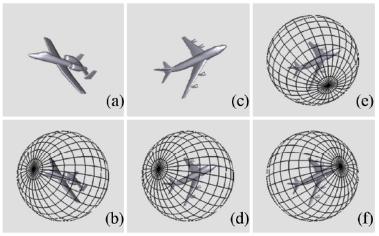
On Visual Similarity Based 3D Model Retrieval

Ding-Yun Chen Xiao-Pei Tian Yu-Te Shen Ming Ouhyoung

Department of Computer Science and Information Engineering
National Taiwan University, Taipei

600.658 - Seminar on Shape Analysis and Retrieval

Visual Similarity Based 3D Model Matching



- Image differences in light fields
- Multiple viewing angles
- Exhaustive search to find best alignment

Outline

Outline

Light Field

Definition

A light field (or plenoptic function) is a 5D function representing the radiance at a given 3D point along a given direction.

- Reduces to 4D in free space
- Collection of 2D images rendered from a 2D array of cameras
- 20 cameras positioned on the vertices of a regular dodecahedron

Light Field

Definition

A light field (or plenoptic function) is a 5D function representing the radiance at a given 3D point along a given direction.

- Reduces to 4D in free space
- Collection of 2D images rendered from a 2D array of cameras
- 20 cameras positioned on the vertices of a regular dodecahedron ▸ Regular Dodecahedron
  - Images rendered as silhouettes
  - Only 10 of the 20 silhouettes need to be used

Light Field

Definition

A light field (or plenoptic function) is a 5D function representing the radiance at a given 3D point along a given direction.

- Reduces to 4D in free space
- Collection of 2D images rendered from a 2D array of cameras
- 20 cameras positioned on the vertices of a regular dodecahedron ▸ Regular Dodecahedron
  - Images rendered as silhouettes
  - Only 10 of the 20 silhouettes need to be used

Light Field

Definition

A light field (or plenoptic function) is a 5D function representing the radiance at a given 3D point along a given direction.

- Reduces to 4D in free space
- Collection of 2D images rendered from a 2D array of cameras
- 20 cameras positioned on the vertices of a regular dodecahedron ▸ Regular Dodecahedron
  - Images rendered as silhouettes
  - Only 10 of the 20 silhouettes need to be used

Light Field

Definition

A light field (or plenoptic function) is a 5D function representing the radiance at a given 3D point along a given direction.

- Reduces to 4D in free space
- Collection of 2D images rendered from a 2D array of cameras
- 20 cameras positioned on the vertices of a regular dodecahedron ▸ Regular Dodecahedron
  - Images rendered as silhouettes
  - Only 10 of the 20 silhouettes need to be used

Light Field

Definition

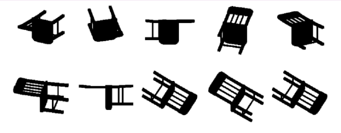
A light field (or plenoptic function) is a 5D function representing the radiance at a given 3D point along a given direction.

- Reduces to 4D in free space
- Collection of 2D images rendered from a 2D array of cameras
- 20 cameras positioned on the vertices of a regular dodecahedron ▸ Regular Dodecahedron
  - Images rendered as silhouettes
  - Only 10 of the 20 silhouettes need to be used

Outline

1. Feature Extraction for Representing 3D Models
   - LightField Descriptor
   - Image Metric
   - Extracting LightField Descriptors for a 3D Model
2. Retrieval of 3D Models Using LightField Descriptors
3. Experimental Results

A LightField Descriptor for 3D Models

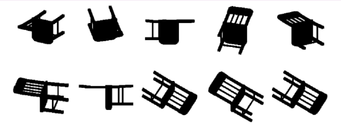


Definition

LightFiled Descriptor is defined as features of 10 images rendered from vertices of dodecahedron over a hemisphere.

- 60 camera positions for each orientation of the dodecahedron
  - Regular Dodecahedron
- $D_A(L_1, L_2) = \min_{i=1}^{60} \sum_{k=1}^{10} d(I_{1k}^i, I_{2k}^i)$

A LightField Descriptor for 3D Models

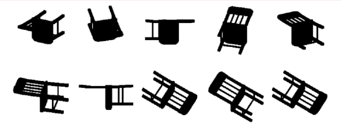


Definition

LightFiled Descriptor is defined as features of 10 images rendered from vertices of dodecahedron over a hemisphere.

- 60 camera positions for each orientation of the dodecahedron
  ▸ Regular Dodecahedron
- $D_A(L_1, L_2) = \min_{i=1}^{60} \sum_{k=1}^{10} d(I_{1k}^{i}, I_{2k}^{i})$

A LightField Descriptor for 3D Models



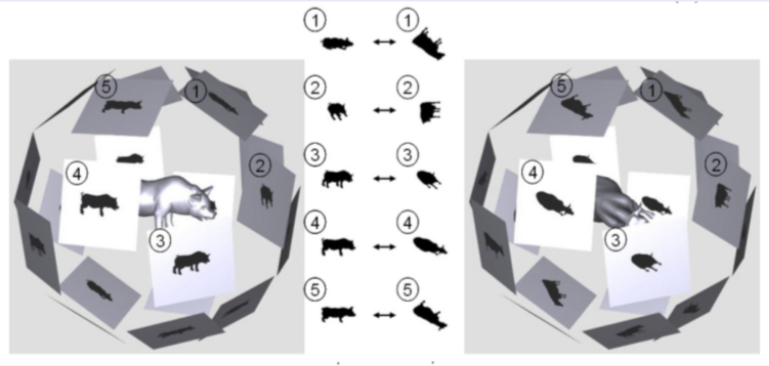
Definition

LightFiled Descriptor is defined as features of 10 images rendered from vertices of dodecahedron over a hemisphere.

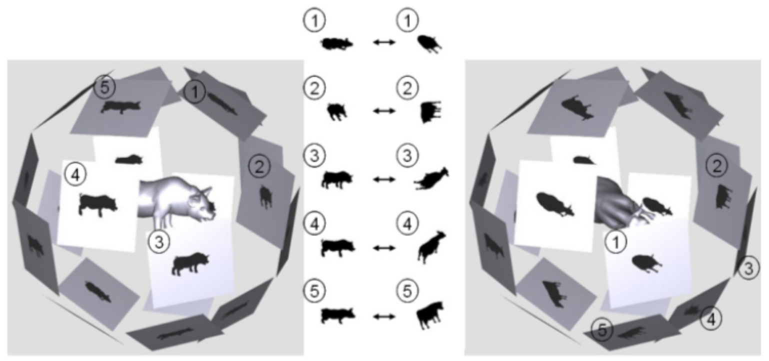
- 60 camera positions for each orientation of the dodecahedron
  ▸ Regular Dodecahedron
- $D_A(L_1, L_2) = \min_{i=1}^{60} \sum_{k=1}^{10} d(I_{1k}^i, I_{2k}^i)$

On orientations of pigs and cows

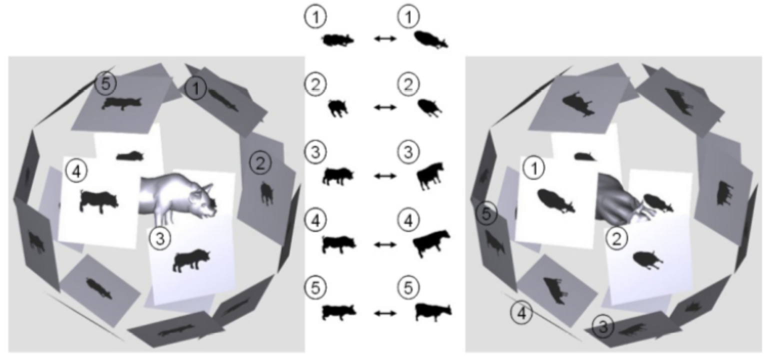
On orientations of pigs and cows



On orientations of pigs and cows



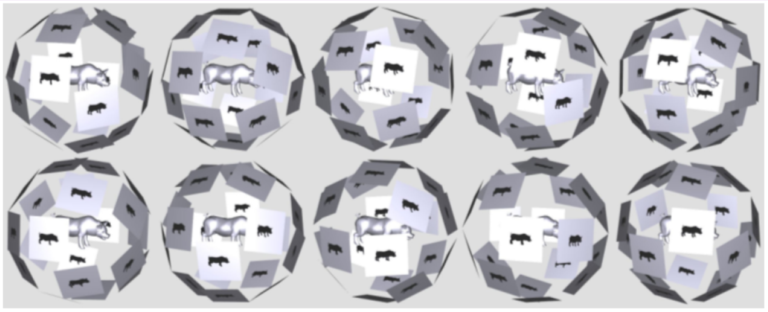
On orientations of pigs and cows



A Set of LightField Descriptors to Counter Rotational Variations

- Use N different camera system orientations
  - Ensure uniform camera distribution
  - Obtains a total of $(N \times (N-1)+1) \times 60$ different rotations between 2 models
  - $D_B = \min_{j,k=1}^{N} D_A(L_{1j}, L_{2k})$
  - Current work uses $N = 10$ (5460 different rotations)

A Set of LightField Descriptors to Counter Rotational Variations



Outline

1. Feature Extraction for Representing 3D Models
   - LightField Descriptor
   - Image Metric
   - Extracting LightField Descriptors for a 3D Model

2. Retrieval of 3D Models Using LightField Descriptors

3. Experimental Results

Combine Different Shape Descriptors

- Combine a "region-based" and a "contour-based" shape descriptor
- Region-based descriptors:
  - Combine information from all pixels within the region
  - Do not emphasize boundary features
  - Zernike Moment Descriptors (ZMD) ▸ Details
- Contour-based descriptors:
  - Exploits only boundary information, ignoring the interior
  - Fourier Descriptors (FD) ▸ Details

- 35 ZMD and 10 FD coefficients (8 bits)
- $d(I_1, I_2) = \sum_{j=1}^{45} |C_{1j} - C_{2j}|$

Combine Different Shape Descriptors

- Combine a "region-based" and a "contour-based" shape descriptor
- Region-based descriptors:
  - Combine information from all pixels within the region
  - Do not emphasize boundary features
  - Zernike Moment Descriptors (ZMD) Details
- Contour-based descriptors:
  - Exploits only boundary information, ignoring the interior
  - Fourier Descriptors (FD) Details

- 35 ZMD and 10 FD coefficients (8 bits)
- $d(I_1, I_2) = \sum_{j=1}^{45} |C_{1j} - C_{2j}|$

Combine Different Shape Descriptors

- Combine a “region-based” and a “contour-based” shape descriptor
- Region-based descriptors:
  - Combine information from all pixels within the region
  - Do not emphasize boundary features
  - Zernike Moment Descriptors (ZMD) ▸ Details
- Contour-based descriptors:
  - Exploits only boundary information, ignoring the interior
  - Fourier Descriptors (FD) ▸ Details

- 35 ZMD and 10 FD coefficients (8 bits)
- $d(I_1, I_2) = \sum_{j=1}^{45} |C_{1j} - C_{2j}|$

Combine Different Shape Descriptors

- Combine a "region-based" and a "contour-based" shape descriptor
- Region-based descriptors:
  - Combine information from all pixels within the region
  - Do not emphasize boundary features
  - Zernike Moment Descriptors (ZMD) ▸ Details
- Contour-based descriptors:
  - Exploits only boundary information, ignoring the interior
  - Fourier Descriptors (FD) ▸ Details

- 35 ZMD and 10 FD coefficients (8 bits)
- $d(I_1, I_2) = \sum_{j=1}^{45} |C_{1j} - C_{2j}|$

Outline

Extracting LightField Descriptors for a 3D Model

Algorithm

1. Normalize for translation and scaling
2. Obtain *LightField Descriptors* for 10 different camera system orientation
3. For each *LightField Descriptor* store 10 views of the image
4. For each of the 100 images store the corresponding image metric

Extracting LightField Descriptors for a 3D Model

Algorithm

1. Normalize for translation and scaling
2. Obtain *LightField Descriptors* for 10 different camera system orientation
3. For each *LightField Descriptor* store 10 views of the image
4. For each of the 100 images store the corresponding image metric

Extracting LightField Descriptors for a 3D Model

Algorithm

1. Normalize for translation and scaling
2. Obtain *LightField Descriptors* for 10 different camera system orientation
3. For each *LightField Descriptor* store 10 views of the image
4. For each of the 100 images store the corresponding image metric

Extracting LightField Descriptors for a 3D Model

Algorithm

1. Normalize for translation and scaling
2. Obtain *LightField Descriptors* for 10 different camera system orientation
3. For each *LightField Descriptor* store 10 views of the image
4. For each of the 100 images store the corresponding image metric

Outline

Retrieval of 3D Models from Large Database

Baseic Idea

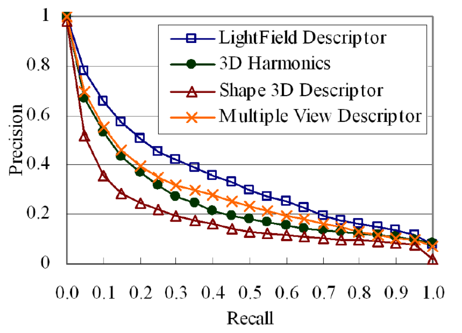
- Use a few *LightField Descriptors* and only a few highly quantized coefficients while comparing all images in the database
- Set the threshold to be the mean of similarity
- Progressively refine the comparison by:
  - Increasing the number of *LightField Descriptors* to compare
  - Increasing the number of coefficients used to calculate similarity

Outline

Test Set

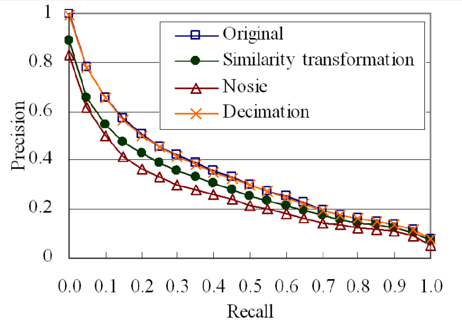
- Database of 1,833 3D models
- Annotated by a single human evaluator
- Models classified according to "functional similarities"
    - 47 "classes", covering 549 models
    - Rest 1,284 models classified as "miscellaneous"

Performance Evaluation



- **3D Harmonics:** discussed yesterday
- **Shape 3D Descriptor:** curvature histograms (MPEG 7 standard)
- **Multiple View Descriptor:** align using PCA

Robustness Evaluation



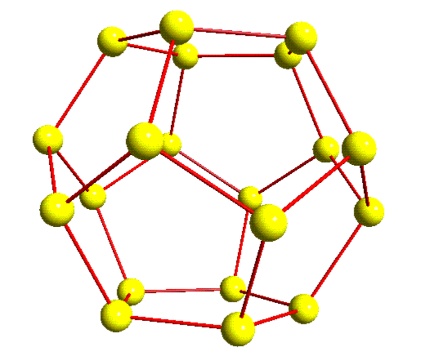
- **Similarity transformation:** rotation, translation & scaling
- **Noise:** vertex coordinates changed
- **Decimation:** randomly delete 20% polygons

Examples

Summary

- Not very concise
- Quick to compute (?)
- Not very efficient to match
- Good discrimination
- Invariant to transformations
- Invariant to deformations
- Insensitive to noise
- Insensitive to topology (?)
- Robust to degeneracies

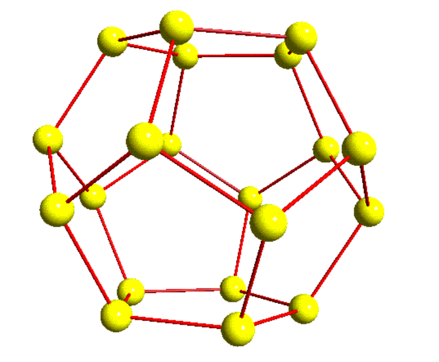
Regular Dodecahedron



Picture courtesy: http://btm2xl.mat.uni-bayreuth.de

Return

Regular Dodecahedron



Picture courtesy: http://btm2xl.mat.uni-bayreuth.de

Return

Zernike Moment Descriptors

- Zernike polynomials
  - A set of complex-valued functions over the unit circle
- Zernike moments of order n with repetition m

- $V_{nm}(x, y) = V_{nm}(\rho \cos\theta, \rho \sin\theta) = R_{nm}(\rho) \exp(jm\theta)$
- $R_{nm}(\rho) = \sum_{s=0}^{(n-|m|)/2} (-1)^s \frac{(n-s)!}{s!(\frac{n+|m|}{2}-s)!(\frac{n-|m|}{2}-s)!} \rho^{n-2s}$
- $A_{nm} = \frac{n+1}{\pi} \sum_{x,y:x^2+y^2 \leq 1} f(x, y) V^*_{nm}(x, y)$

Return

Zernike Moment Descriptors

- Zernike polynomials
  - A set of complex-valued functions over the unit circle
- Zernike moments of order n with repetition m

- $V_{nm}(x, y) = V_{nm}(\rho \cos\theta, \rho \sin\theta) = R_{nm}(\rho) \exp(jm\theta)$
- $R_{nm}(\rho) = \sum_{s=0}^{(n-|m|)/2} (-1)^s \frac{(n-s)!}{s!(\frac{n+|m|}{2}-s)!(\frac{n-|m|}{2}-s)!} \rho^{n-2s}$
- $A_{nm} = \frac{n+1}{\pi} \sum_{x,y:x^2+y^2\leq 1} f(x, y) V^*_{nm}(x, y)$

Return

Fourier Descriptors

- Fourier transform of a shape signature
- Centroid distance used as the shape signature
- Use only the magnitude of Fourier coefficients

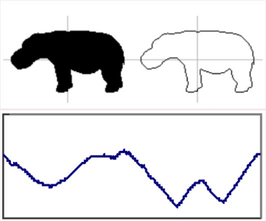
- $r(t) = ([x(t) - x_c]^2 + [y(t) - y_c]^2)^{1/2}$
- $x_c = \frac{1}{N}\sum_{t=0}^{N} x(t), \quad y_c = \frac{1}{N}\sum_{t=0}^{N} y(t)$
- $a_n = \frac{1}{N}\sum_{t=0}^{N} r(t) \exp(\frac{-j2\pi nt}{N})$
- $\mathbf{f} = \left\{ \frac{|a_1|}{|a_0|}, \frac{|a_2|}{|a_0|}, \ldots, \frac{|a_{N/2}|}{|a_0|} \right\}$

Return

Fourier Descriptors

- Fourier transform of a shape signature
- Centroid distance used as the shape signature
- Use only the magnitude of Fourier coefficients

- $r(t) = ([x(t) - x_c]^2 + [y(t) - y_c]^2)^{1/2}$
- $x_c = \frac{1}{N}\sum_{t=0}^{N} x(t), \quad y_c = \frac{1}{N}\sum_{t=0}^{N} y(t)$
- $a_n = \frac{1}{N}\sum_{t=0}^{N} r(t)\exp(\frac{-j2\pi nt}{N})$
- $\mathbf{f} = \left\{\frac{|a_1|}{|a_0|}, \frac{|a_2|}{|a_0|}, \ldots, \frac{|a_{N/2}|}{|a_0|}\right\}$

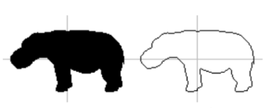


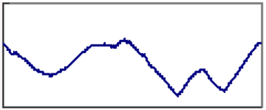
Return

Fourier Descriptors

- Fourier transform of a shape signature
- Centroid distance used as the shape signature
- Use only the magnitude of Fourier coefficients

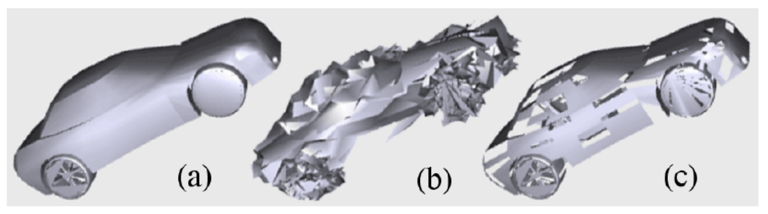
- $r(t) = ([x(t) - x_c]^2 + [y(t) - y_c]^2)^{1/2}$
- $x_c = \frac{1}{N}\sum_{t=0}^{N} x(t), \quad y_c = \frac{1}{N}\sum_{t=0}^{N} y(t)$
- $a_n = \frac{1}{N}\sum_{t=0}^{N} r(t)\exp(\frac{-j2\pi nt}{N})$
- $\mathbf{f} = \left\{\frac{|a_1|}{|a_0|}, \frac{|a_2|}{|a_0|}, \ldots, \frac{|a_{N/2}|}{|a_0|}\right\}$





Return

Model Corruption



- (a) Original 3D model
- (b) Effect of noise
- (c) Effect of decimation

Return