What is Human Motion?

What makes Human Motion Hard to Analyze?

Intent

Muscle Contractions

Surface Motion

Observations

What we usually want

What causes motion

What we can directly measure

This lecture

It's impossible to kiss your elbow
Kinematics vs Dynamics

**Kinematics**: Geometry of Motion (Motion without Cause)

**Dynamics**: Physics of Motion (Motion with Cause)

This lecture
Capturing Human Motion
Holy Grail: Single Video Camera

Cameras are ubiquitous, cheap, and passive

3D Structure

3D Motion

Illustration from IR
This Lecture...

3D Dynamic Surface Reconstruction using Passive Sensing

• How should we represent human body surfaces?
• What can we extract from images?
• A Brief History of Virtualizing Reality
• Volumetric and Point-based 3D Reconstruction Algorithms
• Tour of the Virtualizing Studio 4.0
How do we Represent the Body Surface?

Representation Primitives

Voxel

Mesh

Surfel
Voxels
Volumetric Picture Element
Voxels
Volumetric Picture Elements

- Dynamic Voxels (doxels): Spacetime grid (e.g., 100 cm x 100 cm x 100 cm x 100 sec).
- Memory intensive (if used trivially)
- **Example**: 1 minute capture at 30 frames per second of 10 meter cubed space at centimeter resolution

\[
60 \times 30 \times (100 \times 10)^3 = 1,800,000,000,000
\]

- seconds
- frames per second
- centimeters per meter
- meters
- number of voxels
Mesh

- Continuity constraint embedding
- Limited memory consumption
- Fixed topology
Surfels
Surface Elements

Reconstructing 3D Body Shape and Motion

Representation


image measurables
There is also shading, texture, and other cues. See Shape from X (Marr)
Shading
Surface normals from shading information
Features
Detection/Tracking of Descriptors
Correspondences
Feature-based Matching
Correspondences
Feature-based Matching
Silhouettes
Background subtraction

Silhouettes
Holy Grail: Single Video Camera

Problem is unsolved. Very unsolved.
3D-2D Projection
How are images formed?

World Coordinate System \( \mathbb{R}^3 \)

Image Coordinate System \( \mathbb{R}^2 \)
3D-2D Projection

How are images formed?

World Coordinate System $\mathbb{R}^3$

$\bullet \mathbf{X} = \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$

$\mathbf{P} \rightarrow \mathbf{x}$

$x = \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$ Image Coordinate System $\mathbb{R}^2$
3D-3D Transformation
World Coordinate to Camera Coordinate

3D Rotation 3D Translation
R, t

World Coordinate System \( \mathbb{R}^3 \)

\[
\mathbf{X} = \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \quad \mathbf{X}' = \begin{bmatrix} X' \\ Y' \\ Z' \\ 1 \end{bmatrix}
\]

Camera Coordinate System \( \mathbb{R}^3 \)

Point in Camera Coordinates

Point in World Coordinates

\[
\begin{bmatrix} X' \\ Y' \\ Z' \\ 1 \end{bmatrix} = \begin{bmatrix} \mathbf{R}_{3 \times 3} & \mathbf{t}_{3 \times 1} \\ 0 & 1 \end{bmatrix}_{4 \times 4} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}
\]
Intrinsic Matrix
Camera Coordinate to Image Coordinate

Camera Coordinate System $\mathbb{R}^3$

Image Coordinate System $\mathbb{R}^2$

\[
\lambda x
\]
\[
\lambda y
\]
\[
\lambda
\]

\[
K_{3 \times 3} \begin{bmatrix} I_{3 \times 3} & 0_{3 \times 1} \end{bmatrix}_{3 \times 4} \begin{bmatrix} X' \\ Y' \\ Z' \\ 1 \end{bmatrix}
\]

\[
K = \begin{bmatrix} s_x f & 0 & p_x \\ 0 & s_y f & p_y \\ 0 & 0 & 1 \end{bmatrix}
\]

Pixel scaling factors
Focal length
Principal offset
3D-2D Projection

World to Camera to Image Coordinate

World Coordinate System \( \mathbb{R}^3 \)

Camera Coordinate System \( \mathbb{R}^3 \)

Image Coordinate System \( \mathbb{R}^2 \)

\[
\begin{bmatrix}
\lambda x \\
\lambda y \\
\lambda
\end{bmatrix} = K_{3 \times 3} \begin{bmatrix}
I_{3 \times 3} \\
0_{3 \times 1}
\end{bmatrix}_{3 \times 4} \begin{bmatrix}
R_{3 \times 3} \\
t_{3 \times 1}
\end{bmatrix}_{4 \times 4} \begin{bmatrix}
X \\
Y \\
Z \\
1
\end{bmatrix}
\]

\[
x \cong K [R | t] X
\]

\[
x \cong P_{3 \times 4} X
\]

Find \( P \) using camera calibration

http://www.vision.caltech.edu/bouguetj/calib_doc/
\[ \| \cdot \|_d \]

Normalized Distance in the presence of noise

\[ x \approx P_{3 \times 4}X \]

“equal up to scale” not “equal”

\[
\begin{bmatrix}
\lambda x \\
\lambda y \\
\lambda
\end{bmatrix} = P
\begin{bmatrix}
X \\
Y \\
Z \\
1
\end{bmatrix}
\]

\[ x = \lambda PX \]

Measure of Goodness

Maximum Likelihood Objective (under Gaussian Noise)
Single Image Projection

Invertible?

\[ \mathbf{x} \leftarrow \mathbf{P}_{3 \times 4} \mathbf{X} \]

World Coordinate System
\( \mathbb{R}^3 \)

Image Coordinate System
\( \mathbb{R}^2 \)

\[ \min_{\mathbf{x}} \| \mathbf{x}, \mathbf{P} \mathbf{X} \|_d \]

Ill-posed: 3 unknown, 2 constraints
Reconstruct me! :)

Reconstruct me! :)

Reconstruct me! :)
How do we resolve this?

Multiple Views!

\[ \mathbf{x} \rightarrow P_{3 \times 4} \mathbf{X} \]
Virtualized Reality™
Takeo Kanade
Virtualizing Studio
Kanade, Narayanan, Rander (1995)
Virtualizing Studio
Vedula, Saito, Kanade (1998)
Virtualizing Studio
Blue-C
Gross et al. (2003)

16 cameras
Stanford Multicamera Array

Levoy et al. (2005)

100 VGA Cameras
Lightstage 1-6
Paul Debevec (USC)

8 HD cameras
1200 light sources
Onsite 3D Video Capture

Nobuhara et al. (2009)

16 UXGA cameras
Discussion
Multiple View Reconstruction

Resolve ambiguity

\[ x \leftarrow P_{3 \times 4} X \]

\[ \min_X \sum_i \| x_i, P_i X \|_d \]
Stereoscopic 3D Reconstruction

Correspondence-based

\[
\min_{X} \|x_1, P_1 X\|_d + \|x_2, P_2 X\|_d
\]

Nonlinear least squares
Initialization
Direct Linear Transform Algorithm

\[ \mathbf{x} \cong \mathbf{P}_{3 \times 4} \mathbf{X} \]

Projection Equation --- Equal up to scale

\[ \| \mathbf{x} - \lambda \mathbf{P} \mathbf{X} \|_2 = \| \mathbf{x}, \mathbf{P} \mathbf{X} \|_d \]

Normalized Distance

\[ \mathbf{x} \times \lambda \mathbf{P} \mathbf{X} = 0 \]

Cross product:
\[ \mathbf{x} \times \mathbf{y} = \| \mathbf{x} \| \| \mathbf{y} \| \sin(\theta) \mathbf{n} \]

Cross product of a vector and a scaled version of itself is zero

\[ \mathbf{x} \times \mathbf{P} \mathbf{X} = 0 \]

Underconstrained Homogeneous System

\[ \mathbf{A}_{2 \times 4} \mathbf{X}_{4 \times 1} = 0 \]

Homogeneous System --- Solve using SVD

Function of \( \mathbf{x} \) and \( \mathbf{P} \)

From camera 1

\[ \begin{bmatrix} \mathbf{A}_1 \\ \vdots \\ \mathbf{A}_F \end{bmatrix} \mathbf{X} = 0 \]

From camera \( F \)
The three most important problems in computer vision are registration, registration, registration!

--- Takeo Kanade
# Stereoscopic 3D Reconstruction

<table>
<thead>
<tr>
<th>Pros</th>
<th>Cons</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>
# Stereoscopic 3D Reconstruction

<table>
<thead>
<tr>
<th>Pros</th>
<th>Cons</th>
</tr>
</thead>
<tbody>
<tr>
<td>Can provide temporal correspondence</td>
<td>Requires accurate spatial correspondence</td>
</tr>
<tr>
<td>High accuracy</td>
<td>Sparse reconstruction</td>
</tr>
<tr>
<td>Accuracy depends on the number of cameras</td>
<td>Does not provide normal information</td>
</tr>
<tr>
<td>Can identify concavities</td>
<td></td>
</tr>
</tbody>
</table>
Voxel Carving
Correspondence-free Reconstruction
Silhouettes
Background subtraction

Voxel Carving
Correspondence-free Reconstruction

\[ \mathbf{x} \cong \mathbf{P}_{3 \times 4} \mathbf{X} \]
Visual Hull
Voxel-Carving

Pros

Cons
Voxel-Carving

Pros

- Does not require spatial correspondences
- Trades off density with computation
- Easy to code
- Camera work with few cameras

Cons

- Does not provide temporal correspondence
- Redundant computation
- Requires accurate silhouettes
- Does not provide normal information
- Accuracy depends on the number of cameras
- Convex Hull
Animating 3D Scans

SCAPE: Shape Completion

Anguelov (2005)
The Kitchen Sink

de Aguiar (2008)
Animating 3D Scans

Pros and Cons

Pros
- High resolution
- Can fill missing data
- Temporal continuity

Cons
- Drift
- Topology changes
- Low detail (if generic models are used)
- Baked detail (if specific models are used)
Representation
Reconstructing 3D Body Shape and Motion

Image \rightarrow \text{Voxel Carving} \rightarrow \text{Voxel Carving} \rightarrow \text{Surfel} \rightarrow \text{Model Constraints}

\text{image measurables} \rightarrow \text{Photometric Stereo} \rightarrow \text{Voxel} \rightarrow \text{Mesh}
Conclusion

3D Structure reconstruction is maturing.
3D Motion estimation is primitive.
This Lecture...

3D Dynamic Surface Reconstruction using Passive Sensing

- How should we represent human body surfaces?
- What can we extract from images?
- A Brief History of Virtualizing Reality
- Volumetric and Point-based 3D Reconstruction Algorithms
- Tour of the Virtualizing Studio 4.0
Further Reading
Subtitle

• Matsuyama et al., 3D Video and Its Applications, 2012.
• Vlasic et al., Dynamic Shape Capture using Multi-View Photometric Stereo, SIGGRAPH Asia, 2009
Demo!