Supplementary Material

1 Supplementary Figures and Tables

1.1 Supplementary Figures

Figure 1: Methodology pipeline

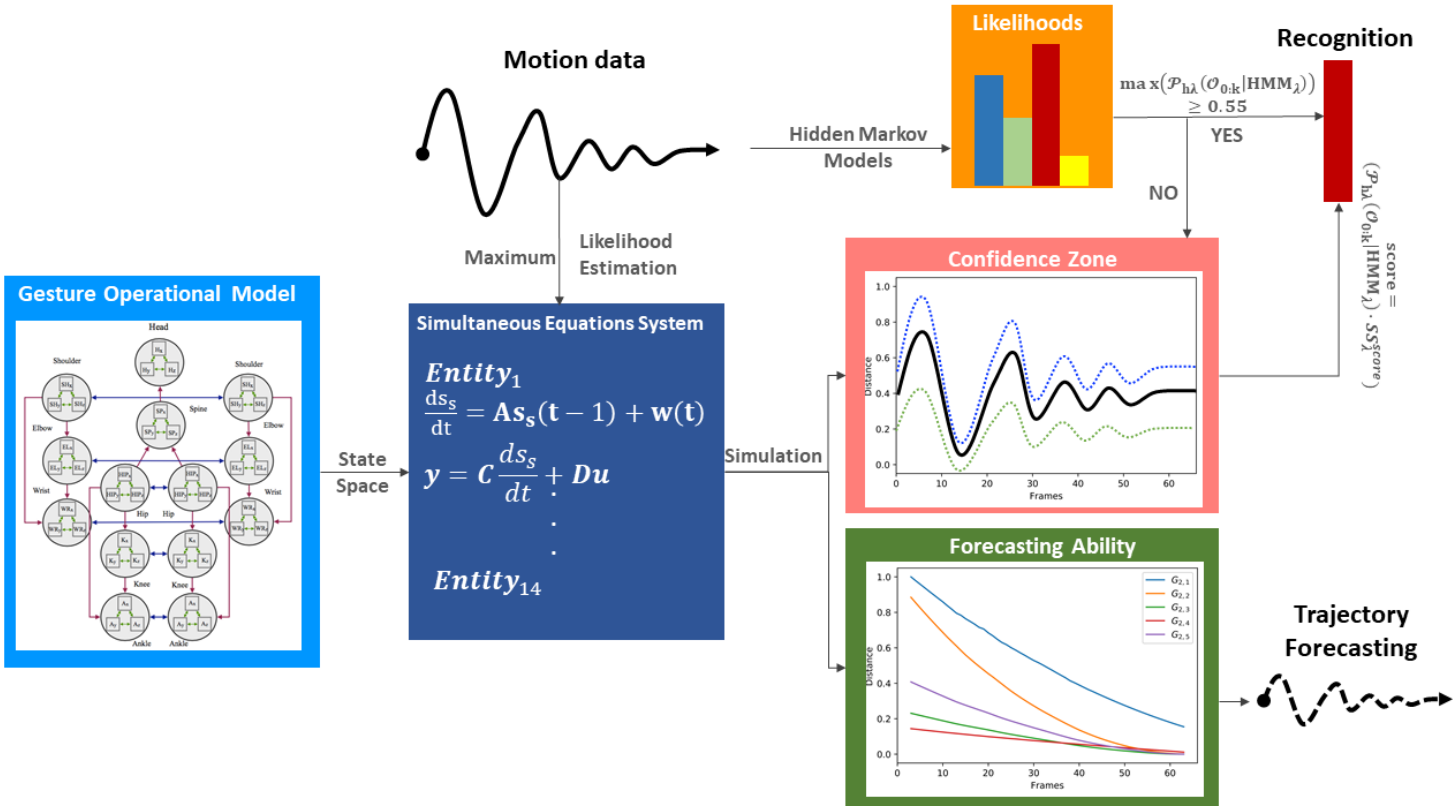


Figure 2: The full-body assumptions of the Gesture Operational Model (GOM) are depicted in the figure. Some relationships happen to be bidirectional, while others not. The relationships of the human body are governed by four different assumptions, intra-joint association, transitioning, inter-limb synergies, and intra-limb mediation. On the down-right of the image, the mapping on body is presented. The numbers in the GOM model, represent the corresponding body part of the joints representation from OpenPose framework.

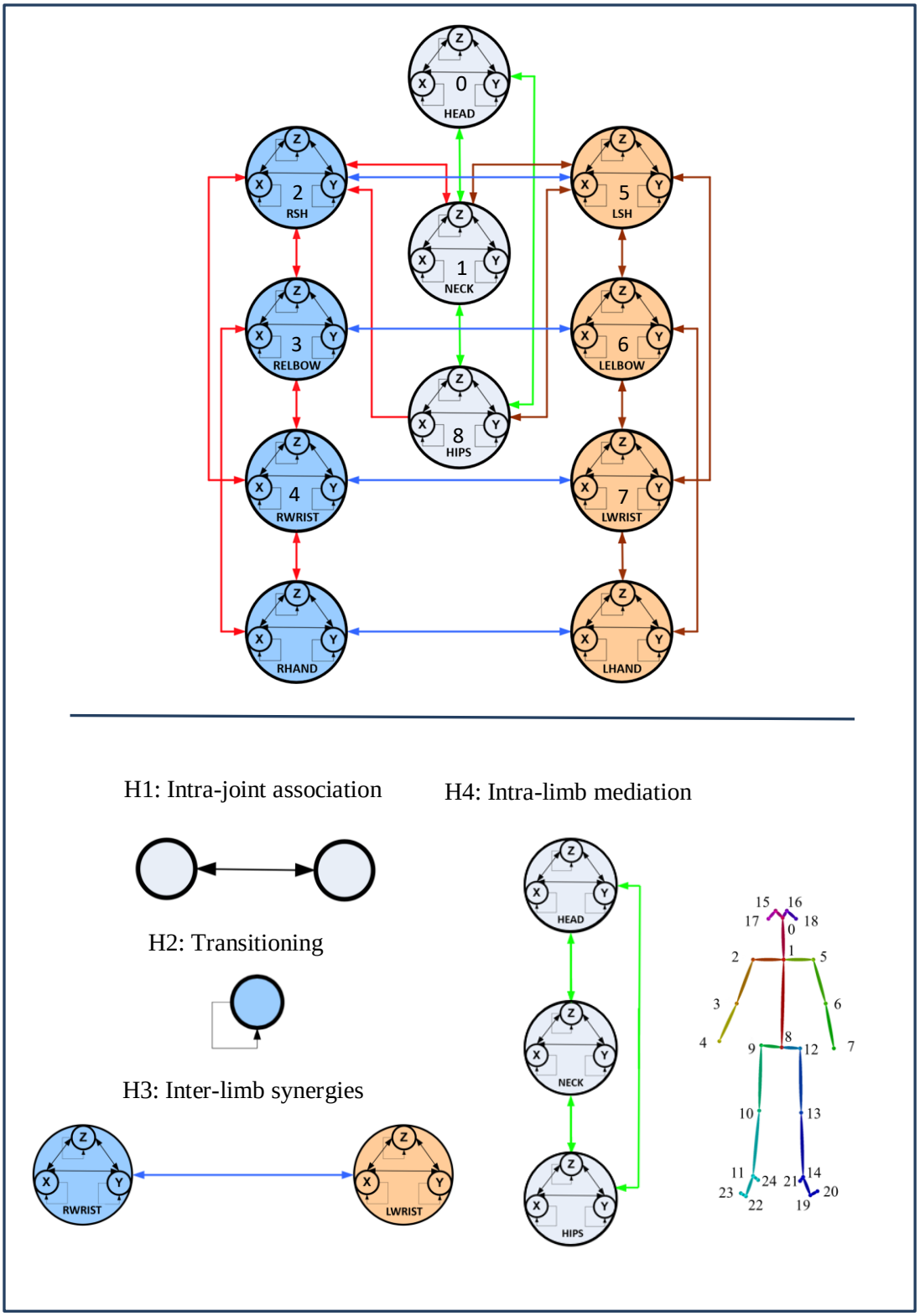


Figure 3: Examples of real motion observations (blue) and simulated values (orange) from the $RHAND_X$ State-Space model of the gesture $G_{1,1}$ (left) and $G_{1,4}$ (right)

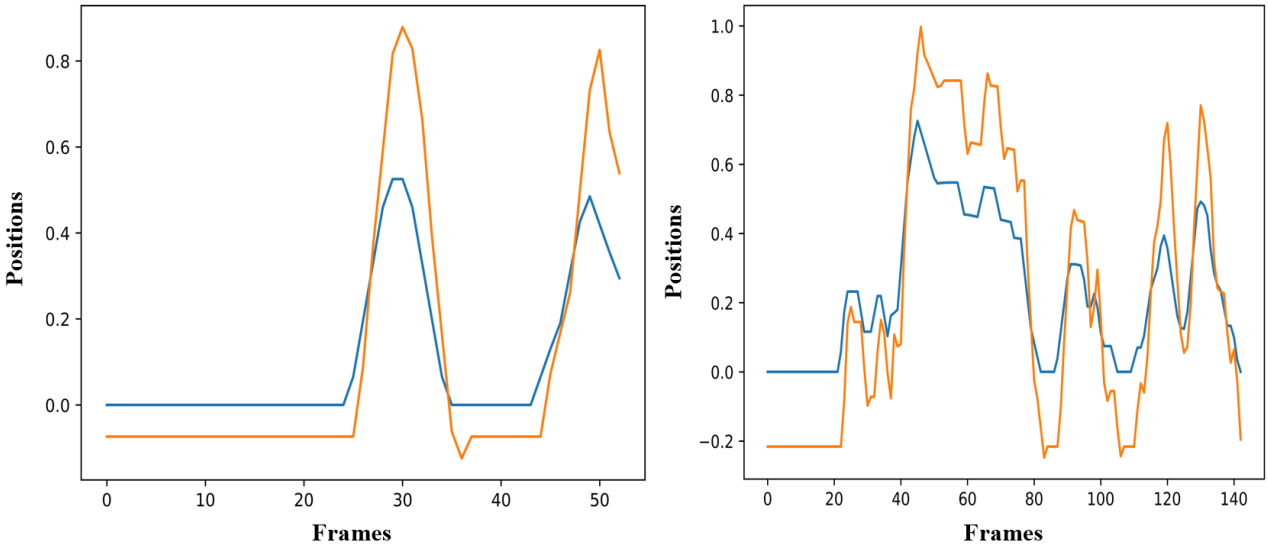


Figure 4: Examples of real motion observations of motion data (blue) and simulated values (orange) from the $RHAND_X$ State-Space model of gesture $G_{2,1}$ (left) and $G_{2,2}$ (right)

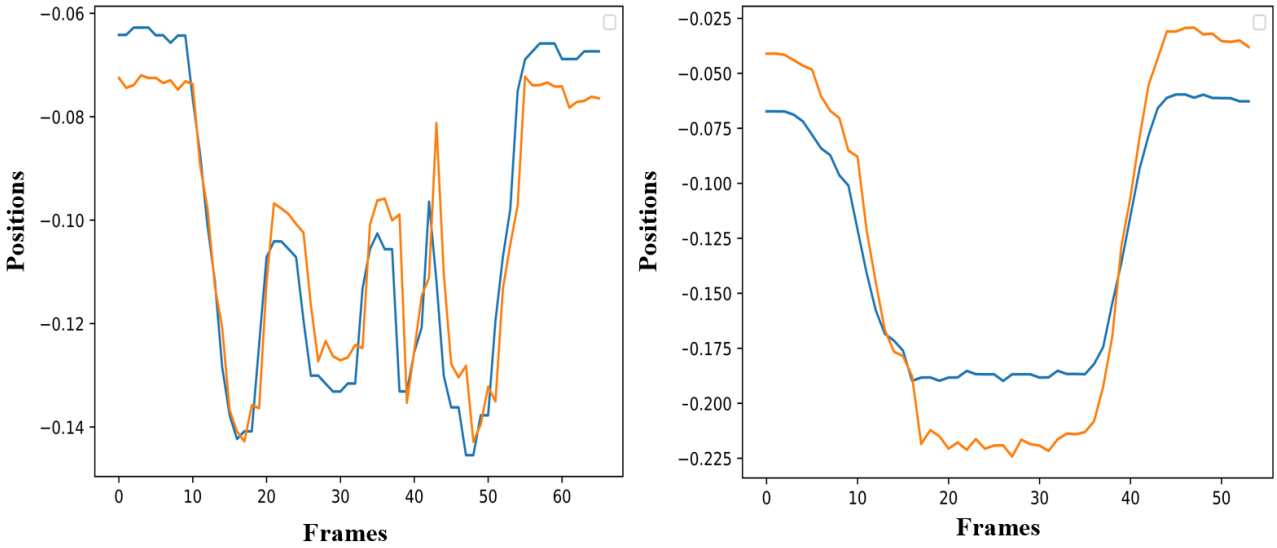


Figure 5: Examples of real motion observations (blue) and simulated values (orange) from the $RHAND_X$ State-Space model of the gesture $G_{3,1}$ (left) and $G_{3,4}$ (right)

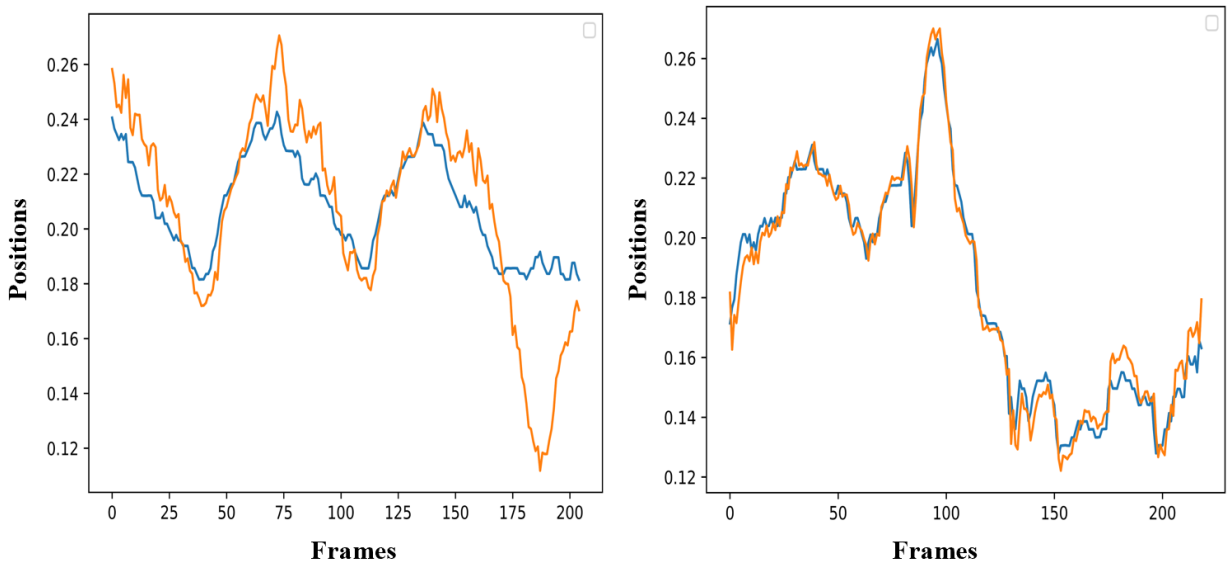


Figure 6: Trajectory forecasting

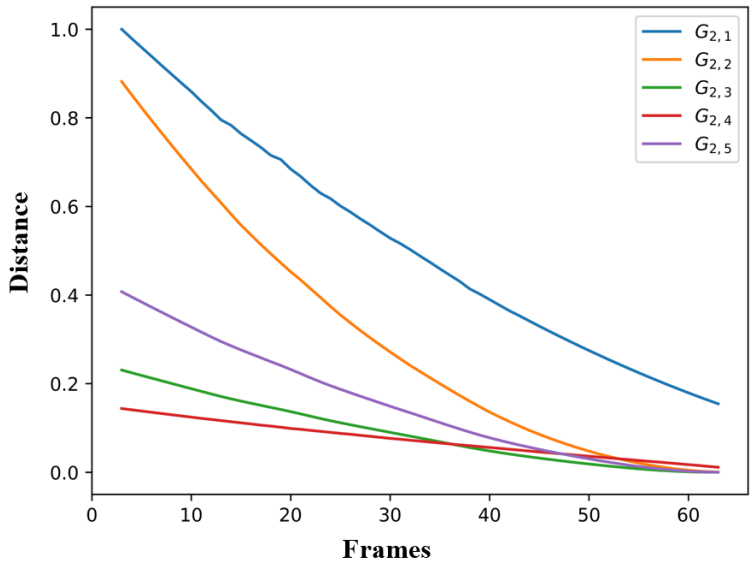
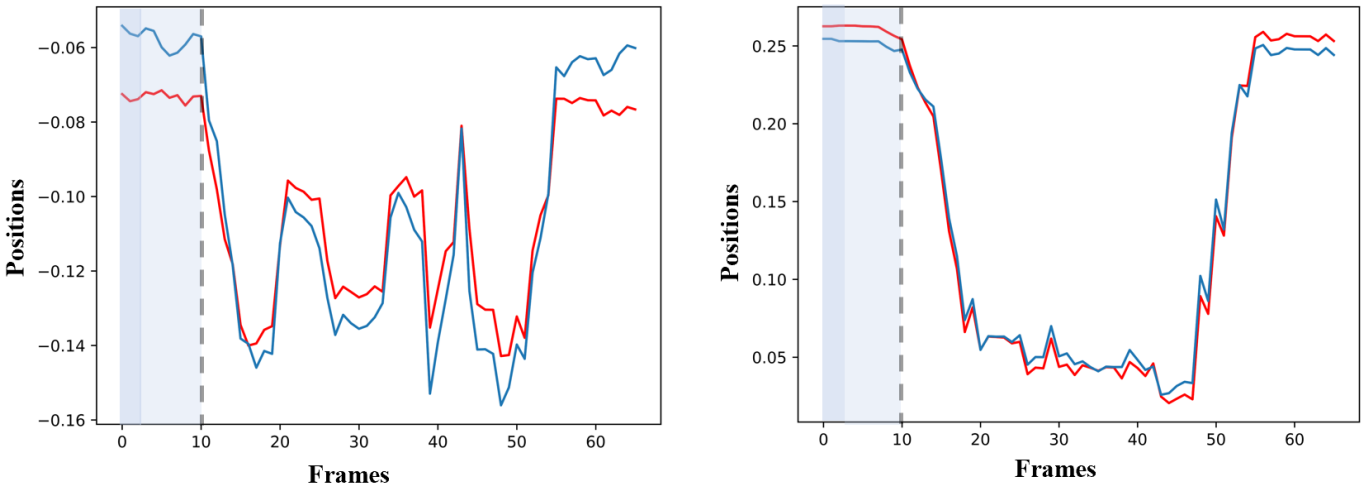


Figure 7: Left: Diagram of the simulated forecasted values of $RWRIST_X$ before the disturbance (red) and simulated forecasted values of $RWRIST_X$ (blue), after the shock on the values of $RWRIST_Y$ by 80% for two frames. Right: Diagram of the simulated forecasted values of $RWRIST_Y$ before the disturbance (red) and simulated forecasted values of $RWRIST_Y$ after the shock on the values of $RWRIST_X$ by 80% for two frames.



1.2 Supplementary tables

Table 1: RMSE between the iterations of the real data of GV_2 and GV_3

GV_2	$RMSE$
$\overline{G_{2,1}}$	0.0565
$\overline{G_{2,2}}$	0.0523
$\overline{G_{2,3}}$	0.0556
$\overline{G_{2,4}}$	0.0330
$\overline{G_{2,5}}$	0.0407
GV_3	$RMSE$
$\overline{G_{3,1}}$	0.0265
$\overline{G_{3,2}}$	0.0302
$\overline{G_{3,3}}$	0.0489
$\overline{G_{3,4}}$	0.0461

Table 2: Gesture vocabulary of TV assembling dataset, AGV commands dataset, Glassblowing and Human-robot collaboration dataset respectively

GV1 – TV assembly

$G_{1,1}$: Take the card from the left side box

$G_{1,2}$: Take the wire from the right-side box

$G_{1,3}$: Connect the wire with the card

$G_{1,4}$: Place the card on the TV chassis

GV2 - AGV commands

$G_{2,1}$: Hello

$G_{2,2}$: Left

$G_{2,3}$: Right

$G_{2,4}$: Speed up

$G_{2,5}$: Speed down

GV3 - Glassblowing

$G_{3,1}$: Fix details with pliers

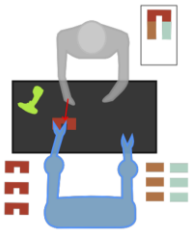
$G_{3,2}$: Tighten base of glass

$G_{3,3}$: Make shape with paper

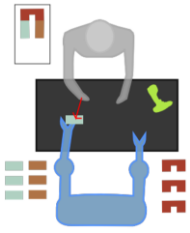
$G_{3,4}$: Fix shape

GV4 – Human-robot collaboration

$G_{4,1}$: Take a motor hose part in the robot right claw

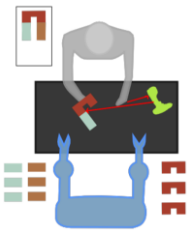


$G_{4,2}$: Take a motor hose part in the robot left claw



$G_{4,3}$: Join two parts of the motor hose

$G_{4,4}$: Screw



$G_{4,5}$: Put the final motor hose in a box

Table 3: Confusion matrix using HMM, HMM^{SS} and 3DCNN based on the model of Tran et al. [54] approaches for *GV1*

	$HMM_{1,1}$	$HMM_{1,2}$	$HMM_{1,3}$	$HMM_{1,4}$	Recall (%)
$G_{1,1}$	48	0	0	0	100
$G_{1,2}$	0	44	0	2	95.65
$G_{1,3}$	1	4	36	3	81.81
$G_{1,4}$	0	0	0	46	100
Precision (%)	97.9	91.66	100	90.2	
	$HMM_{1,1}^{SS}$	$HMM_{1,2}^{SS}$	$HMM_{1,3}^{SS}$	$HMM_{1,4}^{SS}$	Recall (%)
$G_{1,1}$	47	0	0	1	97.91
$G_{1,2}$	1	45	0	0	97.82
$G_{1,3}$	0	1	43	0	97.72
$G_{1,4}$	0	4	0	42	91.3
Precision (%)	97.91	90	100	97.67	
	$3DCNN_{1,1}$	$3DCNN_{1,2}$	$3DCNN_{1,3}$	$3DCNN_{1,4}$	Recall (%)
$G_{1,1}$	48	0	0	0	100
$G_{1,2}$	0	43	0	5	89.5
$G_{1,3}$	0	0	44	0	100
$G_{1,4}$	0	7	0	39	84.7
Precision (%)	100	86	100	88.6	

Table 4: Confusion matrix using HMM, HMM^{SS} and 3DCNN based on the model of Tran et al. [54] approach for *GV2*

	$HMM_{1,1}$	$HMM_{1,2}$	$HMM_{1,3}$	$HMM_{1,4}$	$HMM_{1,5}$	Recall (%)
$G_{2,1}$	14	1	0	1	0	87.5
$G_{2,2}$	3	13	0	0	0	81.25
$G_{2,3}$	0	0	16	0	0	100
$G_{2,4}$	5	0	0	11	0	68.75
$G_{2,5}$	2	0	0	2	12	75
Precision (%)	58.3	92.8	100	78.5	100	
	$HMM_{2,1}^{SS}$	$HMM_{2,2}^{SS}$	$HMM_{2,3}^{SS}$	$HMM_{2,4}^{SS}$	$HMM_{2,5}^{SS}$	Recall (%)
$G_{2,1}$	16	0	0	0	0	100
$G_{2,2}$	4	12	0	0	0	75
$G_{2,3}$	0	0	16	0	0	100
$G_{2,4}$	2	0	0	14	0	87.5
$G_{2,5}$	3	0	0	3	10	62.5
Precision (%)	64	100	100	82.3	100	
	$3DCNN_{2,1}$	$3DCNN_{2,2}$	$3DCNN_{2,3}$	$3DCNN_{2,4}$	$3DCNN_{2,5}$	Recall (%)
$G_{2,1}$	16	0	0	0	0	100
$G_{2,2}$	0	16	0	0	0	100
$G_{2,3}$	0	0	16	0	0	100
$G_{2,4}$	0	0	0	12	4	75
$G_{2,5}$	8	0	0	0	8	50
Precision (%)	66.6	100	100	100	66.66	

Table 5: Confusion matrix using HMM, HMM^{SS} and 3DCNN based on the model of Tran et al. [54] approach for GV_3

	$HMM_{3,1}$	$HMM_{3,2}$	$HMM_{3,3}$	$HMM_{3,4}$	Recall (%)
$G_{3,1}$	31	2	1	1	88.57
$G_{3,2}$	0	33	1	0	97.05
$G_{3,3}$	2	2	16	1	76.19
$G_{3,4}$	0	0	0	27	100
Precision (%)	93.93	89.18	88.88	93.1	
	$HMM_{3,1}^{SS}$	$HMM_{3,2}^{SS}$	$HMM_{3,3}^{SS}$	$HMM_{3,4}^{SS}$	Recall (%)
$G_{3,1}$	31	2	1	1	88.57
$G_{3,2}$	0	33	1	0	97.05
$G_{3,3}$	1	1	17	2	80.95
$G_{3,4}$	0	0	0	27	100
Precision (%)	96.87	91.66	89.47	90	
	$3DCNN_{3,1}$	$3DCNN_{3,2}$	$3DCNN_{3,3}$	$3DCNN_{3,4}$	Recall (%)
$G_{3,1}$	35	0	0	0	100
$G_{3,2}$	0	27	7	0	79.4
$G_{3,3}$	2	2	17	0	80.9
$G_{3,4}$	0	0	0	27	100
Precision (%)	94.5	93.1	70.8	100	

Table 6: Confusion matrix using HMM and HMM^{SS} approach for GV_4

	$HMM_{4,1}$	$HMM_{4,2}$	$HMM_{4,3}$	$HMM_{4,4}$	$HMM_{4,5}$	Recall (%)
$G_{4,1}$	42	1	0	0	1	95.4
$G_{4,2}$	3	86	0	0	1	95.5
$G_{4,3}$	0	0	75	2	12	84.2
$G_{4,4}$	1	0	0	43	0	97.7
$G_{4,5}$	2	0	3	0	75	93.7
Precision (%)	87.5	98.8	96.15	95.5	86.2	
	$HMM_{4,1}^{SS}$	$HMM_{4,2}^{SS}$	$HMM_{4,3}^{SS}$	$HMM_{4,4}^{SS}$	$HMM_{4,5}^{SS}$	Recall (%)
$G_{4,1}$	42	1	0	0	1	95.5
$G_{4,2}$	3	86	0	0	1	95.5
$G_{4,3}$	0	0	87	0	2	89.8
$G_{4,4}$	5	2	0	37	0	84
$G_{4,5}$	1	0	5	0	74	92.5
Precision (%)	82.3	96.6	94.5	100	94.8	

Table 7: Comparison of mean *f-scores* and final accuracies of each *GV* for HMM and $\mathrm{HMM}^{\mathrm{SS}}$ approach. For the datasets GV_1, GV_2 and GV_3 also the 3DCNN results are presented, based on the model of Tran et al. [55]. For GV_4 the results are compared to those presented in Coupeté et al. [54] using a k-means and HMM approach, with 25 clusters and discrete HMMs with 12 hidden states.

	Datasets				
Mean $f-score$		GV_1	GV_2	GV_3	GV_4
	HMM	94.34 %	83.1 %	90.64 %	92.1%
	$\mathrm{HMM}^{\mathrm{SS}}$	96.21 %	85 %	91.57 %	92.29%
	$\mathrm{k-means}\ +\mathrm{HMM}$	-	-	-	80%
	3DCNN	93.4 %	84%	90%	-
	Datasets				
Total accuracy		GV_1	GV_2	GV_3	GV_4
	HMM	94.56%	82.5%	91.45%	92.5%
	$\mathrm{HMM}^{\mathrm{SS}}$	96.19%	85%	92.3%	93.94%
	$\mathrm{k-means}\ +\mathrm{HMM}$	-	-	-	82%
	3DCNN	93.5%	87%	89%	-

Table 8: Theil inequality coefficient, root mean squared error, mean absolute error, mean absolute percentage error for one example of the X coordinate of the right wrist per dataset

Gestures	Theil Inequality U	Bias proportion U^B	Variance proportion U^V	Covariance proportion U^C	RMSE
$G_{1,1}$	0.018388	0.009178	0.081456	0.909366	0.028904
$G_{2,1}$	0.0000373	0	0.017247	0.983653	0.007461
$G_{3,1}$	0.0000161	0	0.008713	1.041715	0.003277
$G_{4,1}$	0.010059	0	0.039551	0.960449	0.018053