Motion Capture with Linear Blend Skinning (LBS)

王雁刚 副教授
东南大学
www.yangangwang.com

About me

姓　　名：　王雁刚
工作单位：　东南大学 － 自动化学院
职　　称：　副教授
毕业院校：　清华大学
个人主页：　www.yangangwang.com
联系电话：　15110264815
电子邮件：　yangangwang@seu.edu.cn

学习及工作经历：

- 2017.09 － 现在：　东南大学自动化学院，副教授
- 2014.07 － 2017.08：微软亚洲研究院，副研究员
- 2012.10 － 2013.03：瑞士洛桑联邦理工大学，访问学者
- 2009.09 － 2014.07：清华大学，博士
- 2005.09 － 2009.07：东南大学，学士

Today

- I What is Motion Capture?
- II Linear Blend Skinning (LBS)
- III Motion Capture with LBS
- IV Advanced Techniques
- V Conclusion

Today

- I What is Motion Capture?
- II Linear Blend Skinning (LBS)
- III Motion Capture with LBS
- IV Advanced Techniques
- V Conclusion

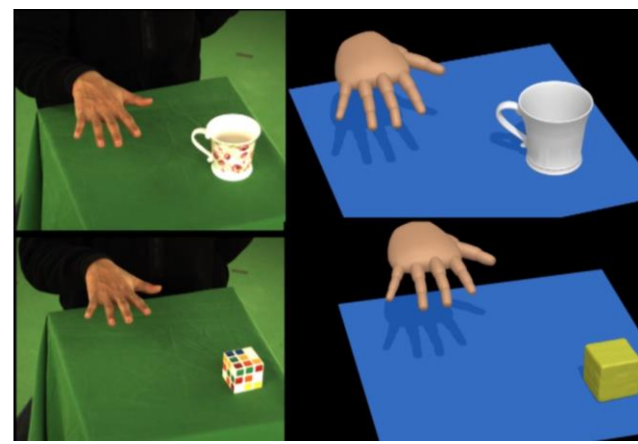
What is Motion Capture?

"… is a technique of digitally recording 3D human movements for immediate or delayed analysis and playback…"

Wikipedia

Full body

Fine detail face



Complex interaction

Applications of Motion Capture

Film industry

Robotics

Biomechanics

Video games

Natural user interaction

Virtual reality

Sensor & Maker System

- Pros
  - accurate
  - real time and robust
- Cons
  - sparse tracker and low spatial resolution
  - expensive hardware
  - cumbersome to wear special suits

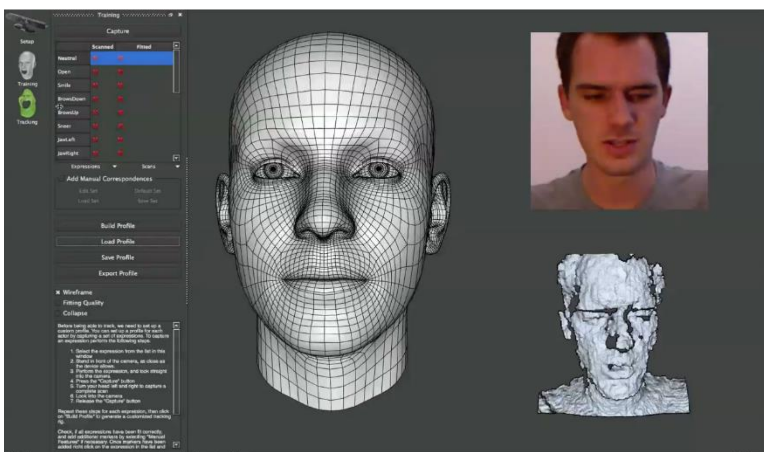
Electromechanical system

Hand mocap system

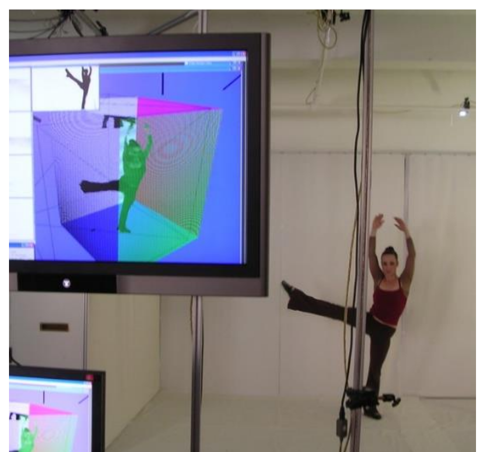
Marker system

Video-based System

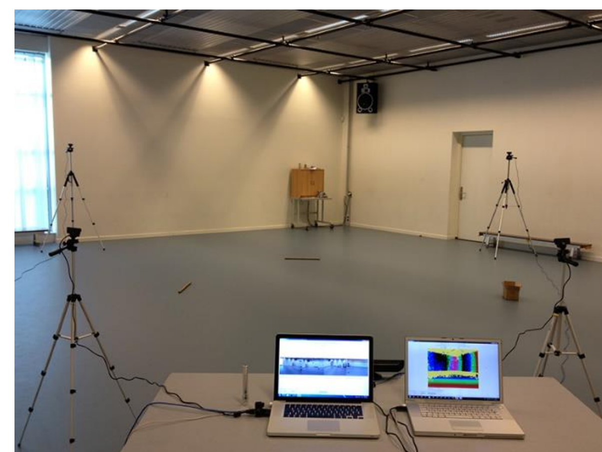
- Pros
  - no markers, no sensors or no special suits
  - small and inexpensive hardware configurations
  - high spatial resolution
- Cons
  - not as accurate as marker systems
  - not as stable as marker and sensor systems



Faceshift®



Organic Motion®



iPiSoft®

Challenges for Video-based System

- Capture high spatial resolution motion
  - few discernible features
  - complex background
- Difficult and nontrivial temporal tracking
  - significant occlusions
  - lighting changes

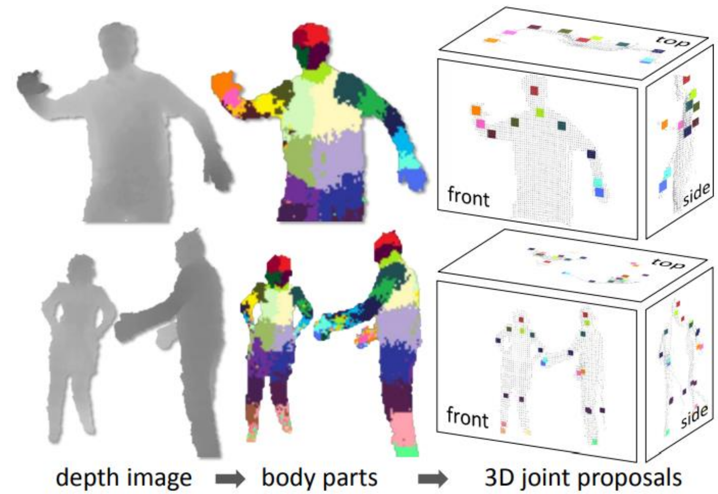
Few features

Occlusions

nontrivial temporal tracking

Main Solutions

Generative Method



Discriminative Method

How to perform ***model deformation***?

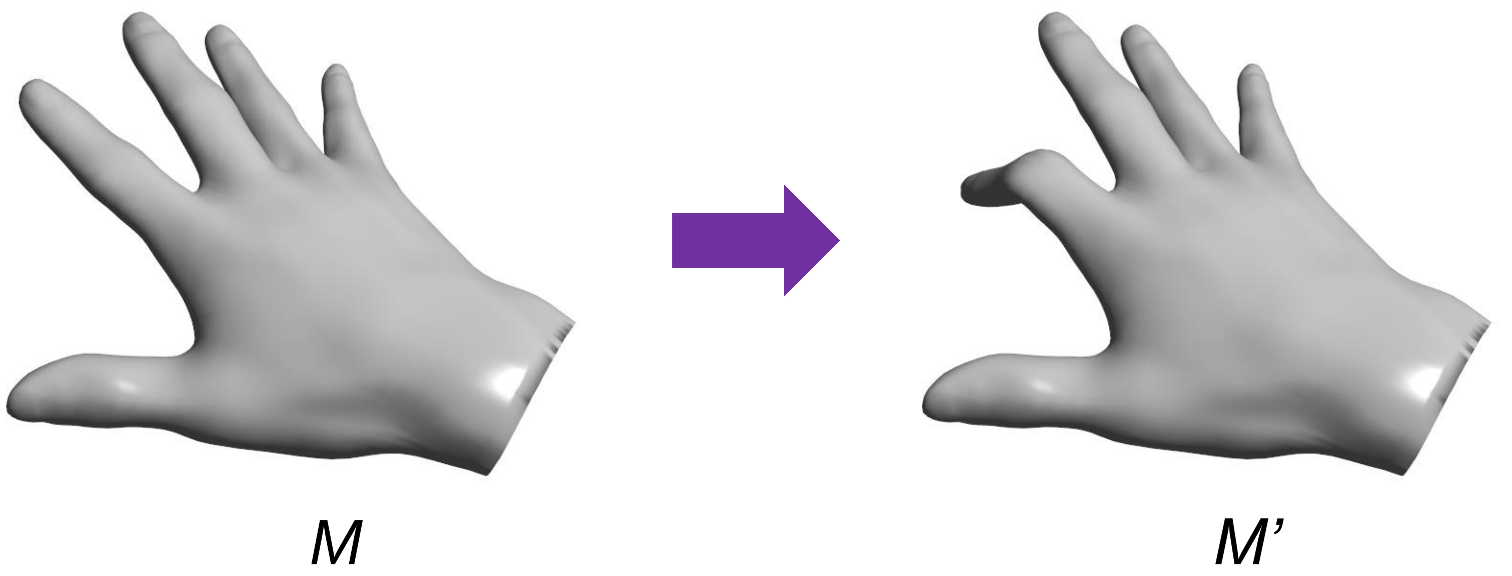
Hybrid Method

Today

- I What is Motion Capture?
- II Linear Blend Skinning (LBS)
- III Motion Capture with LBS
- IV Advanced Techniques
- V Conclusion

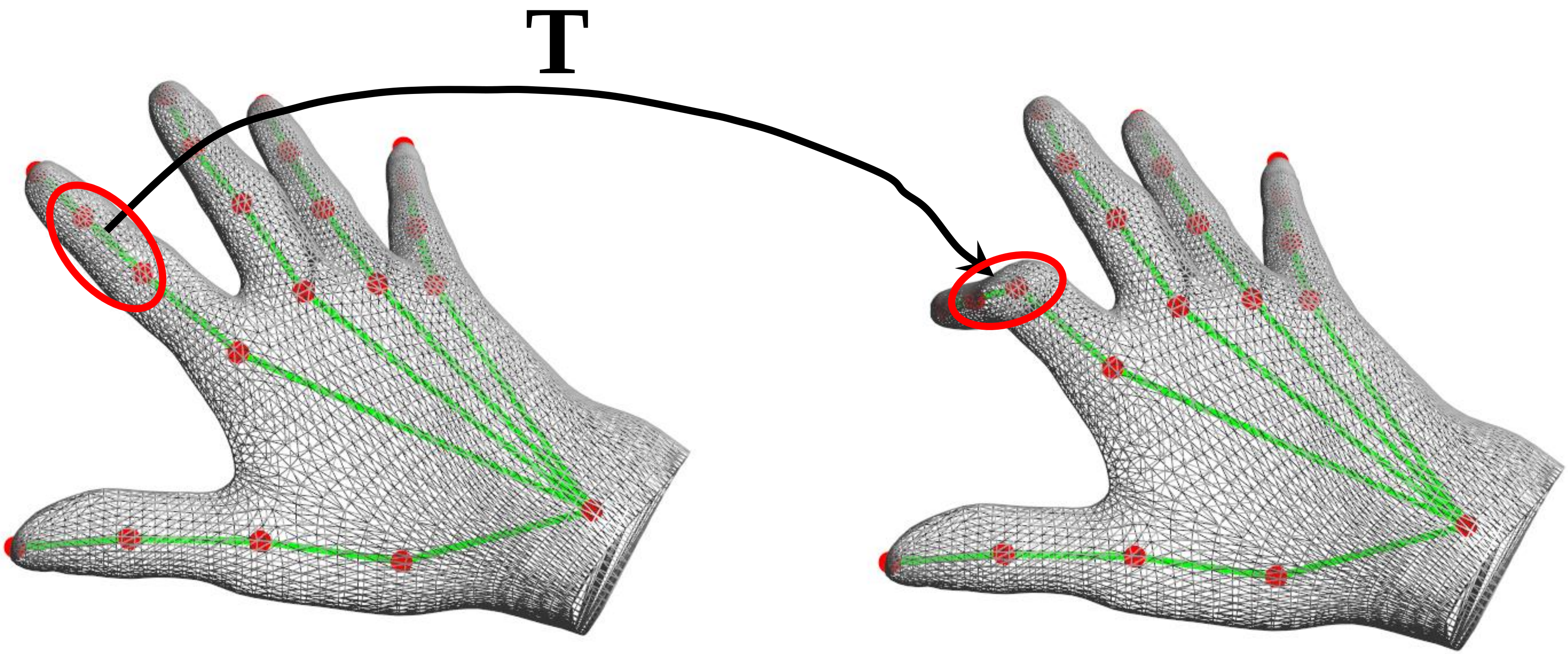
Articulate Model Deformation

- Consider the following problem
  - we want to deform the mesh model *M* into mesh model *M'*
  - implicit skeleton structure
- Articulate shape deformation is very common in real life



Basic Concepts

- Skinning transformations

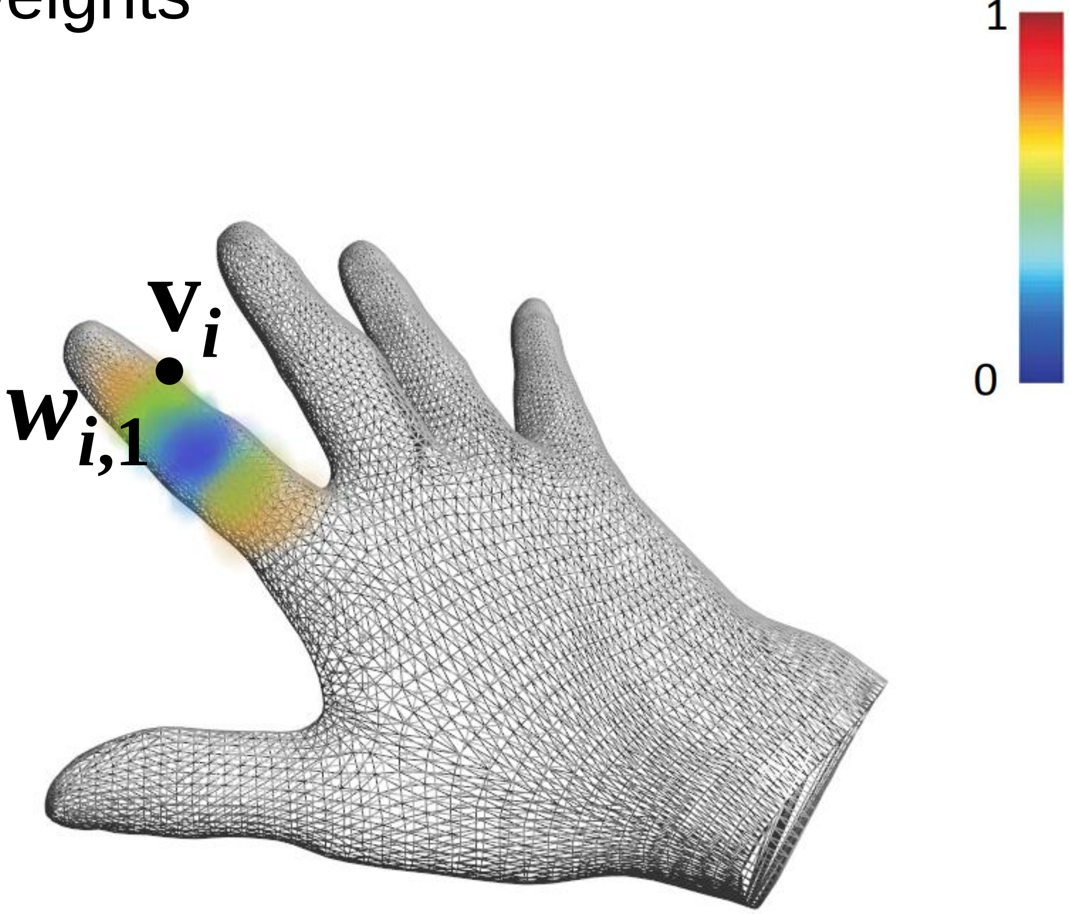


Rest pose *M*

Target pose *M'*

Basic Concepts

- Skinning weights



Linear Blend Skinning (LBS)

$$v_i' = \sum_{j=1}^{m} w_{i,j}\mathbf{T}_j v_i = \left(\sum_{j=1}^{m} w_{i,j}\mathbf{T}_j\right) v_i$$

Vertex in model *M'*

Vertex in model *M*

Skinning weights

Skinning transformation

Linear Blend Skinning (LBS)

$$v_i' = \sum_{j=1}^{m} w_{i,j}\mathbf{T}_j v_i = \left(\sum_{j=1}^{m} w_{i,j}\mathbf{T}_j\right) v_i$$

- How to compute the Skinning weights $w_{i,j}$?
  - "Automatic Rigging and Animation of 3D Characters", SIGGRAPH 2007

http://www.mit.edu/~ibaran/autorig/pinocchio.html

  - Key Idea
    - heat equilibrium to find the weights

$$-\Delta w_j + H w_j = H p_j$$

Today

- I What is Motion Capture?
- II Linear Blend Skinning (LBS)
- III Motion Capture with LBS
- IV Advanced Techniques
- V Conclusion

Motion Capture with LBS

Search optimal skinning transformations $\mathbf{T}$ to drive the articulate mesh model with $lbs(\mathbf{T})$ to best match the observed image data O

$$\min_T E\left(lbs\left(T\right), O\right)$$

- How to perform the skinning transformations T to drive the articulate mesh model M ?
- How to measure the inconsistency between the deformed model and observed image data?
- How to find an optimal skinning transformations?

Motion Capture with LBS

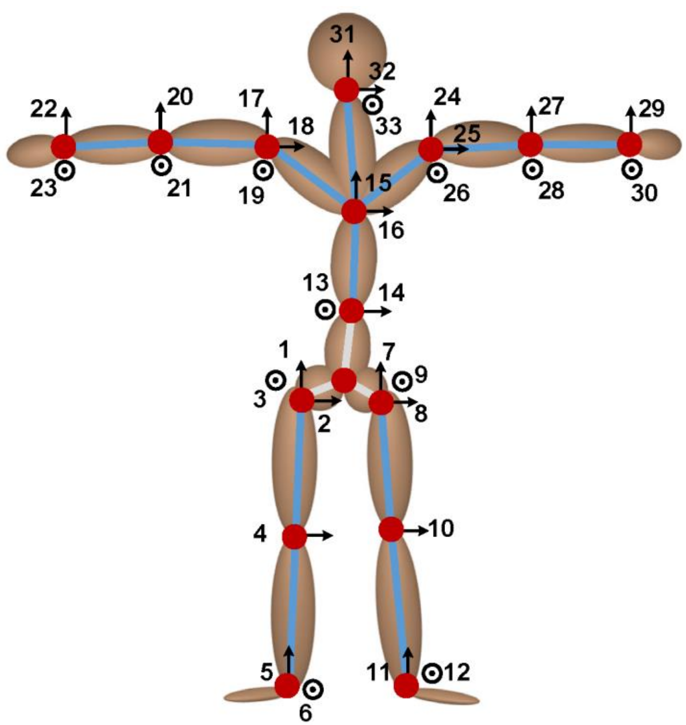
Search optimal skinning transformations $\mathbf{T}$ to drive the articulate mesh model with $lbs(\mathbf{T})$ to best match the observed image data O

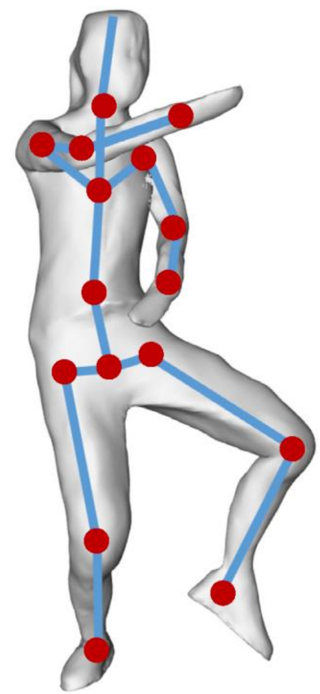
$$\min_{T} E\left(lbs\left(T\right), O\right)$$

- How to perform the skinning transformations *T* to drive the articulate mesh model *M* ?
- How to measure the inconsistency between the deformed model and observed image data?
- How to find an optimal skinning transformations?

Skinning Transformations for Articulate Pose

- Articulate pose is represented as **tree structure**
- N skinning transformations: $(\mathbf{T_1}, \mathbf{T_2}, \ldots, \mathbf{T_n})$
  - Each $\mathbf{T_i}$ joint has several Degree of Freedom (DoFs)
  - All children will be influenced by their parents

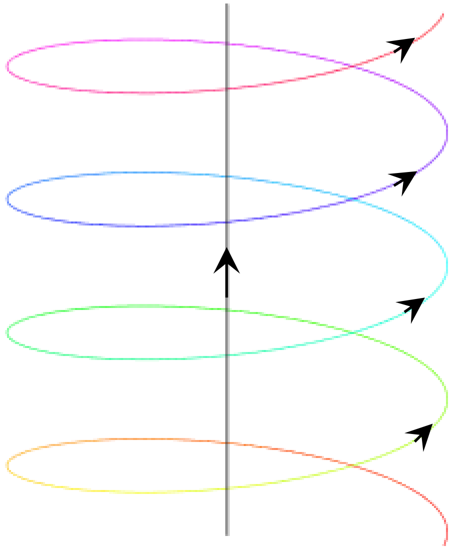




Skinning Transformations for Articulate Pose

- Skinning transformations for rigid body movement
  - Rotation + translation, *a.k.a*, **screw displacement**
  - A screw is an 6-dimension ordered pair

$$S = (\mathbf{S}, \mathbf{V})$$



- **Twist** of rigid body (actually, twist is a screw)

$$T = (\vec{\omega}, \mathbf{v} + \mathbf{d} \times \vec{\omega})$$

$\vec{\omega}$ is the angular velocity vector

$\mathbf{d}$ is the translation vector

$\mathbf{v}$ is the velocity of the origin of the moving frame

https://en.wikipedia.org/wiki/Screw_theory

Skinning Transformations for Articulate Pose

- **Revolute joints** in the article pose
  - Joint position is $\mathbf{P}$, one of rotation axis is $\vec{\boldsymbol{\omega}}$

$$\xi = (\mathbf{P} \times \vec{\boldsymbol{\omega}}, \vec{\boldsymbol{\omega}}) = (\mathbf{v}, \vec{\boldsymbol{\omega}})$$

  - It can also be represented it as a matrix

$$\hat{\xi} = \begin{bmatrix} 0 & -\omega_z & \omega_y & v_1 \\ \omega_z & 0 & -\omega_x & v_2 \\ -\omega_y & \omega_x & 0 & v_3 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

- Homogenous transformation $\mathbf{T^h}_i$ is

$$\mathbf{T^h}_i = e^{\theta_0 \hat{\xi}_0} \cdot \prod_{j \in Parent(i), j \neq 0} e^{\theta_j \hat{\xi}_j}$$

---- From appendix B in my Ph.D. Thesis

Motion Capture with LBS

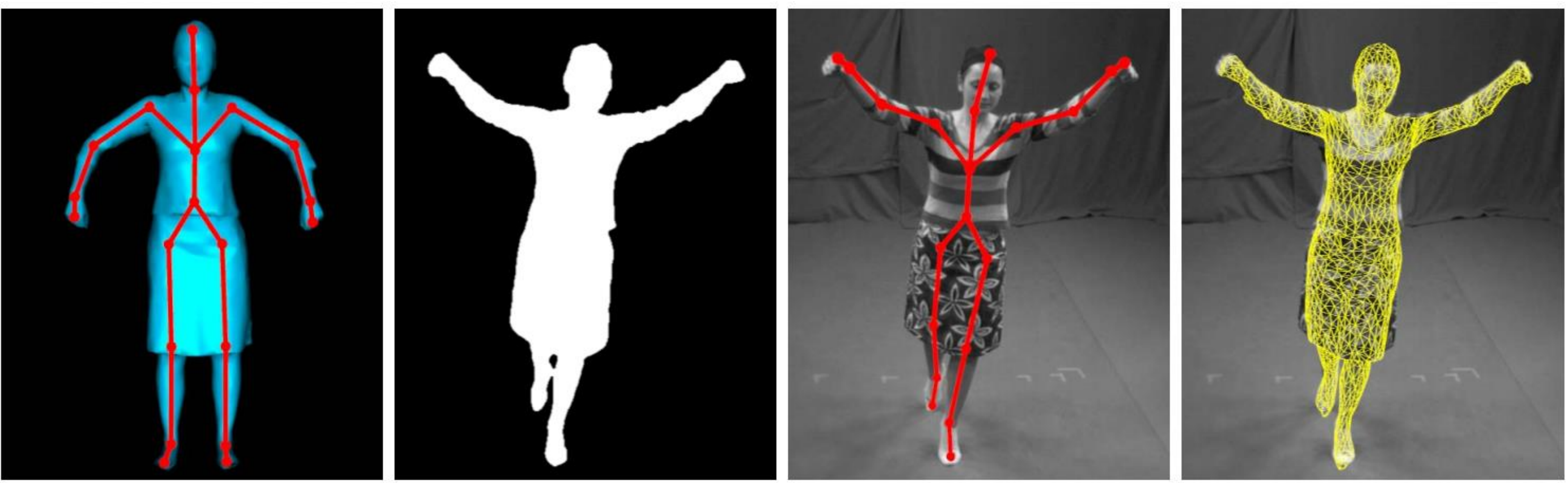
Search optimal skinning transformations $\mathbf{T}$ to drive the articulate mesh model with $lbs(\mathbf{T})$ to best match the observed image data O

$$\min_{T} E\left(lbs\left(T\right),O\right)$$

- How to perform the skinning transformations T to drive the mesh model M ?
- How to measure the inconsistency between the deformed model and observed image data?
- How to find an optimal skinning transformations?

Analysis-by-synthesis

- Deform the mesh model *M* (synthesize)
  - Articulation pose is driven by $\boldsymbol{\theta} = (\theta_0, \theta_1, \ldots, \theta_n)$
- Compare the differences between the **projected** synthesis and real captured image data (analysis)



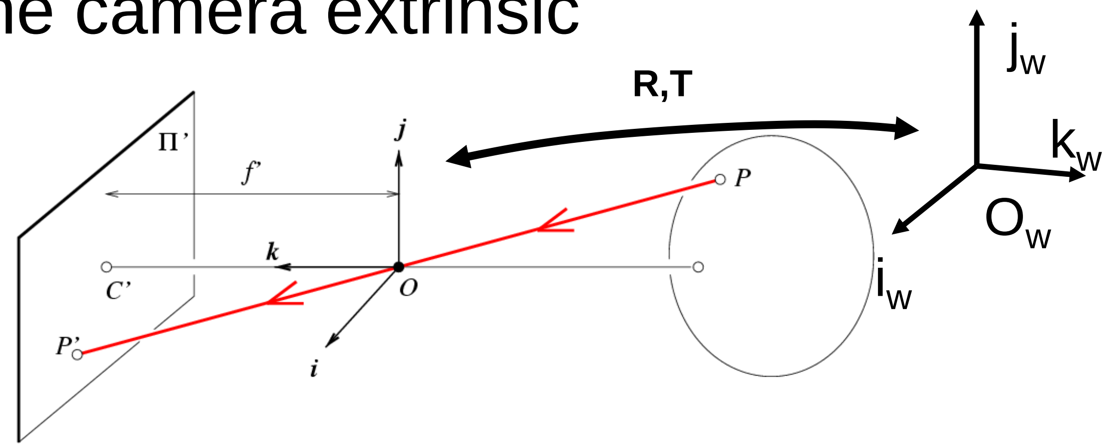
(a) Mesh model (b) Segmented images (c) Estimated skeleton (d) Deformed surface

Camera Projection Equation

- Suppose the camera parameters are K, R, t

  K is the camera intrinsic

  R, t is the camera extrinsic



- Camera projection equation for 3D point x and its corresponding 2D point X (homogeneous)

$$sX = K(Rx + t)$$

  s is the scale factor

3D-2D Consistency

- For reducing the scale factor s, we use X to cross the two sides of projection equation

$$\boxed{X \times sX} = X \times K(Rx + T)$$

$\boxed{X \times sX} \rightarrow 0$

$$[X]_\times K(Rx + t) = 0$$

or

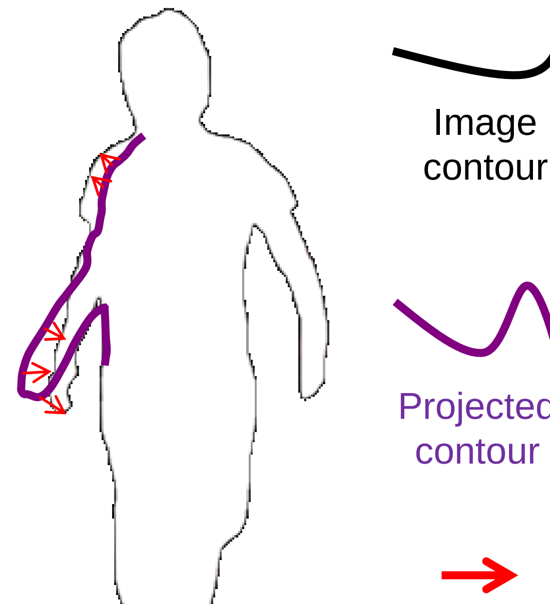
$$[X]_\times KRx = -[X]_\times Kt$$

x: 3D point

Here, $[X]_\times = \begin{bmatrix} 0 & -1 & Y \\ 1 & 0 & -X \\ -Y & X & 0 \end{bmatrix}$

3D-2D Consistency

- Take **silhouette** as example

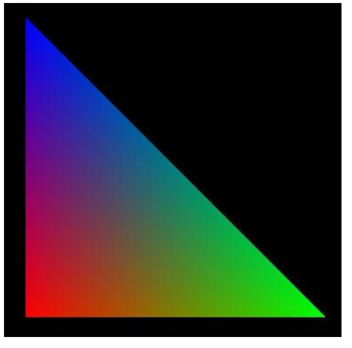


How do we know the 3D vertex position in the projected contour?

Rendering to Find 3D Position

- Two pass rendering
- First pass: encode the index of mesh triangles into the vertex color for each mesh face with (R, G, B) → 2^{24}-1=8388607
- Second pass: encode the (1,0,0), (0,1,0), (0,0,1) three colors for each mesh face to find the barycentric coordinate

1st pass: color with face index



2nd pass: color for barycentric

More about 3D-2D Consistency

- **Silhouette** and **contour matching** is successful for model driven motion capture
  - Motion Capture Using Joint Skeleton Tracking and Surface Estimation, CVPR2009
  - Markerless Motion Capture of Interacting Characters Using Multi-view Image Segmentation, CVPR2011
- Obtain the silhouette or mask is a very challenge topic → Deep learning!!! e.g., Mask-rcnn
- Model Rendering with texture
  - SIFT feature matching
  - Color consistency
  - Edge consistency

Motion Capture with LBS

Search optimal skinning transformations $\mathbf{T}$ to drive the articulate mesh model with $lbs(\mathbf{T})$ to best match the observed image data O

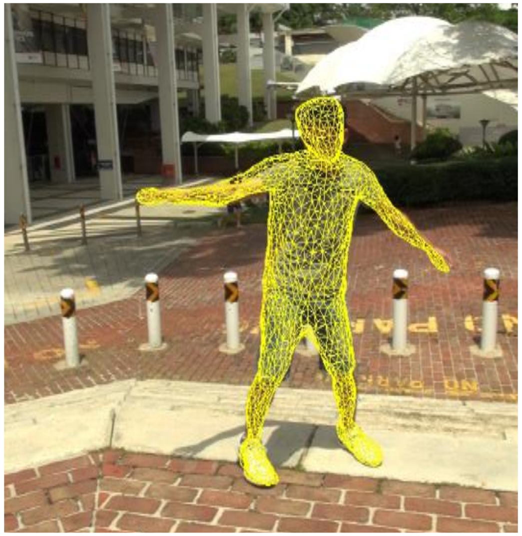
$$\min_{T} E\left(lbs\left(T\right), O\right)$$

- How to perform the skinning transformations T to drive the mesh model M ?
- How to measure the inconsistency between the deformed model and observed image data?
- How to find an optimal skinning transformations?

Global Optimization

- Sampling-based optimization

Interacting simulated annealing (ISA)



Deformed mesh

Observed image

https://www.yangangwang.com/papers/wang-omm-2017-04

Outdoor Markerless Motion Capture with Sparse Handheld Video Cameras, TVCG, 24(5), 2018

Local Optimization

- Consider the LBS deformation of articulated shape

$$v_i' = \left(\sum_{j=1}^{m} w_{i,j} \mathbf{T}_j \right) v_i$$

v_i: original vertex; v_i': deformed vertex

- It is also noted that

$$\mathbf{T^h}_i = e^{\theta_0 \hat{\xi}_0} \cdot \prod_{j \in Parent(i), j \neq 0} e^{\theta_j \hat{\xi}_j}$$

Local Optimization

- With the Taylor series at v_i and performing the linearization, we could have

$$v_i' = \left(I + \theta_0 \hat{\xi}_0 + \sum_{j=1}^{m} \overline{w}_j \theta_j \hat{\xi}_j \right) v_i$$

Where, $\overline{w}_j = \sum_{k \in children(j)} w_k$

- The upper equation can also be represented as

$$v_i' = v_i + J\chi$$

→ (B-8) in appendix B of my Ph.D. Thesis

Where, χ *has all the articulation parameters*

---- From appendix B in my Ph.D. Thesis

Local Optimization

- Our goal is to find some χ to perform the 3D-2D matching
- Since we have

$$[X]_\times KRx = -[X]_\times Kt$$

- And

$$v_i' = v_i + J\chi$$

- Solve **Ax = b** is easy and fast

Today

I What is Motion Capture?

II Linear Blend Skinning (LBS)

III Motion Capture with LBS

IV Advanced Techniques

V Conclusion

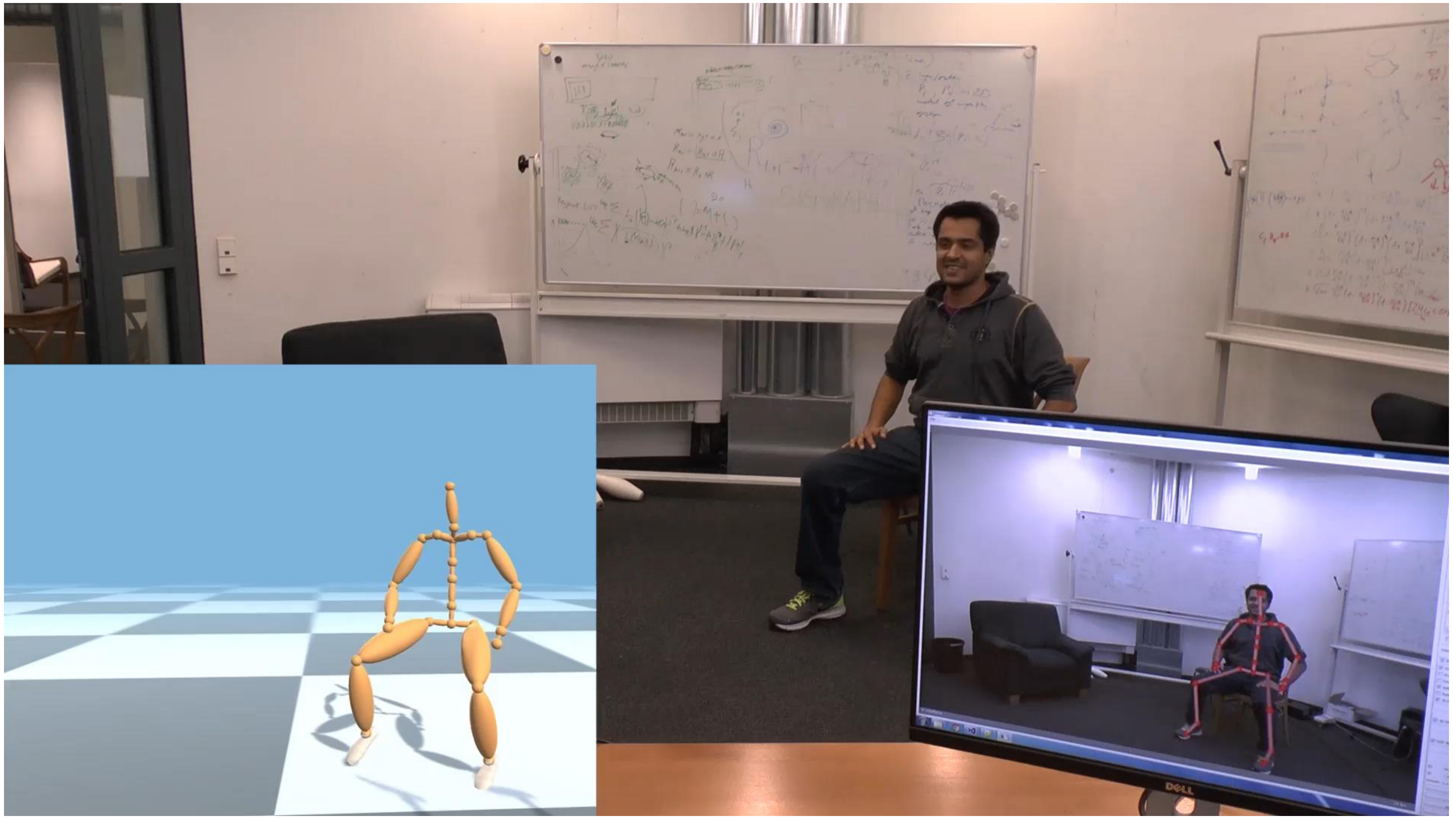
Advanced Techniques

- We have discussed about the articulated mesh deformation
- How about non-rigid?

- Robust Non-Rigid Motion Tracking and Surface Reconstruction Using L0 Regularization, *IEEE Transactions on Visualization and Computer Graphics (**TVCG**), 2018*

Advanced Techniques

- How about motion capture without LBS?



- VNect: Real-time 3D Human Pose Estimation with a Single RGB Camera, SIGGRAPH 2017

Advanced Techniques

- How about motion capture without LBS?

- Dense human pose estimation aims at mapping all human pixels of an RGB image to the 3D surface of the human, CVPR2018

Today

I What is Motion Capture?

II Linear Blend Skinning (LBS)

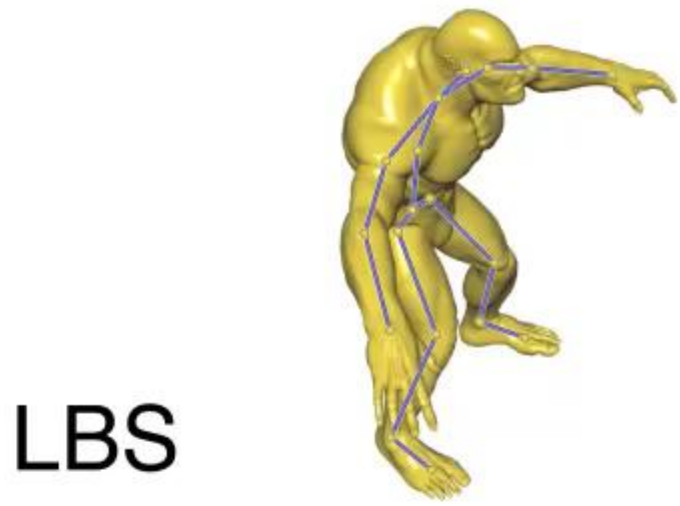
III Motion Capture with LBS

IV Advanced Techniques

V Conclusion

Conclusion

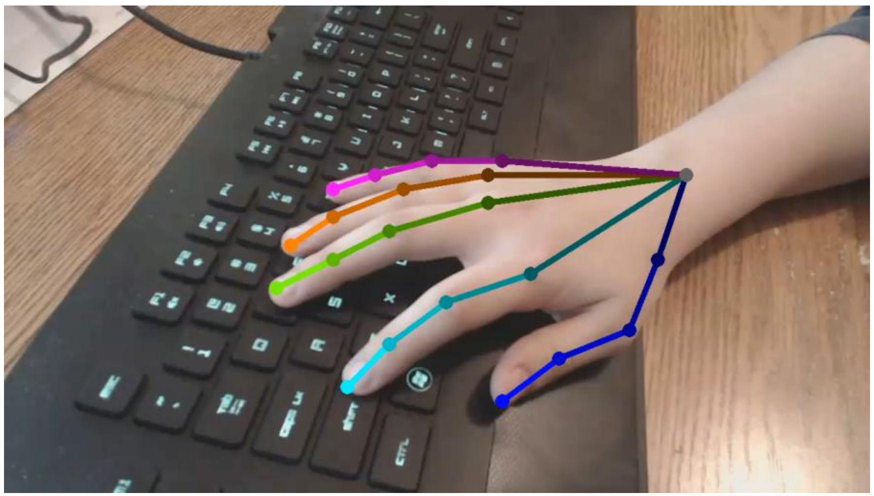
- LBS is a very basic and widely used technique for articulated motion capture
- LBS suffers from the candy-wrapper effect



http://igl.ethz.ch/projects/skinning/stretchable-twistable-bones/

Easter Egg

- Currently, I am focusing on real-time motion capture for hand gesture



- The dataset (***OneHand10K***) will be released in my personal website soon! ☺
  - More than **10K** with 21 2D joints and **Mask**

- Mask-pose Cascaded CNN for 2D Hand Pose Estimation from Single Color Image, *IEEE Transactions on Circuits and Systems for Video Technology (**TCSVT**), 2018. Accepted*

Any Questions?