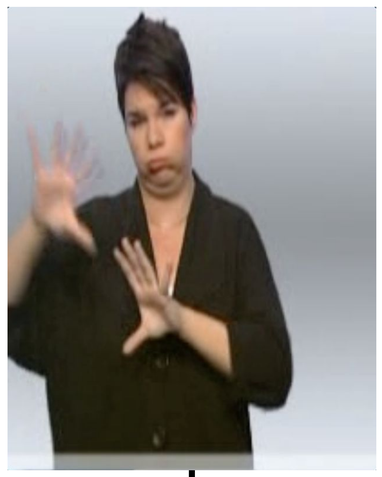
# Evaluation of Pose Estimation Systems for Sign Language Translation

Catherine O'Brien*, Gerard Sant*, Mathias Müller, Sarah Ebling
Department of Computational Linguistics | University of Zürich

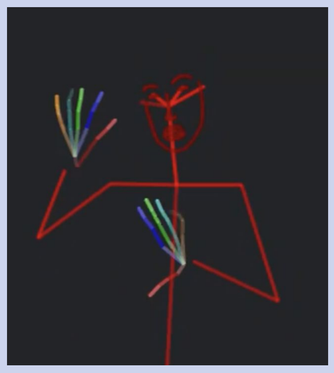
*These authors contributed equally

## Which estimator yields the best translations?

**Original video frame from PHOENIX-2014**



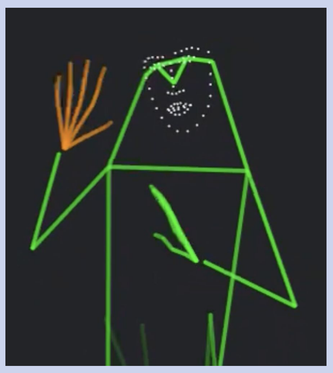
**MediaPipe**



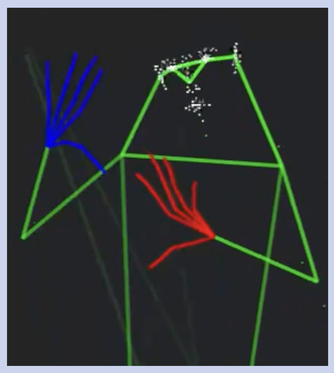
**OpenPose**

**MMPose Wholebody**

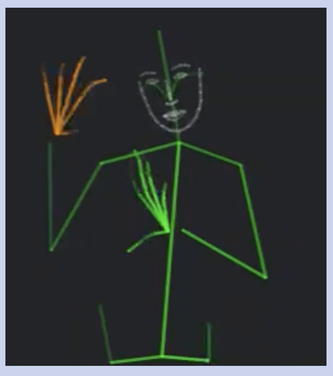
**OpenPifPaf**



**SDPose**



**Sapiens**



**AlphaPose**

**SMPLest-X**

## Translation Scores

**Best BLEU**

**Sapiens and SDPose**

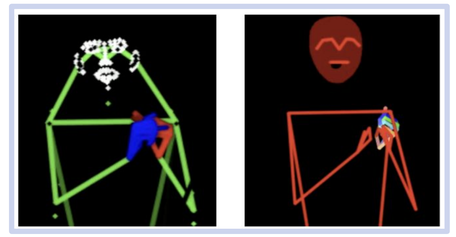
**Best BLEU With Lower Compute**

**AlphaPose and MMPose**

| Estimator | BLEU (↑) | speed (fps) |
|---|---|---|
| MediaPipe | $10.327 \pm 0.269$ | 0.89 / 3.15 |
| OpenPose | $10.606 \pm 0.251$ | - / 4.40 |
| MMPose Wholebody | $10.901 \pm 0.299$ | 0.89 / 3.81 |
| OpenPifPaf | $9.365 \pm 0.263$ | 1.21 / 4.42 |
| SDPose | $11.681 \pm 0.415$ | 0.07 / 0.84 |
| Sapiens | $11.525 \pm 0.222$ | 0.04 / 3.29 |
| AlphaPose | $11.251 \pm 0.241$ | - / 22.89 |
| SMPLest-X | $9.709 \pm 0.334$ | - / 8.36 |

## Occlusion

| estimator | correctness (%) |
|---|---|
| Mediapipe | 73.33 |
| OpenPose | 66.66 |
| MMPose Wholebody | 33.33 |
| OpenPifPaf | 6.66 |
| SDPose | 46.66 |
| Sapiens | 100.00 |
| AlphaPose-136 | 40.00 |
| SMPLest-X | 0.00 |

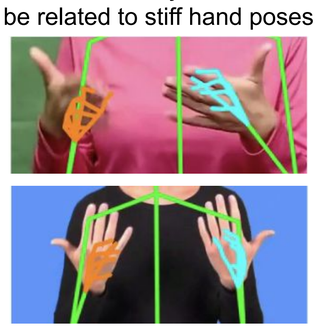


Sapiens MediaPipe

Prior research shows that MediaPipe struggles to handle occlusion. When checking the different estimators on 15 instances of occlusion, it appeared that **several estimators (notably, Sapiens) may handle occlusion better than MediaPipe.**

## Jitter Analysis

SMPLest-X's low jitter score may be related to stiff hand poses



| | Phoenix | SignSuisse |
|---|---|---|
| MediaPipe | **2.46** (2.01–2.81) | 6.51 (5.10–8.38) |
| OpenPose | 6.84 (5.90–8.18) | 26.36 (19.57–29.55) |
| MMPose Wholebody | 8.58 (7.32–10.35) | 14.47 (11.66–18.59) |
| OpenPifPaf | 8.94 (6.35–10.20) | 15.35 (13.12–18.63) |
| SDPose | 14.97 (10.70–21.79) | 13.24 (10.80–17.10) |
| Sapiens | 7.77 (6.62–9.16) | 10.86 (9.62–12.88) |
| AlphaPose | 7.36 (5.15–11.50) | 10.32 (8.35–12.47) |
| SMPLest-X | 2.74 (2.16–3.14) | **3.97** (3.19–4.85) |

Jerk scores represent the the third-order temporal difference, penalizing rapid changes in acceleration. Lower scores indicate smoother poses.

## Missing Hands

(a) Original frame (b) Sapiens (c) OpenPifPaf (d) OpenPose (e) MediaPipe

| estimator | left | right | both |
|---|---|---|---|
| MediaPipe | 20.22 | 23.61 | 8.84 |
| OpenPose | 8.43 | 3.88 | 0.22 |
| OpenPifPaf | 67.65 | 59.19 | 40.63 |

On an analysis of the SignSuisse dataset, only MediaPipe, OpenPose, and OpenPifPaf ever have an entirely missing hand (a confidence score of 0 for >50% of hand keypoints)

University of Zurich[UZH]

poster paper pipelines repo video-to-pose repo