

Translating Motion to Notation: Hand Labanotation for Intuitive and Comprehensive Hand Movement Documentation

Anonymous Authors

1 HAND LABANOTATION SYMBOL

In our article, we mention that Hand Labanotation encompasses 26 fundamental symbols, representing the spatial positions of regional vectors. As illustrated in Figure 1, these 26 fundamental symbols span 9 scales in the horizontal direction and 3 in the vertical direction. The horizontal scales are determined by the range of zenith angles θ derived from the polar coordinate transformation of the regional vectors. Similarly, the vertical scales are defined by the range of azimuth angles ϕ obtained from the same transformation. Notably, the ‘place’ scale in the horizontal dimension is unique, indicating a regional vector zenith angle of 0. Since regional vectors are formed by connecting adjacent nodes of the human hand, there is no scenario where the regional vector is a zero vector. Hence, the basic Hand Labanotation symbols do not include a ‘Place Normal’ symbol. The naming of all basic symbols in Hand Labanotation is derived from the combinations of all horizontal and vertical scales of regional vectors.

2 EXPERIMENTAL DETAILS

Experimental Parameters. We utilized 2 Nvidia RTX 3090 GPUs in our experiments and constructed a foundational network with ResNet152 [1]. Additionally, we developed the MHLFormer based on Multi-View Transformers (MVT) [3] for handling multi-view features. The detailed parameter settings used in the experiments are comprehensively presented in Table 1. Furthermore, the experimental parameter settings for several comparative methods are the same as those listed in Table 1.

Method	Loss	LR	Epochs	Batch	Optimizer
ResNet-50	\mathcal{L}_{CE}	1×10^{-4}	50	16	AdamW
ResNet-50	\mathcal{L}_{HXE}	1×10^{-4}	50	16	AdamW
ResNet-50	\mathcal{L}_{LH}	1×10^{-4}	50	16	AdamW
ResNet-50 + Cross Fusion	\mathcal{L}_{CE}	1×10^{-4}	50	16	AdamW
ResNet-50 + Cross Fusion	\mathcal{L}_{HXE}	1×10^{-4}	50	16	AdamW
ResNet-50 + Cross Fusion	\mathcal{L}_{LH}	1×10^{-4}	50	16	AdamW
ResNet-152	\mathcal{L}_{CE}	1×10^{-4}	50	16	AdamW
ResNet-152	\mathcal{L}_{HXE}	1×10^{-4}	50	16	AdamW
ResNet-152	\mathcal{L}_{LH}	1×10^{-4}	50	16	AdamW
ResNet-152 + Cross Fusion	\mathcal{L}_{CE}	1×10^{-4}	50	16	AdamW
ResNet-152 + Cross Fusion	\mathcal{L}_{HXE}	1×10^{-4}	50	16	AdamW
ResNet-152 + Cross Fusion	\mathcal{L}_{LH}	1×10^{-4}	50	16	AdamW

Table 1: Experimental parameter settings for ablation study.

Gesture Motion Reconstruction. In Section 4.4 of our experiment, we provide additional details on the process of reconstructing human hand movement skeletons using Hand Labanotation. In this process, each Hand Labanotation symbol represents the direction of a hand region vector and does not record the shape and coordinates of the hands. During reconstruction, we assign vector lengths based on the classic length proportions of each joint in the human hand. When reconstructing the azimuth ϕ and zenith angles θ corresponding to the regional vectors using Hand Labanotation, we uniformly

assign the median values of the quantization intervals. For example, if a regional vector corresponds to the Hand Labanotation symbol “Forward Normal”, the vector $\theta = 90, \phi = 0$ is assigned a length based on the proportional skeleton of the hand. Finally, following the node sequence in the hand skeleton structure, all reconstructed regional vectors are connected to form the restored hand skeleton.

3 MORE DETAILS ON HLD

Our Hand Labanotation Dataset (HLD) boasts a rich source of data, featuring multi-view, multi-scene, multi-actor, and multi-action image data. It primarily comprises the Interhand2.6M dataset [2] and the FreiHand dataset [4], with example images from these datasets shown in Figure 2 and Figure 3, respectively. Figure 2 presents data from the Interhand dataset, while Figure 3 is sourced from the FreiHand dataset. The Interhand portion of the HLD dataset was captured in a studio equipped with multiple cameras, including 80 to 140 cameras, operating at frame rates ranging from 30 to 90 frames per second (fps). The setup involved 350 to 450 directional LED point lights targeting the hands to ensure uniform illumination. The captured images have a resolution of 4096×2668 , but the resolution used in practice is 512×334 . This dataset encompasses a total of 36 videos, involving 26 different participants, of which 19 are male and 7 are female. These videos include both single-hand and double-hand types of hand sequences.

The FreiHand portion of the dataset includes data from both indoor and outdoor environments, capturing hand gestures from 32 subjects of different genders and ethnic backgrounds. Each subject was asked to perform actions with and without objects. The data was captured using 8 calibrated and temporally synchronized RGB cameras, with the actual image resolution used being 224×224 .

4 DETAILS ON ROBOTIC HAND CONTROL

In Experiment 4.6, we employed Hand Labanotation as a universal symbolic representation to control a robotic hand to perform actions as indicated by Hand Labanotation symbols, accompanied by visualized images. The following supplementary details are provided for this experiment.

The robotic hand used in the experiment is produced by *Beijing Yinchuang Technology Co., Ltd.*, with the product name and code being “*Lingqiao Hand RH56DF3-XR/L*”. It has a total of 12 joints and 6 degrees of freedom. Each of the 4 fingers has a 1 degree of freedom, while the thumb has 2 degrees of freedom, with each finger possessing only one movable joint. For more details, please refer to Robotic Hand User Manual. Since Hand Labanotation symbols quantify the azimuth and zenith angles of each regional vector of hand movements, and the robotic hand’s degrees of freedom in the joints correspond to the human hand’s skeletal structure, controlling the robotic hand is feasible. For joints with 1 single degree of freedom, it is necessary to extract the Labanotation symbol corresponding to the regional vector with the joint as the root node.

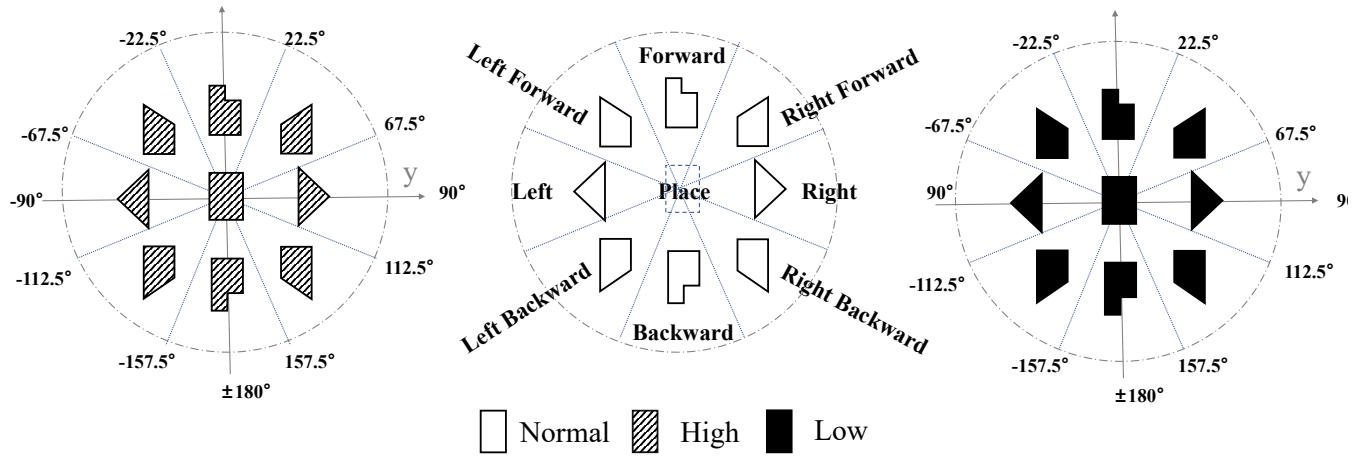


Figure 1: Display of 26 basic Hand Labanotation symbols.

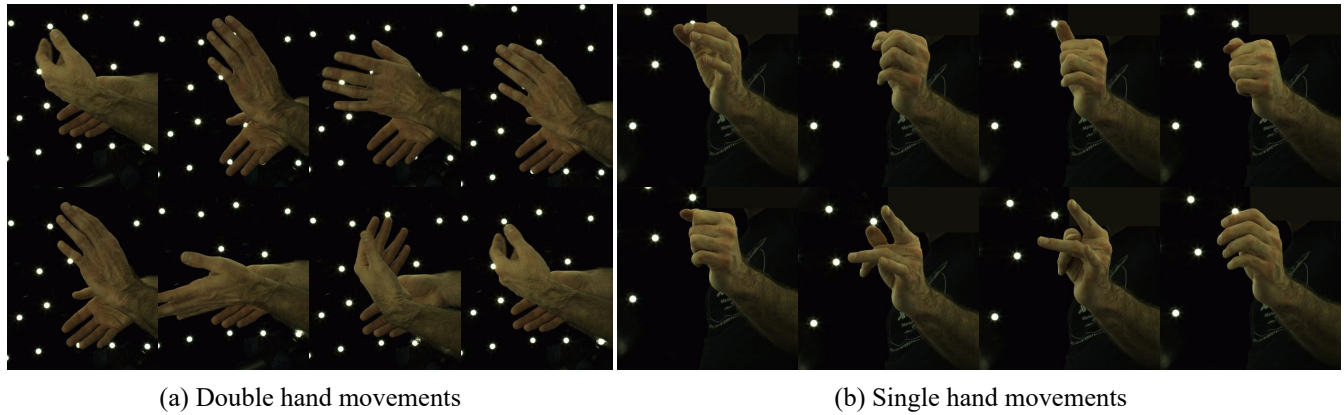


Figure 2: Partial data display from the Interhand dataset, including both bilateral and unilateral hand movements.

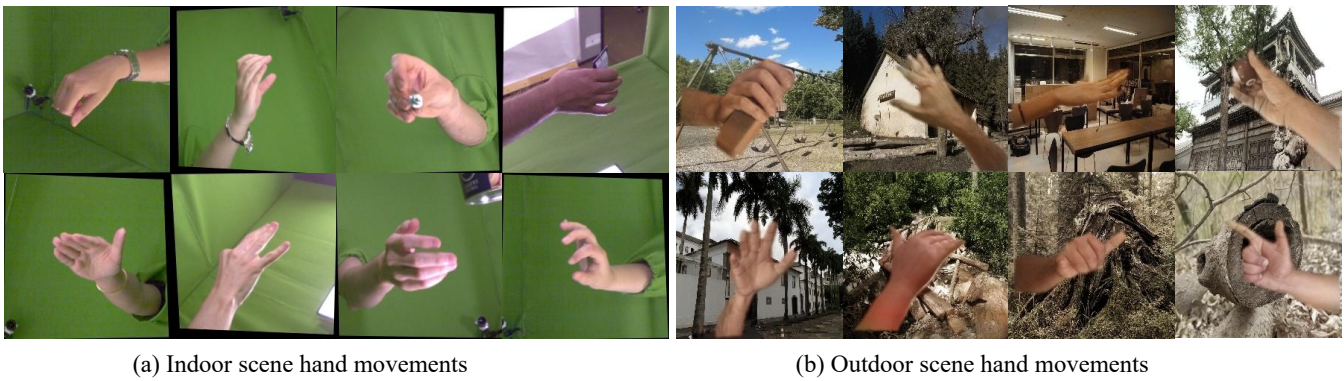
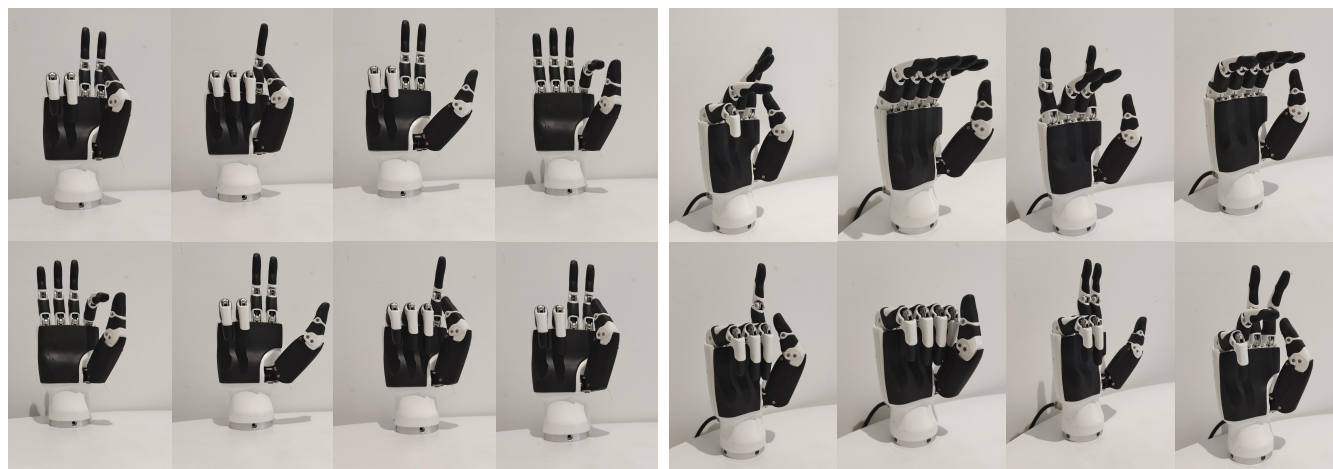


Figure 3: Partial data display from the Freihand dataset, including hand movements in indoor and outdoor scenes.

This symbol is then converted into a rotational angle for the robotic hand, allowing the hand region controlled by that joint to replicate the hand region represented by Hand Labanotation. For joints with

2 degrees of freedom, 2 rotation parameters are used to control the joint's movement. After obtaining the corresponding azimuth and rotation angles from Hand Labanotation, a transformation matrix



(a) Hand Labanotation-controlled robotic hand frontal view display

(b) Hand Labanotation-controlled robotic hand lateral view display

Figure 4: Further experimental demonstrations are provided, showcasing the control of a robotic hand via hand Labanotation to execute specified actions. These are represented through both frontal and lateral view illustrations.

is used to align the control parameters of the robotic hand with the parameters of the Labanotation symbols, enabling real-time control.

In addition, we present further experimental results showcasing the manipulation of a robotic hand controlled by hand Labanotation to perform corresponding actions, as illustrated in Figure 4.

5 OTHER MATERIALS

We have released the HLD dataset with all its annotation files HLD Dataset. And we have submitted our code.

REFERENCES

- [1] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep Residual Learning for Image Recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 770–778.
- [2] Gyeongsik Moon, Shou-I Yu, He Wen, Takaaki Shiratori, and Kyoung Mu Lee. 2020. InterHand2.6M: A Dataset and Baseline for 3D Interacting Hand Pose Estimation from a Single RGB Image. In *Proceedings of the IEEE European Conference on Computer Vision*. 548–564.
- [3] Shen Yan, Xuehan Xiong, Anurag Arnab, Zhichao Lu, Mi Zhang, Chen Sun, and Cordelia Schmid. 2022. Multiview Transformers for Video Recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 3333–3343.
- [4] Christian Zimmermann, Duygu Ceylan, Jimei Yang, Bryan Russell, Max Argus, and Thomas Brox. 2019. FreiHAND: A Dataset for Markerless Capture of Hand Pose and Shape From Single RGB Images. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 813–822.