Computer Vision

CSE 455
Stereo and 3D

Linda Shapiro
Professor of Computer Science & Engineering
Professor of Electrical Engineering
Camera Calibration

The idea is to snap images at different depths and get a lot of 2D-3D point correspondences.

\[ x_1, y_1, z_1, u_1, v_1 \]
\[ x_2, y_2, z_1, u_2, v_2 \]
\[ \vdots \]
\[ x_n, y_n, z_n, u_n, v_n \]

Then solve a system of equations to get camera parameters.
Camera Parameters

A camera is described by several parameters:

- Translation $T$ of the optical center from the origin of world coords
- Rotation $R$ of the image plane
- Focal length $f$, principal point $(x'_c, y'_c)$, pixel size $(s_x, s_y)$
- Blue parameters are called “extrinsics,” red are “intrinsics”
A camera is described by several parameters:

- Translation $T$ of the optical center from the origin of the world coords
- Rotation $R$ of the image plane
- Focal length $f$, principal point $(x'_c, y'_c)$, pixel size $(s_x, s_y)$
- Blue parameters are called "extrinsics," red are "intrinsics"

Projection equation:

$$x = \begin{bmatrix} wx \\ wy \\ w \end{bmatrix} = \begin{bmatrix} * & * & * & * \\ * & * & * & * \\ * & * & * & * \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} = \Pi X$$

- The projection matrix models the cumulative effect of all parameters
- Useful to decompose into a series of operations
Camera Parameters

A camera is described by several parameters:

- Translation $T$ of the optical center from the origin of world coords
- Rotation $R$ of the image plane
- Focal length $f$, principal point $(x'_c, y'_c)$, pixel size $(s_x, s_y)$
- Blue parameters are called “extrinsics,” red are “intrinsics”

**Projection equation**

$$x = \begin{bmatrix} wx \\ wy \\ w \end{bmatrix} = \begin{bmatrix} * & * & * \\ * & * & * \\ * & * & * \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \Pi X \quad y' \quad (x'_c, y'_c)$$

- Useful to decompose into a series of operations

$$\Pi = \begin{bmatrix} -fs_x & 0 & x'_c \\ 0 & -fs_y & y'_c \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} R_{3x3} & 0_{3x1} \\ 0_{1x3} & 1 \end{bmatrix} \begin{bmatrix} I_{3x3} \\ 0_{1x3} \end{bmatrix} T_{3x1} \quad \left[ tx, ty, tz \right]^T$$

- The definitions of these parameters are **not** completely standardized
  - especially intrinsics—varies from one book to another
Stereo
Amount of horizontal movement is ...

...inversely proportional to the distance from the camera
Depth from Stereo

• Goal: recover depth by finding image coordinate $x'$ that corresponds to $x$
Disparity is inversely proportional to depth.

\[
\frac{x - x'}{O - O'} = \frac{f}{z}
\]

\[
\text{disparity} = x - x' = \frac{B \cdot f}{z}
\]

See Chapter 12 of Shapiro and Stockman Text.
Depth from Stereo

• Goal: recover depth by finding image coordinate $x'$ that corresponds to $x$

• Sub-Problems
  1. Calibration: How do we recover the relation of the cameras (if not already known)?
  2. Correspondence: How do we search for the matching point $x'$?
Correspondence Problem

- We have two images taken from cameras with different intrinsic and extrinsic parameters.

- How do we match a point in the first image to a point in the second? How can we constrain our search?
Key idea: Epipolar constraint

Potential matches for $x$ have to lie on the corresponding line $l’$. 

Potential matches for $x’$ have to lie on the corresponding line $l$. 

• **Baseline** – line connecting the two camera centers

• **Epipoles**
  = intersections of baseline with image planes
  = projections of the other camera center

• **Epipolar Plane** – plane containing baseline (1D family)
Epipolar geometry: notation

- **Baseline** – line connecting the two camera centers
- **Epipoles**
  = intersections of baseline with image planes
  = projections of the other camera center
- **Epipolar Plane** – plane containing baseline (1D family)
- **Epipolar Lines** - intersections of epipolar plane with image planes (always come in corresponding pairs)
Example: Converging cameras
Example: Motion parallel to image plane
Epipolar constraint: Calibrated case

- Assume that the intrinsic and extrinsic parameters of the cameras are known.
- We can multiply the projection matrix of each camera (and the image points) by the inverse of the calibration matrix to get *normalized* image coordinates.
- We can also set the global coordinate system to the coordinate system of the first camera. Then the projection matrices of the two cameras can be written as \([I \mid 0]\) and \([R \mid t]\).
Simplified Matrices for the 2 Cameras

\[
\begin{pmatrix}
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0
\end{pmatrix}
= \begin{pmatrix}
I \\
0
\end{pmatrix}
\]

\[
\begin{pmatrix}
R \\
0
\end{pmatrix}
= \begin{pmatrix}
R \\
T
\end{pmatrix}
\]
Epipolar constraint: Calibrated case

The vectors $Rx$, $t$, and $x'$ are coplanar
Epipolar constraint: Calibrated case

\[ x' \cdot [t \times (Rx)] = 0 \quad \Rightarrow \quad x'^T E x = 0 \quad \text{with} \quad E = [t \times]R \]

The vectors \( Rx, t, \) and \( x' \) are coplanar.

Essential Matrix \( E \)
(Longuet-Higgins, 1981)
Epipolar constraint: Calibrated case

- $E x$ is the epipolar line associated with $x$ ($l' = E x$)
- $E^T x'$ is the epipolar line associated with $x'$ ($l = E^T x'$)
- $E e = 0$ and $E^T e' = 0$
- $E$ is singular (rank two)
- $E$ has five degrees of freedom

$x' \cdot [t \times (Rx)] = 0 \implies x'^T E x = 0$ with $E = [t \times] R$
Moving on to stereo...

Fuse a calibrated binocular stereo pair to produce a depth image

Dense depth map

Many of these slides adapted from Steve Seitz and Lana Lazebnik
Stereo image rectification

• Reproject image planes onto a common plane parallel to the line between camera centers

• Pixel motion is horizontal after this transformation

• Two homographies (3x3 transform), one for each input image reprojection

Example

Unrectified

Rectified
• Slide a window along the right scanline and compare contents of that window with the reference window in the left image

• Matching cost: SSD, SAD, or normalized correlation
Correspondence search

Left

Right

scanline

Norm. corr
Results with window search

Data

Window-based matching

Ground truth
Using more than two images

Multi-View Stereo for Community Photo Collections
M. Goesele, N. Snavely, B. Curless, H. Hoppe, S. Seitz
Proceedings of ICCV 2007,
3D model

- “Digital copy” of real object
- Allows us to
  - Inspect details of object
  - Measure properties
  - Reproduce in different material
- Many applications
  - Cultural heritage preservation
  - Computer games and movies
  - City modelling
  - E-commerce
Applications: cultural heritage

SCULPTEUR European project
Applications: art

Block Works Precipitate III 2004
*Mild steel blocks* 80 x 46 x 66 cm

Domain Series Domain VIII Crouching
1999 *Mild steel bar* 81 x 59 x 63 cm
Applications: structure engineering

BODY / SPACE / FRAME, Antony Gormley, Lelystad, Holland
Applications: 3D indexation
Applications: archaeology

- “forma urbis romae” project

**Fragments of the City: Stanford’s Digital Forma Urbis Romae Project**
David Koller, Jennifer Trimble, Tina Najbjerg, Natasha Gelfand, Marc Levoy


1186 fragments
Applications: large scale modelling

[Pollefeys08]

[Furukawa10]

[Goesele07]

[Cornelis08]
Applications: Medicine

<table>
<thead>
<tr>
<th>expert's order</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>images</td>
<td>learning</td>
<td>a-Imk</td>
<td>mirror</td>
<td>m-Imk</td>
<td>plane</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>learning</td>
<td>1</td>
<td>2</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>8</td>
<td>9</td>
<td>7</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>a-Imk</td>
<td>1</td>
<td>2</td>
<td>5</td>
<td>6</td>
<td>4</td>
<td>8</td>
<td>7</td>
<td>9</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>mirror</td>
<td>1</td>
<td>2</td>
<td>8</td>
<td>5</td>
<td>6</td>
<td>9</td>
<td>3</td>
<td>7</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>m-Imk</td>
<td>1</td>
<td>2</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>9</td>
<td>7</td>
<td>10</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>plane</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>5</td>
<td>4</td>
<td>6</td>
<td>7</td>
<td>9</td>
<td>10</td>
<td>8</td>
</tr>
</tbody>
</table>
Scanning technologies

• Laser scanner, coordinate measuring machine
  – Very accurate
  – Very Expensive
  – Complicated to use

Minolta

Contura CMM
Medical Scanning System
The “Us” Data Set (subset)
3d shape from photographs

“Estimate a 3d shape that would generate the input photographs given the same material, viewpoints and illumination”

geometry

viewpoint

material

illumination

image
Photometric Stereo

• Estimate the surface normals of a given scene given multiple 2D images taken from the same viewpoint, but under different lighting conditions.

• Basic photometric stereo required a Lambertian reflectance model:

\[ I = \rho \mathbf{n} \cdot \mathbf{v} \]

where \( I \) is pixel intensity, \( \mathbf{n} \) is the normal, \( \mathbf{v} \) is the lighting direction, and \( \rho \) is diffuse albedo constant, which is a reflection coefficient.
Basic Photometric Stereo
Basic Photometric Stereo

Scene

Diffuse sphere

Projector 1

High speed camera

Projector 2

Projector 3
Basic Photometric Stereo

• **K** light sources
• Lead to **K** images \( R_1(p,q), \ldots, R_K(p,q) \) each from just one of the light sources being on
• For any \((p,q)\), we get **K** intensities \( I_1, \ldots I_K \)
• Leads to a set of **linear equations** of the form
  \[
  I_k = \rho n \cdot v_k
  \]
• Solving leads to a **surface normal map**.
Photometric Stereo

Inputs

3D normals
3d shape from photographs

Photograph based 3d reconstruction is:

- practical
- fast
- non-intrusive
- low cost
- Easily deployable outdoors
- “low” accuracy
- Results depend on material properties
Reconstruction

• Generic problem formulation: given several images of the same object or scene, compute a representation of its 3D shape
Reconstruction

• **Generic problem formulation:** given several images of the same object or scene, compute a representation of its 3D shape

• “Images of the same object or scene”
  – Arbitrary number of images (from two to thousands)
  – Arbitrary camera positions (camera network or video sequence)
  – Calibration may be initially unknown

• “Representation of 3D shape”
  – Depth maps
  – Meshes
  – Point clouds
  – Patch clouds
  – Volumetric models
  – Layered models
Multiple-baseline stereo

Reconstruction from silhouettes

- Can be computed robustly
- Can be computed efficiently

\[ \text{background + foreground} - \text{background} = \text{foreground} \]
Reconstruction from Silhouettes

- The case of binary images: a voxel is photo-consistent if it lies inside the object’s silhouette in all views.
Reconstruction from Silhouettes

• The case of binary images: a voxel is **photo-consistent** if it lies inside the object’s silhouette in all views

Finding the silhouette-consistent shape (visual hull):

• *Backproject* each silhouette
• Intersect backprojected volumes
Calibrated Image Acquisition

Calibrated Turntable

Selected Dinosaur Images

Selected Flower Images
Space Carving in General

• Space Carving Algorithm
  • Initialize to a volume $V$ containing the true scene
  • Choose a voxel on the outside of the volume
  • Project to visible input images
  • Carve if not photo-consistent (inside object’s silhouette)

Our 4-camera light-striping stereo system

(now deceased)
Calibration Object

The idea is to snap images at different depths and get a lot of 2D-3D point correspondences.
Surface Modeling and Display from Range and Color Data

Kari Pulli  UW
Michael Cohen  MSR
Tom Duchamp  UW
Hugues Hoppe  MSR
John McDonald  UW
Linda Shapiro  UW
Werner Stuetzle  UW

UW = University of Washington
    Seattle, WA USA
MSR = Microsoft Research
    Redmond, WA USA
Introduction

Goal

- develop robust algorithms for constructing 3D models from range & color data
- use those models to produce realistic renderings of the scanned objects
Surface Reconstruction

Step 1: Data acquisition
Obtain range data that covers the object. Filter, remove background.

Step 2: Registration
Register the range maps into a common coordinate system.

Step 3: Integration
Integrate the registered range data into a single surface representation.

Step 4: Optimization
Fit the surface more accurately to the data, simplify the representation.
Carve space in cubes

Label cubes

- Project cube to image plane (hexagon)
- Test against data in the hexagon
3D space is made up of many cubes.

OUTSIDE one of many cubes in virtual 3D cube space

(u,v,d)

image plane

depth map

(x,y,z)

one of many cubes in virtual 3D cube space
Several views

Processing order:
FOR EACH cube
FOR EACH view

Rules:
any view thinks cube's out => it's out
every view thinks cube's in => it's in
else
=> it's at boundary
Hierarchical space carving

- Big cubes => fast, poor results
- Small cubes => slow, more accurate results
- Combination = octrees

RULES:  
- cube's out => done
- cube's in => done
- else => recurse
Hierarchical space carving

- Big cubes => fast, poor results
- Small cubes => slow, more accurate results
- Combination = octrees

RULES:  
- cube's out => done  
- cube's in => done  
- else => recurse
The rest of the chair
Same for a husky pup
Optimizing the dog mesh

Registered points

Initial mesh

Optimized mesh
View dependent texturing
More: Space Carving Results: African Violet

Input Image (1 of 45)

Reconstruction

Reconstruction

Reconstruction

Source: S. Seitz
More: Space Carving Results: Hand

Input Image (1 of 100)

Views of Reconstruction
Stereo from community photo collections

- Up to now, we’ve always assumed that camera calibration is known
- For photos taken from the Internet, we need *structure from motion* techniques to reconstruct both camera positions and 3D points.
Head Reconstruction from Uncalibrated Internet Photos

• Input: Internet photos in different poses and expressions

• Output: 3D model of the head
Recognizing Deformable Shapes

Salvador Ruiz Correa
(CSE/EE576 Computer Vision I)
Goal

- We are interested in developing algorithms for recognizing and classifying deformable object shapes from range data.

- This is a difficult problem that is relevant in several application fields.
What Kind Of Deformations?

- Toy animals
- 3D Faces
- Normal
- Abnormal
- Shape classes: significant amount of intra-class variability
Component-Based Methodology

1. Numeric Signatures
   define

2. Components

3. Symbolic Signatures

Overcomes the limitations of the alignment-verification approach

Architecture of Classifiers

4. Recognition And Classification Of Deformable Shapes

Describe spatial configuration
Numeric Signatures

1. Numeric Signatures
2. Components
3. Symbolic Signatures
4. Encode Local Surface Geometry of an Object
   +
   →
   →
   Architecture of Classifiers
The Spin Image Signature

P is the selected vertex.

X is a contributing point of the mesh.

$\alpha$ is the perpendicular distance from X to P’s surface normal.

$\beta$ is the signed perpendicular distance from X to P’s tangent plane.
Spin Image Construction

- A spin image is constructed
  - about a specified oriented point $o$ of the object surface
  - with respect to a set of contributing points $C$, which is controlled by maximum distance and angle from $o$.

- It is stored as an array of accumulators $S(\alpha, \beta)$ computed via:

- For each point $c$ in $C(o)$
  1. compute $\alpha$ and $\beta$ for $c$.
  2. increment $S(\alpha, \beta)$
Numeric Signatures: Spin Images

- Rich set of surface shape descriptors.
- Their spatial scale can be modified to include local and non-local surface features.
- Representation is robust to scene clutter and occlusions.
Components

1. Numeric Signatures

2. Components
   - Equivalent Numeric Signatures: Encode Local Geometry of a Shape Class

3. Symbolic Signatures

4. Architecture of Classifiers
How To Extract Shape Class Components?

Select Seed Points → Compute Numeric Signatures → Region Growing Algorithm → Component Detector

Training Set

Grown components around seeds
Component Extraction Example

Selected 8 seed points by hand

Grow one region at the time (get one detector per component)

Region Growing

Labeled Surface Mesh

Detected components on a training sample
How To Combine Component Information?

Note: Numeric signatures are invariant to mirror symmetry; our approach preserves such an invariance.
Symbolic Signatures

1. Numeric Signatures
2. Components
3. Symbolic Signatures

Encode Geometrical Relationships Among Components

Architecture of Classifiers
Symbolic Signature

Labeled Surface Mesh

Encode Geometric Configuration

Symbolic Signature at P

Critical Point P

Matrix storing component labels
Symbolic Signatures Are Robust To Deformations

Relative position of components is stable across deformations: experimental evidence
Architecture of Classifiers

1. Numeric Signatures
2. Components
3. Symbolic Signatures

Learns Components And Their Geometric Relationships

Architecture of Classifiers
Proposed Architecture

- Input
- Labeled Mesh

Two classification stages

Verify spatial configuration of the components

Identify Components

Labeled Mesh

Identify Symbolic Signatures

Class Label

-1 (Abnormal)
Experimental Validation

Recognition Tasks: 4 (T1 - T4)
Classification Tasks: 3 (T5 - T7)
No. Experiments: 5470
Shape Classes
Enlarging Training Sets Using Virtual Samples

Morphs

Originals

Global Morphing Operators

University of Washington

Variables

Physical Modeling

Operators

Enlarging Training Sets Using Virtual Samples

Morphs

Originals

Global Morphing Operators

University of Washington

Variables

Physical Modeling

Operators

Enlarging Training Sets Using Virtual Samples

Morphs

Originals

Global Morphing Operators

University of Washington

Variables

Physical Modeling

Operators

Enlarging Training Sets Using Virtual Samples

Morphs

Originals

Global Morphing Operators

University of Washington

Variables

Physical Modeling

Operators

Enlarging Training Sets Using Virtual Samples

Morphs

Originals

Global Morphing Operators

University of Washington

Variables

Physical Modeling

Operators

Enlarging Training Sets Using Virtual Samples

Morphs

Originals

Global Morphing Operators

University of Washington

Variables

Physical Modeling

Operators
Task 1: Recognizing Single Objects (1)

- No. Shape classes: 9.
- Training set size: 400 meshes.
- Testing set size: 200 meshes.
- No. Component detectors: 3.
- No. Symbolic signature detectors: 1.
- Numeric signature size: 40x40.
- Symbolic signature size: 20x20.
- No clutter and occlusion.
Task 1: Recognizing Single Objects (2)

- Snowman: 93%.
- Rabbit: 92%.
- Dog: 89%.
- Cat: 85.5%.
- Cow: 92%.
- Bear: 94%.
- Horse: 92.7%.
- Human head: 97.7%.
- Human face: 76%.

Recognition rates (true positives)
(No clutter, no occlusion, complete models)
Main Contributions (2)

• A region growing algorithm for learning shape class components.

• A novel architecture of classifiers for abstracting the geometry of a shape class.

• A validation of our methodology in a set of large scale recognition and classification experiments aimed at applications in scene analysis and medical diagnosis.