

## Hybrid Skeleton Driven Surface Registration for Temporally Consistent Volumetric Video

João Regateiro, Marco Volino and Adrian Hilton

Centre for Vision, Speech and Signal Processing  
University of Surrey

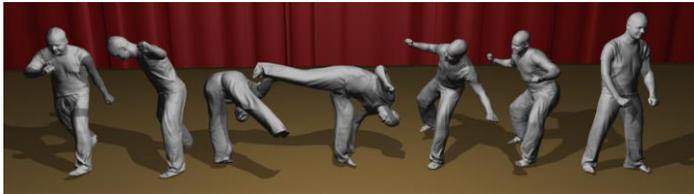
# Motivation

---

# Motivation

## 3D video capture

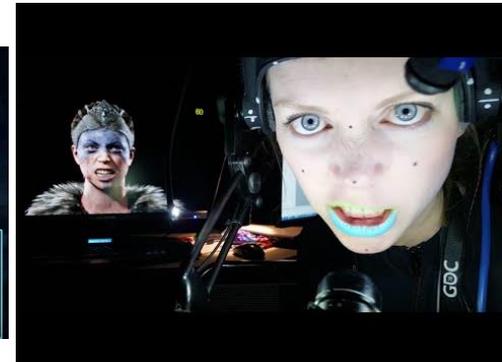
### Performance capture



### Film and animation



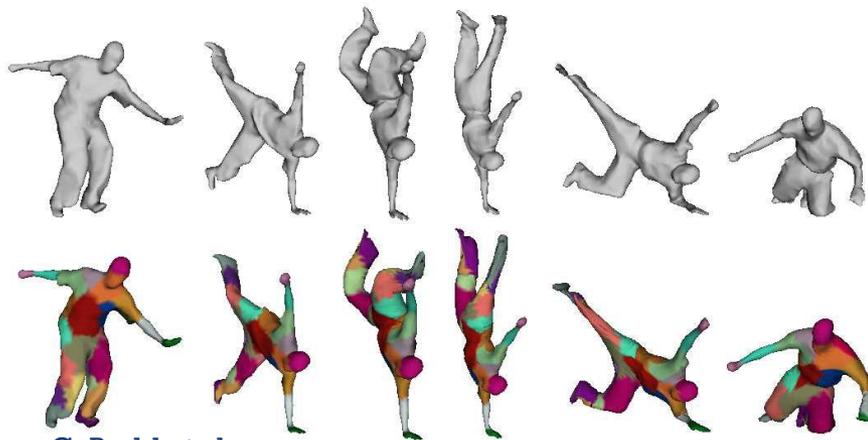
### Games



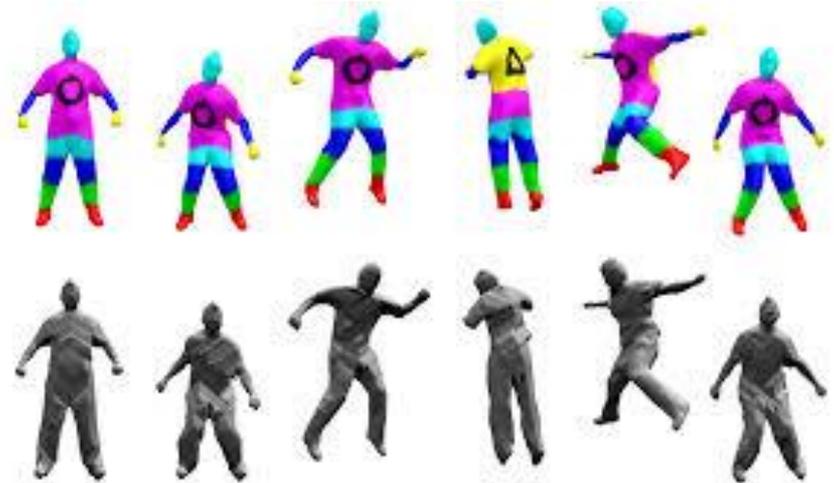
# Temporally Consistent Volumetric Video

## Alignment background

3D reconstruction with temporal coherence remains a challenge



C. Budd et al.



Cagniard et al.



Collet et al.



W. Xu et al.

# Temporally Consistent Volumetric Video

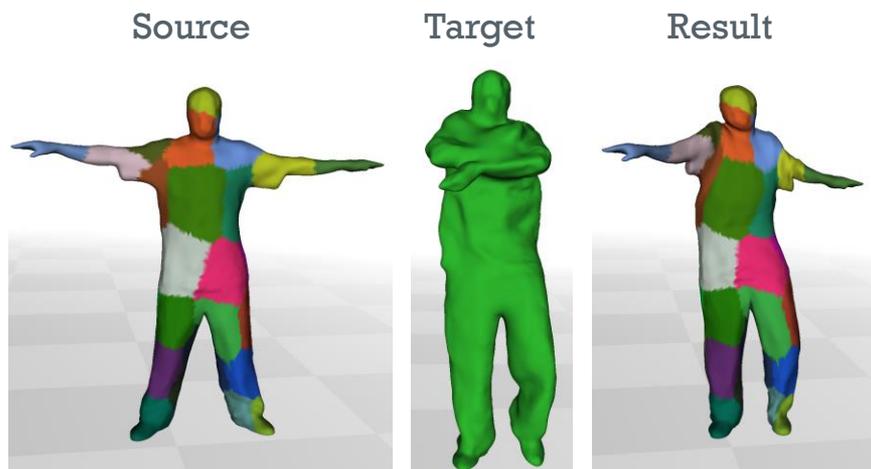
## Problems

### Model-free sequential alignment

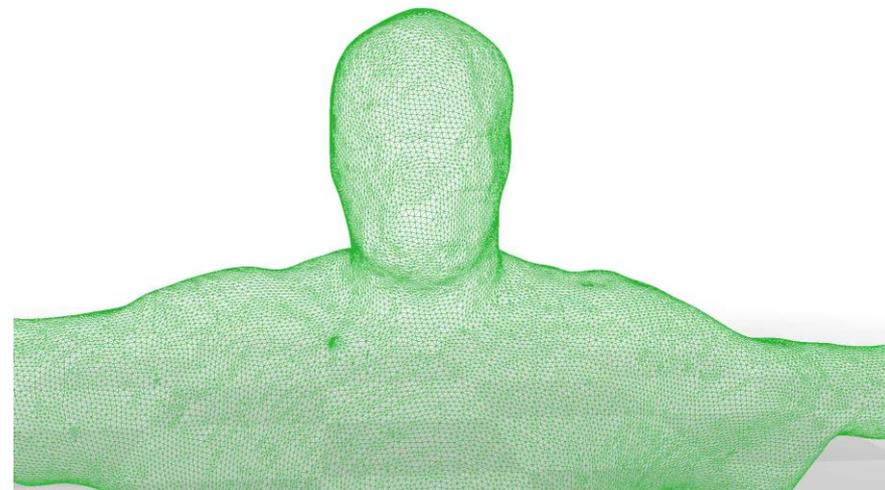
- Accumulation of errors resulting in drift in correspondences over time
- Gross-errors for large non-rigid deformations
- Do not allow alignment across multiple sequences

### Model-based alignment

- Consistent structured representation
- Dynamic surfaces and detail are not preserved



Inconsistent topology



# Approach

---

# Approach

## Input data

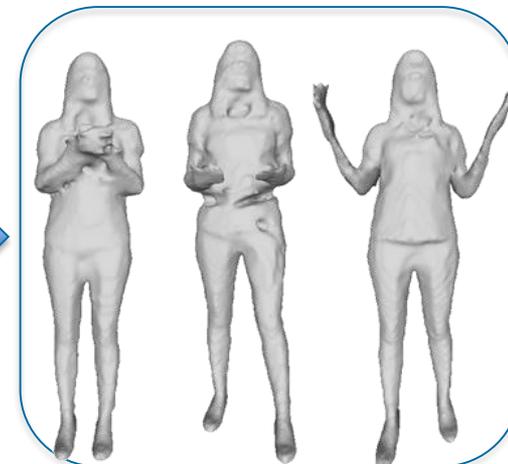
### Performance capture system [Starck and Hilton]

- Calibrated cameras
- Shape reconstruction via visual hull refinement
- Temporal inconsistent 3D mesh surfaces

## Visual hull



## Stereo reconstruction



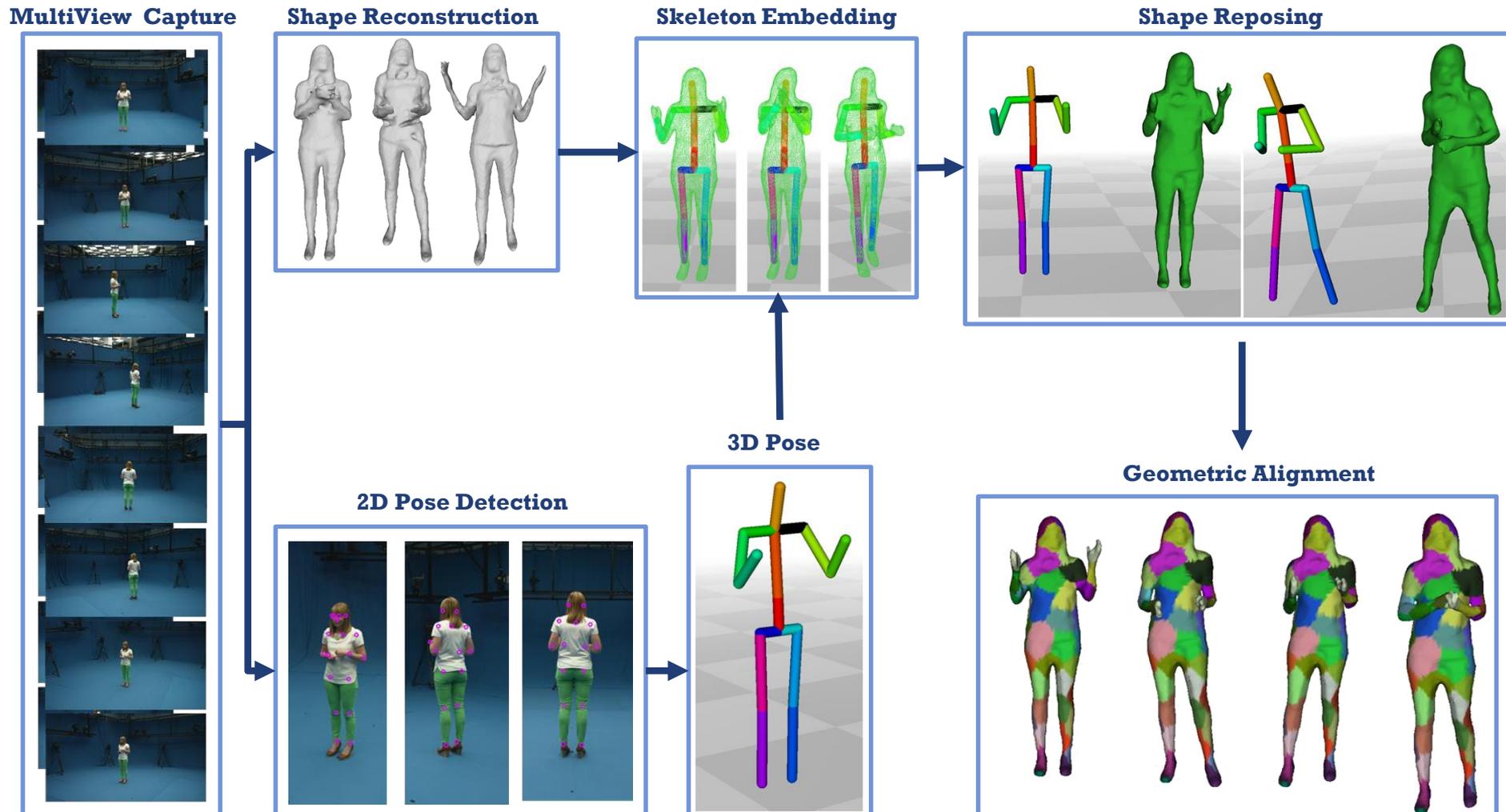
## 2D pose detection

- Single view video
- State-of-the-art convolutional pose machine (CPM) detector [Z. Cao et al].
  - Labelled keypoint position
  - Detection confidences



# Overview

## Hybrid skeleton driven surface registration pipeline

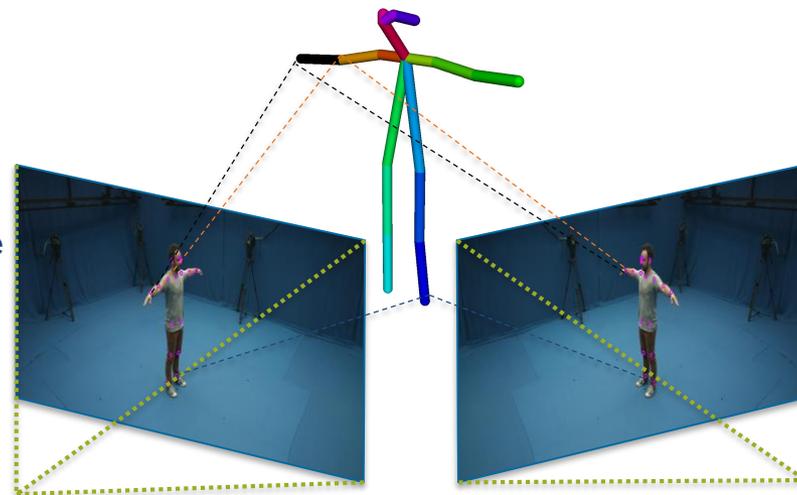


# Overview

## Pose estimation

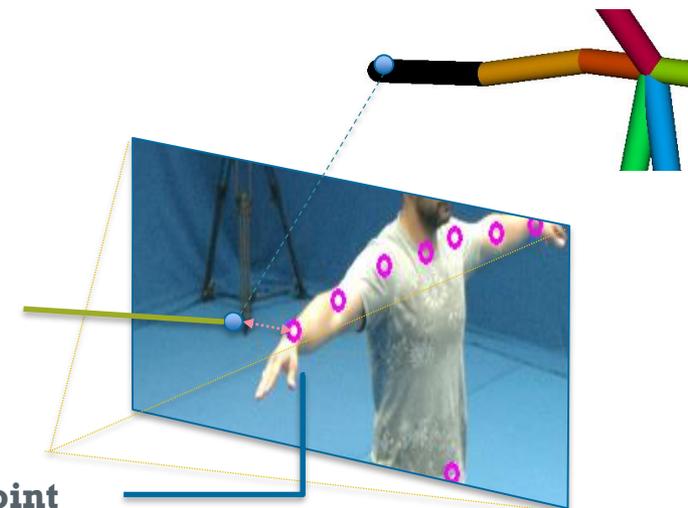
### 2D pose detector

- CPM [Z. Cao et al.] detector applied to a multiple view camera system.



### 3D pose triangulation

- Projection of 2D key point into the 3D space
- Minimizes the distance between the projected 3D location and the 2D key point detection.



$$\underset{j_i}{\operatorname{argmin}} \sum_{c=1}^{N_c} \omega_i^c |P(c, j_i) - p_i^c|$$

┌ Camera  
└ 3D Projected target  
└ 2D key point  
└ Joint confidence

# Overview

## Shape reposing

### Skeleton embedding

- Skin attachment weights are given by Pinocchio frame [I. Baran et al.].
- State-of-the-art Linear Blend Skinning (LBS) for mesh deformation.

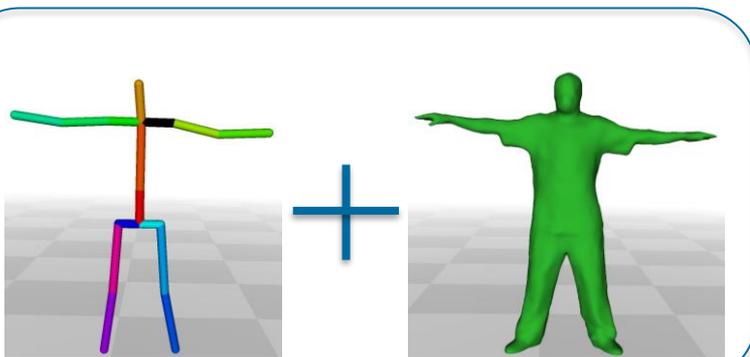
### Reference frames

- LBS results in a reference frame for every frame of the sequence.
- Provides a closer approximation of the desired target pose.

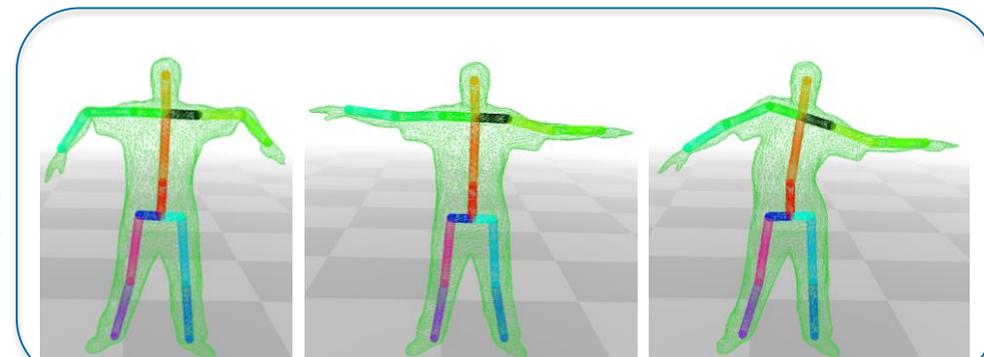
Reference Skeleton

Reference Mesh

LBS Results



=



# Patch-based iterative closest point

## Correspondence search

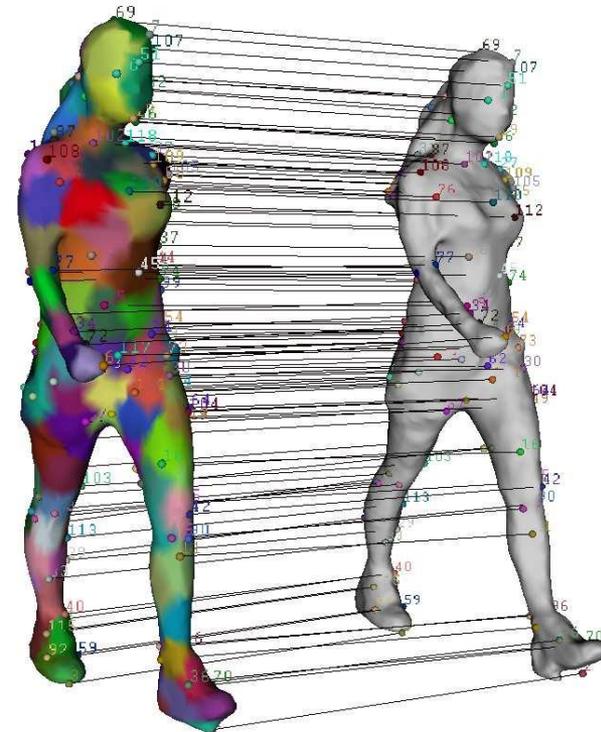
**Geodesic lloyd's algorithm** [C. Budd et al., C. Cagniart et al., S.Lloyd]

- Usability for any surface shape
- Fully automated
- Provides even segmented patches



**Iterative closest point (ICP)** [S.Rusinkiewicz et al., Z. Zhang, C. Budd et al.]

- ICP to **solve** a matching problem.
- Data has no prior correspondences.



# Geometry alignment

## Laplacian deformation framework

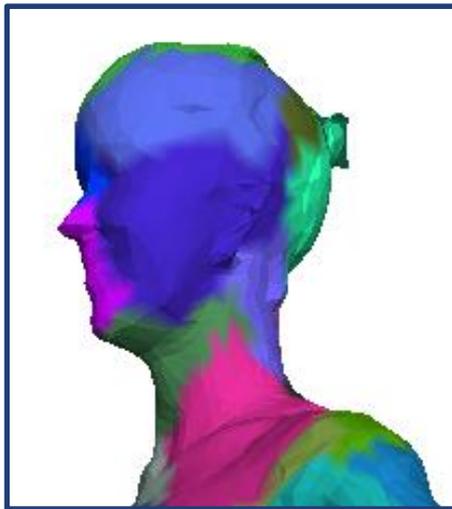
### Input

- Result from patch-based ICP
- Target surface mesh

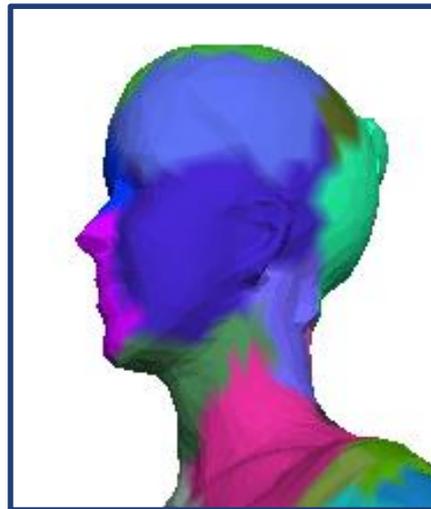
### Laplacian framework [Y. Lipman et al., L. Yaron et al.]

- Regularizes source surface
- Preserves fine surface detail and geometric shape

ICP Result



Laplacian Result



$$\vec{v}_u = \operatorname{argmin}_v E_r + E_c$$



$$E_r = \|Lv_u - \delta(v_k)\|^2$$

$$E_c = \|W_c(v_u - v_k)\|^2$$

# Results

---

# Results

## Vlasic samba

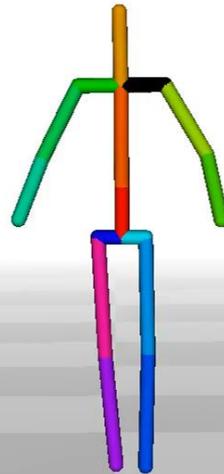
**Template**



**Reconstruction**



**Skeleton**



**LBS**



**Proposed**



## SurfCap street dance

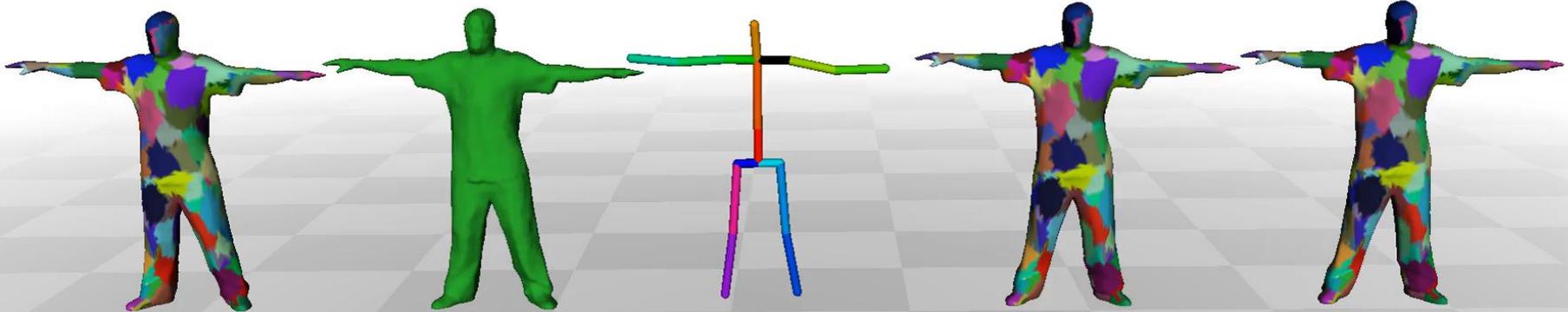
**Template**

**Reconstruction**

**Skeleton**

**LBS**

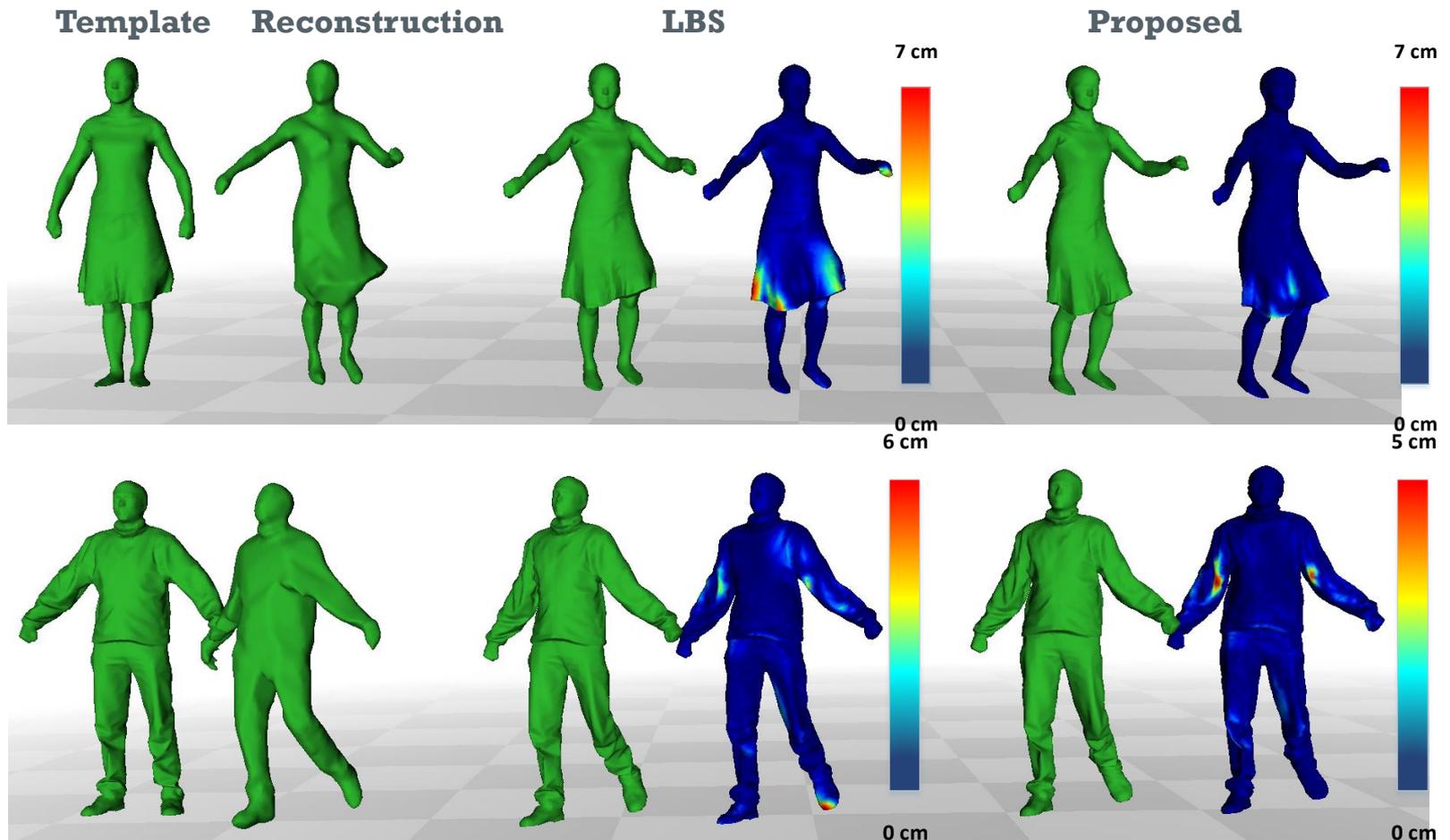
**Proposed**



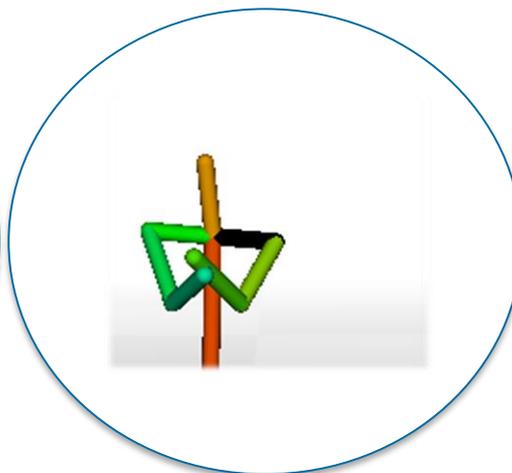
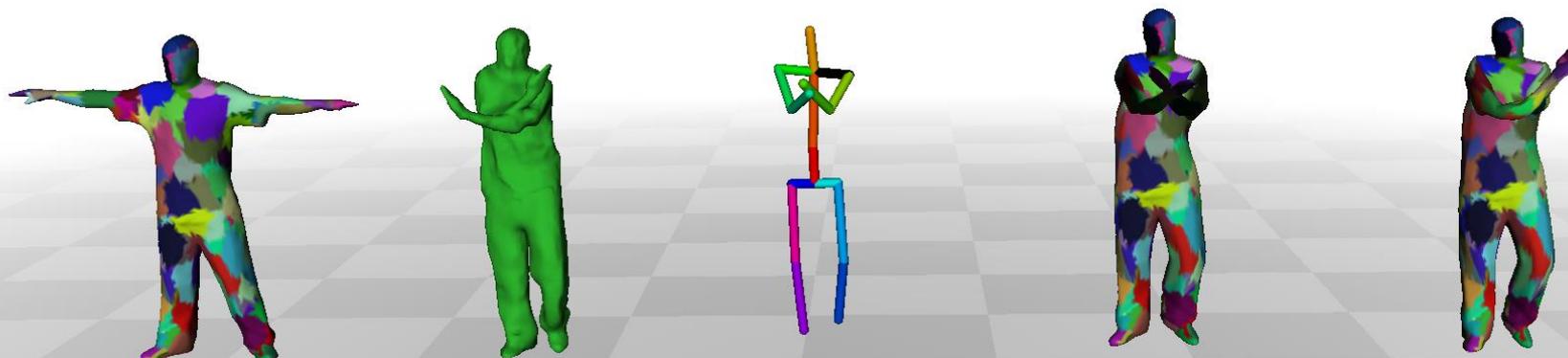
# Results

## Evaluation

The following evaluation is performed using one-sided **Hausdorff** distance



# Limitations



# Applications

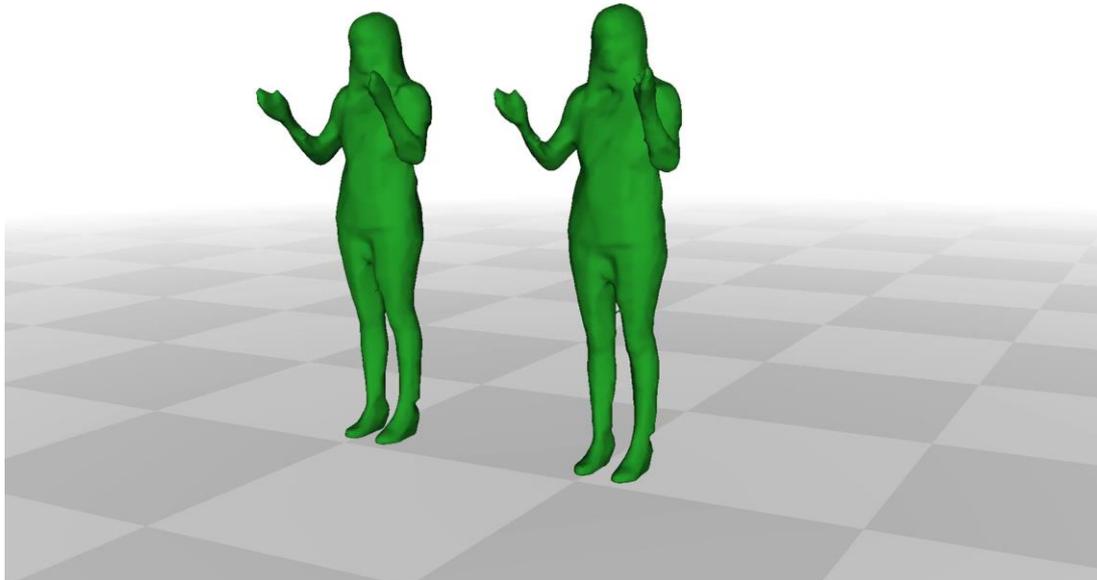
---

## Mesh compression

- Temporally aligned meshes reduce storage and transmission requirements

## Editing of volumetric video

- Allows artist to manipulate volumetric video in intuitive way
- Creation of novel sequence without having to re-capture the scene



# Conclusion

---

# Conclusions

## **Hybrid skeleton driven surface registration**

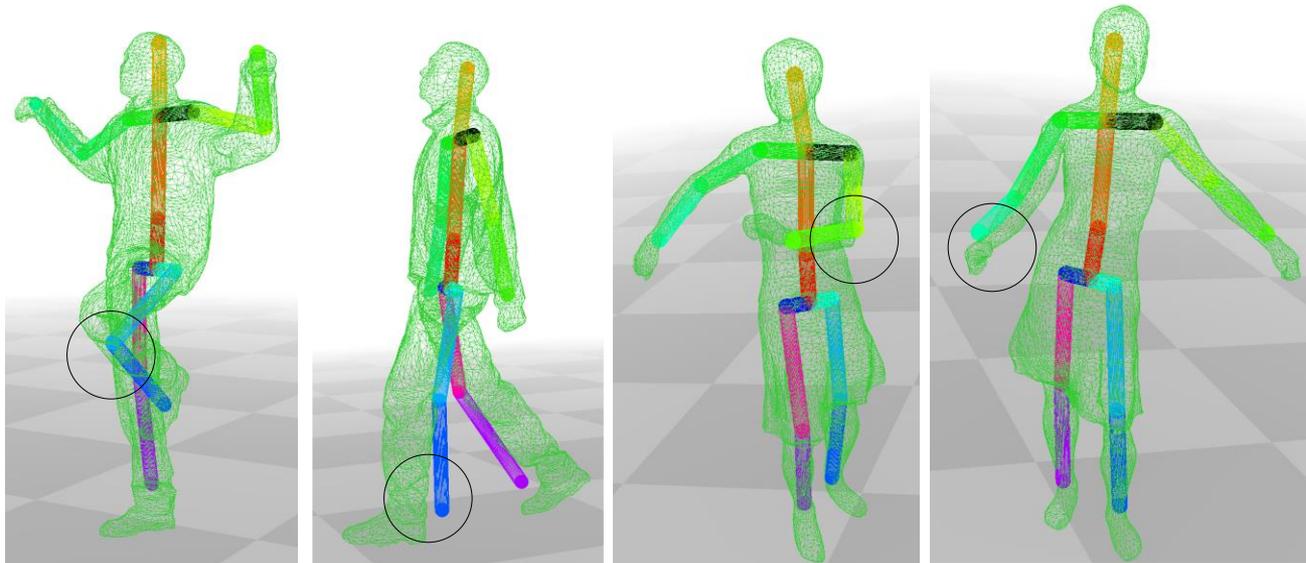
- Unconstrained skeleton pose tracking
- Skeletal tracking of rapid motion and large deformation
- Preserves dynamics and surface details
- Temporally aligned surfaces for complex dynamic sequences

## **Key frame-based editing for volumetric video**

- Skeleton manipulation
- Creates novel sequences and motion

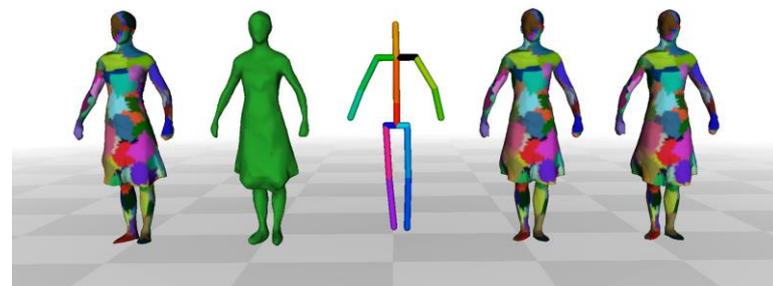
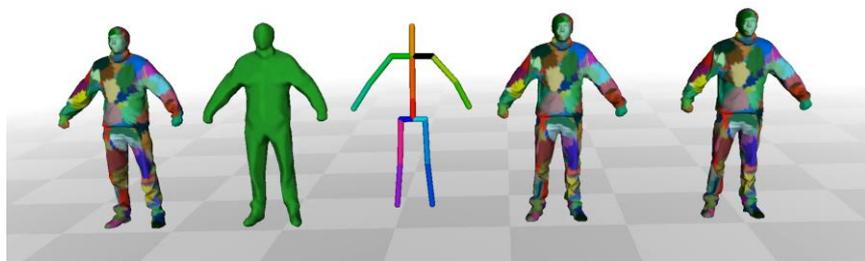
## Future work

- Improve on skeleton tracking, and human body kinematics
- Extending this approach for multiple subjects
- Intuitive editing of volumetric video



# Temporally Consistent 3D Video

## Questions?



Project webpage: [cvssp.org/projects/4d/HSDSR/](http://cvssp.org/projects/4d/HSDSR/)