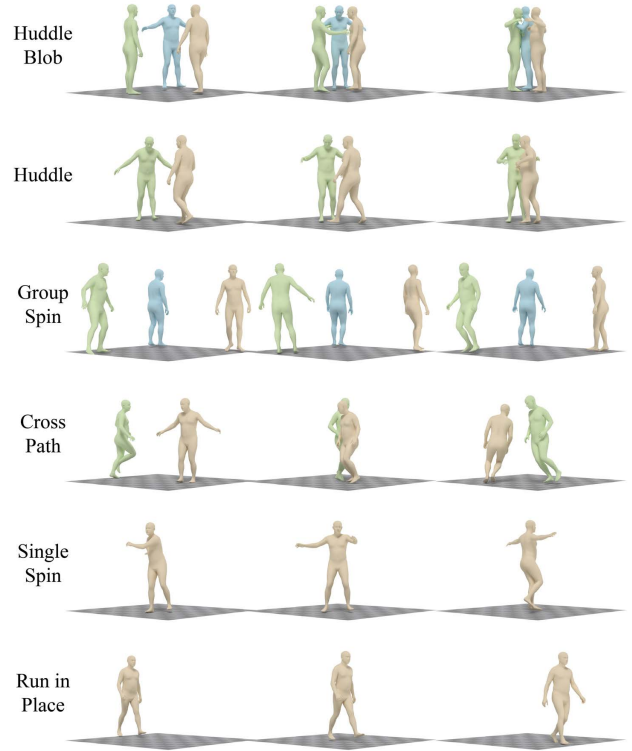


Figure 8. Representative examples of motion activities in HUM4D grouped by motion type.

(ii) **Occlusion.** includes activities in which body parts become partially or heavily hidden due to self-occlusion, interaction overlap, or close formation changes. These sequences are intended to evaluate robustness when visual evidence is incomplete.

- **Single Furniture Sit Stand:** A single subject repeatedly sits on and stands up from furniture, causing partial body occlusion and self occlusion
- **Group Huddle:** Multiple participants gather closely together, producing severe interaction overlap and limited visibility of individual body parts.
- **Group Huddle Blob:** Multiple subjects form a dense cluster, creating heavy body overlap and strong ambiguity in person body association.
- **Group Break Formation:** Multiple participants begin in a compact formation and then separate, resulting in changing visibility, overlapping limbs, and dynamic occlusion patterns.

These activities create frequent visibility loss and overlapping limbs that are challenging for both 2D keypoint de-

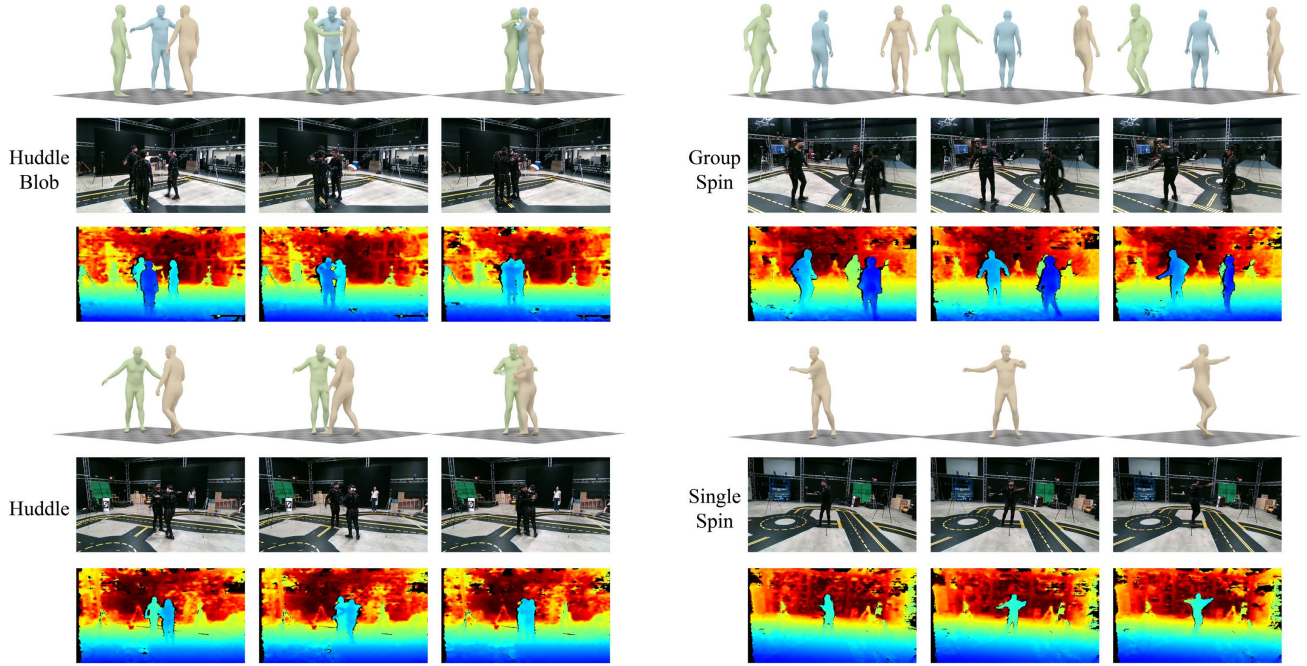


Figure 9. Representative examples from HUM4D. Each example includes the body-model or MoCap visualization, the synchronized RGB image, and the corresponding depth map.

tection and 3D reconstruction.

(iii) **Near Far Camera.** captures situations where subjects move toward or away from the cameras, producing substantial scale and depth variation. This category is designed to evaluate robustness to perspective effects and changing camera-relative distance.

- **Group Walk Toward Camera:** Multiple subjects walk toward the camera, creating large depth changes, increasing apparent body scale, and introducing viewpoint dependent variation across time.

This motion is challenging because large depth changes affect depth estimation and reconstruction quality.

(iv) **ID Swap.** refers to motion situations in which multiple people move in close proximity and exchange relative positions, making identity tracking difficult over time. This category is intended to reveal failures in temporal association and person identity consistency.

- **Group Run Around:** Multiple participants run around one another, causing frequent changes in relative position and making person identity association challenging.
- **Group Switch Location:** Multiple subjects exchange their spatial locations, explicitly testing whether methods can preserve person identity across motion.
- **Group Hide Each Other:** Participants move in ways that partially or fully block another, creating temporary disappearance and reappearance that can lead to identity confusion.

These sequences often expose failures in methods that depend on consistent person association across frames.

7. Dataset Arrangement

In this section, we describe how HUM4D is organized for convenient access. As illustrated in Fig. 10 and Fig. 11, the dataset follows a hierarchical structure from motion type to action category, recording setting, take index, and camera streams and annotation files.

At the top level, the dataset is divided into four motion type groups: **Jittering**, **Occlusion**, **Near Far Camera**, and **ID Swap**. Each of these groups contains multiple activity folders that correspond to the dominant motion pattern shown in that category. For example, the **Jittering** group contains activities such as Group Jump, Group RunInPlace Stop, Group Spin, Single Jump, Single RunInPlace Stop, and Single Spin. Similarly, the remaining motion groups contain their own activity folders, such as Single Furniture Sit Stand, Group Huddle, Group Huddle Blob, and Group Break Formation under **Occlusion**, Group Walk Toward Camera under **Near Far Camera**, and Group Run Around, Group Switch Location, and Group Hide Each Other under **ID Swap**.

Within each activity folder, the data is further organized by recording setting. This level reflects differences in cap-

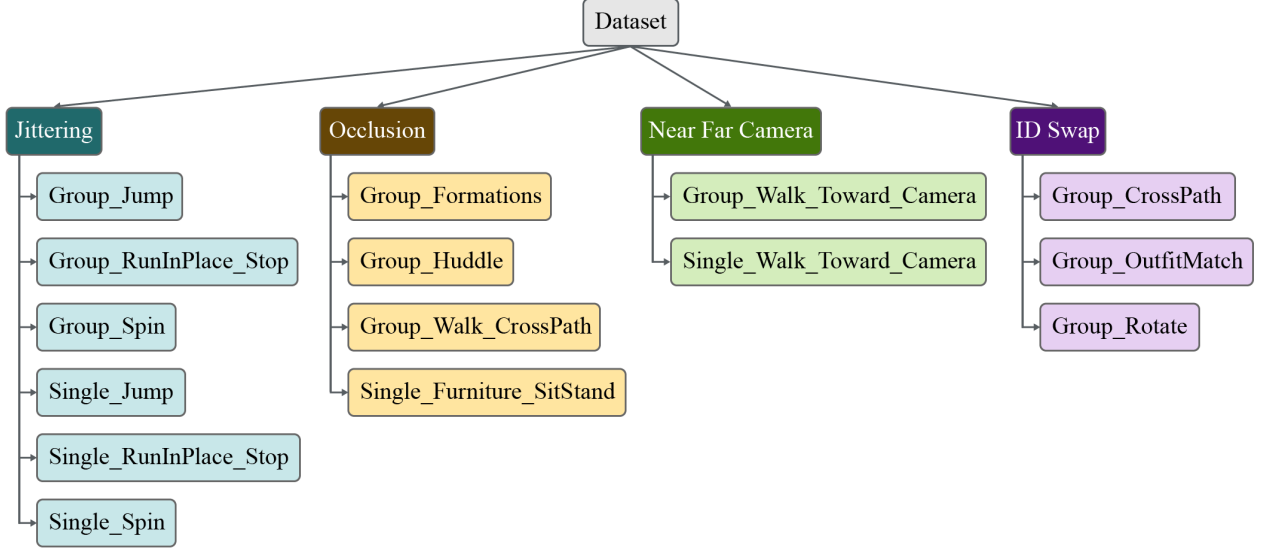


Figure 10. Top level hierarchy of HUM4D. The dataset is first grouped by motion type, and each motion type contains a set of activity folders.

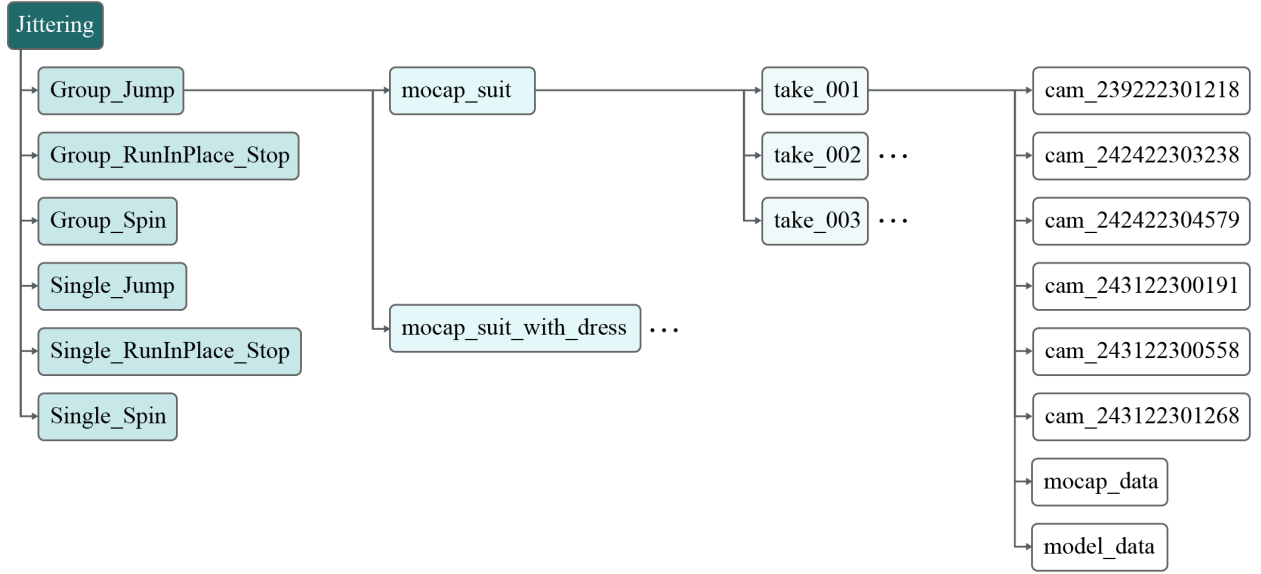


Figure 11. Example lower level hierarchy of HUM4D. Within each activity, the data is further organized by recording setting, take index, multi-view camera streams, and associated annotation files such as mocap data and model data.

ture configuration or subject appearance, such as `mocap_suit` and `mocap_suit with dress`. Each recording setting then contains multiple repeated captures indexed by take number. For example, `take_001`, `take_002`, and `take_003`.

Inside each take, the dataset contains synchronized multi-view camera streams together with motion annotations and processed model outputs. As shown in Fig. 11, each take includes several camera entries identified by camera-specific names such as `cam_239222301218`. In

addition to these image streams, each take also contains annotation entries such as `mocap_data`.

8. Motion Type Analysis

To further analyze method behavior on HUM4D, we report a breakdown of reconstruction performance by motion type. Since HUM4D is organized around four challenging motion categories, namely **Occlusion**, **ID Swap**, **Near-Far Camera**, and **Jittering**, this evaluation offers a more specific view of model behavior. As shown in Table 4, **ID Swap**

Motion Type	PARE	SPIN	HMR2.0	PersPose
Jittering	177.6	175.6	181.1	197.3
Occlusion	157.1	168.1	148.4	166.1
Near-Far Camera	178.6	170.6	205.7	209.9
ID Swap	265.3	268.7	260.8	267.7
Overall	185.7	189.2	184.9	199.2

Table 4. PA-MPJPE (mm) broken down by motion type on HUM4D.

is the most challenging motion type, producing the highest PA-MPJPE across all evaluated methods. By contrast, **Occlusion** yields the lowest average error overall, making it the least challenging category among the four in the benchmark. These results suggest that identity preservation under close multi-person interactions remain particularly difficult for existing methods, and they complement the overall benchmark score with more fine-grained insight.