CS 468
Data-driven Shape Analysis

Scene Analysis and Synthesis

May 27, 2014
Motivation
Motivation

Retrieval

Fisher et al. 2010
Motivation

Retrieval

Input Scenes  Database  Synthesized Results

Synthesis

Fisher et al. 2010

Fisher et al. 2012
Scenes vs 3D models

Similarity

- A model consists of parts, a scene consists of objects
- Object-to-object & part-to-part relationships are not random
- Probabilistic models / matching techniques still apply
Scenes vs 3D models

Similarity
- A model consists of parts, a scene consists of objects
- Object-to-object & part-to-part relationships are not random
- Probabilistic models / matching techniques still apply

Differences
- Scene structure is more diverse (e.g. chairs vs living rooms)
- Typical scenes have plenty of object types
- Geometric similarity often is not descriptive enough
  (e.g. cabinet, nightstands, shelves, etc...)
 Scenes vs 3D models

Similarity
- A model consists of parts, a scene consists of objects
- Object-to-object & part-to-part relationships are not random
- Probabilistic models / matching techniques still apply

Differences
- Scene structure is more diverse (e.g. chairs vs living rooms)
- Typical scenes have plenty of object types
- Geometric similarity often is not descriptive enough
  (e.g. cabinet, nightstands, shelves, etc…)

What can help us?
Scenes vs 3D models

Similarity

- A model consists of parts, a scene consists of objects
- Object-to-object & part-to-part relationships are not random
- Probabilistic models / matching techniques still apply

Differences

- Scene structure varies more (e.g. chairs vs living rooms)
- Typical scenes have plenty of object types

Context is very important for scenes!

Fisher et al. 2010
Shape retrieval in scenes

Probability of an object $a$:

$$p(a|b, f_{ab}) \propto \sum_{c \in M} S_{cb}(\sigma_0) \Psi(a, c, f_{ab})$$
Shape retrieval in scenes

Probability of an object $a$:

$$p(a|b, f_{ab}) \propto \sum_{c \in M} S_{cb}(\sigma_0) \Psi(a, c, f_{ab})$$

Fisher et al. 2010

an object in a database
Shape retrieval in scenes

Probability of an object $a$:

$$p(a|b, f_{ab}) \propto \sum_{c \in M} S_{cb}(\sigma_0) \Psi(a, c, f_{ab})$$

- an object in a database
- the object in the scene
- relationship between a new object and existing object

Fisher et al. 2010
Shape retrieval in scenes

Probability of an object $a$:

$$p(a|b, f_{ab}) \propto \sum_{c \in M} S_{cb}(\sigma_0) \Psi(a, c, f_{ab})$$

- an object in a database
- the object in the scene
- for every object $c$ that's similar to $b$
- relationship between a new object and existing object

Fisher et al. 2010
Shape retrieval in scenes

Probability of an object $a$:

$$p(a|b, f_{ab}) \propto \sum_{c \in M} S_{cb}(\sigma_0) \Psi(a, c, f_{ab})$$

$$S_{st}(\sigma_{size}) = G(\text{Size}(s), \text{Size}(t), \sigma_{size}) K_{\text{model}}(s, t)$$

Similarity in model scales
Gaussian Kernel
Shape retrieval in scenes

Probability of an object $a$:

$$p(a|b, f_{ab}) \propto \sum_{c \in M} S_{cb}(\sigma_0) \Psi(a, c, f_{ab})$$

$$S_{st}(\sigma_{size}) = G(\text{Size}(s), \text{Size}(t), \sigma_{size}) K_{model}(s, t)$$

Similarity in model scales  
Gaussian Kernel  
Similarity after scaling

$$K_{model}(s, t) = 0.1 \delta_{st} + 0.6 K_{tag}(T_s, T_t) + 0.3 K_{geo}(s, t)$$

- exactly same model
- same text tag
- same shape descriptors
Shape retrieval in scenes

Probability of an object $a$:

$$p(a|b, f_{ab}) \propto \sum_{c \in M} S_{cb}(\sigma_0) \Psi(a, c, f_{ab})$$

For all $(u,v)$ that are similar to $(a,c)$ check if their spatial arrangement is similar (i.e. context similarity)

$$\Psi(a, c, f_{ab}) = \frac{\sum_{(u,v) \in O} S_{au}(\sigma_1) S_{cv}(\sigma_2) K(f_{ab}, f_{uv})}{\sum_{(u,v) \in O} S_{au}(\sigma_1) S_{cv}(\sigma_2)}$$
Shape retrieval in scenes

Probability of an object $a$:

$$p(a|b, f_{ab}) \propto \sum_{c \in M} S_{cb}(\sigma_0) \Psi(a, c, f_{ab})$$

For all $(u,v)$ that are similar to $(a,c)$ check if their spatial arrangement is similar (i.e. context similarity)

$$\Psi(a, c, f_{ab}) = \frac{\sum_{(u,v) \in O} S_{au}(\sigma_1) S_{cv}(\sigma_2) K(f_{ab}, f_{uv})}{\sum_{(u,v) \in O} S_{au}(\sigma_1) S_{cv}(\sigma_2)}$$

model-to-model similarity
Shape retrieval in scenes

Probability of an object $a$:

$$p(a|b, f_{ab}) \propto \sum_{c \in M} S_{cb}(\sigma_0) \Psi(a, c, f_{ab})$$

For all $(u,v)$ that are similar to $(a,c)$ check if their spatial arrangement is similar (i.e. context similarity)

$$K_{\text{spatial}}(f_{st}, f_{uv}) = G(z_{st}, z_{uv}, \sigma_z)G(r_{st}, r_{uv}, \sigma_r)$$

$$\Psi(a, c, f_{ab}) = \frac{\sum_{(u,v) \in O} S_{au}(\sigma_1) S_{cv}(\sigma_2) K(f_{ab}, f_{uv})}{\sum_{(u,v) \in O} S_{au}(\sigma_1) S_{cv}(\sigma_2)}$$
Shape retrieval in scenes

Probability of an object $a$:

$$p(a|b, f_{ab}) \propto \sum_{c \in M} S_{cb}(\sigma_0) \Psi(a, c, f_{ab})$$

For all $(u,v)$ that are similar to $(a,c)$ check if their spatial arrangement is similar (i.e. context similarity)

$$\Psi(a, c, f_{ab}) = \frac{\sum_{(u,v) \in O} S_{au}(\sigma_1) S_{cv}(\sigma_2) K(f_{ab}, f_{uv})}{\sum_{(u,v) \in O} S_{au}(\sigma_1) S_{cv}(\sigma_2)}$$

normalization
Shape retrieval in scenes

Probability of an object $a$:

$$p(a|b, f_{ab}) \propto \sum_{c \in M} S_{cb}(\sigma_0) \Psi(a, c, f_{ab})$$

Best $q=5$ neighbors

$$p(a|b_1, ... b_K, f_{ab_1}, ... f_{ab_K}) \propto \prod_{i=1}^{q} P_i$$

Fisher et al. 2010
Training scenes

Probability of an object $a$:

$$p(a|b, f_{ab}) \propto \sum_{c \in M} S_{cb}(\sigma_0) \Psi(a, c, f_{ab})$$

Training Examples?

Fisher et al. 2010
Training scenes

Leverage tags, scene graphs, and material similarity from Google Warehouse

<table>
<thead>
<tr>
<th>Name</th>
<th>Processed Tags</th>
</tr>
</thead>
<tbody>
<tr>
<td>LuminAiria</td>
<td>None</td>
</tr>
<tr>
<td>rouage</td>
<td>Cog</td>
</tr>
<tr>
<td>SimpleDome</td>
<td>Simple, Dome</td>
</tr>
<tr>
<td>11 by 8 paper</td>
<td>Paper</td>
</tr>
<tr>
<td>FineLiner</td>
<td>Fine, Liner</td>
</tr>
<tr>
<td>Box Deckel</td>
<td>Box, Cover, Lid</td>
</tr>
<tr>
<td>Box halter</td>
<td>Box, Halter</td>
</tr>
<tr>
<td>Stift</td>
<td>Pin, Pencil</td>
</tr>
<tr>
<td>Stiel</td>
<td>Handle, Stem</td>
</tr>
<tr>
<td>Eames Softpad Mgmt.  Chair</td>
<td>Soft, Pad, Management, Chair Seat</td>
</tr>
<tr>
<td>HSeat</td>
<td>None (frame is a stop word)</td>
</tr>
<tr>
<td>HFrame</td>
<td>iPhone</td>
</tr>
<tr>
<td>iPhone</td>
<td>Macbook, Pro, Open</td>
</tr>
</tbody>
</table>
Retrieval results

a) Keyword Search

Google 3D warehouse Fork

Fisher et al. 2010
Retrieval results

b) Context Search
Retrieval results

c) Context Search

Fisher et al. 2010
Retrieval results
Retrieval results

Fisher et al. 2010
How do we represent the entire scene?
Scenes as graphs

Fisher et al. 2011
Scenes as graphs

containment
contact
attachment
piercing
above

horizontal support
viewport equality
encircled

surface contact
adhesion
impaled
below
front
circlement
support
hanging
proximity
vertical equality
behind

Scene graph parent-child relationship
Surface contact relationship

Fisher et al. 2011
Scenes as graphs

Similarity of graphs as similarity of walks

Fisher et al. 2011
Scenes as graphs

\[ k^p_R(G_a, G_b, r, s) = \sum_{(r_1, e_1, ..., e_{p-1}, r_p) \in W^p_{G_a}(r)} \sum_{(s_1, f_1, ..., f_{p-1}, s_p) \in W^p_{G_b}(s)} k_{\text{node}}(r_p, s_p) \prod_{i=1}^{p-1} k_{\text{node}}(r_i, s_i) k_{\text{edge}}(e_i, f_i) \]

p-th order rooted walk graph kernel (measures similarity of walks originating in \( r \in G_a \) and \( s \in G_b \))

Fisher et al. 2011
Scenes as graphs

The p-th order rooted walk graph kernel (measures similarity of walks originating in \( r \in G_a \) and \( s \in G_b \)) is defined as:

\[
k^p_R(G_a, G_b, r, s) = \sum_{(r_1, e_1, \ldots, e_{p-1}, r_p) \in W^p_{G_a}(r)} \sum_{(s_1, f_1, \ldots, f_{p-1}, s_p) \in W^p_{G_b}(s)} \prod_{i=1}^{p-1} k_{\text{node}}(r_i, s_i) k_{\text{edge}}(e_i, f_i)
\]

Efficient dynamic programming algorithm to fill values:

\[
k^p_R(G_a, G_b, r, s) = k_{\text{node}}(r, s) \times \sum_{r' \in N_{G_a}(r)} \sum_{s' \in N_{G_b}(s)} k_{\text{edge}}(e, f) k^{p-1}_R(G_a, G_b, r', s')
\]

Fisher et al. 2011
Scenes as graphs

E.g. walk similarity:
\[0.6 \times 1 \times 0.9 \times 1 \times 0.4 = 0.22\]

Fisher et al. 2011
Scenes as graphs

Graph kernel:

\[
 k^p_G(G_a, G_b) = \sum_{r \in V_{G_a}} \sum_{s \in V_{G_b}} k^p_R(G_a, G_b, r, s)
\]

Graph distance:

\[
 d(G_a, G_b) = \sqrt{k^p_G(G_a, G_a) - 2k^p_G(G_a, G_b) + k^p_G(G_b, G_b)}
\]

Fisher et al. 2011
Scene retrieval results

Query

Result

Fisher et al. 2011
Scene retrieval results

Query

Result

Fisher et al. 2011
Scene retrieval results

Query

Result

Fisher et al. 2011
Context-aware suggestions

Query Scene

Suggested Models
Scene synthesis
Scene synthesis

Contextual categories

- Objects A, B that appear in the same context
Scene synthesis

Contextual categories

- Objects $A, B$ that appear in the same context

\[
n(A, B) = 1\{\text{isLeaf}(A) = \text{isLeaf}(B)\} \cdot G\left(\frac{|A| - |B|}{\min(|A|, |B|)}, \sigma_s\right) \cdot \sum_{(a, b) \in M} k(a, b)
\]
Scene synthesis

Contextual categories

- Objects A, B that appear in the same context

\[ n(A, B) = \begin{cases} 1 \{ \text{isLeaf}(A) = \text{isLeaf}(B) \} \cdot G \left( \frac{|A| - |B|}{\min(|A|, |B|)} \cdot \sigma_s \right) \cdot \sum_{(a, b) \in M} k(a, b) \end{cases} \]
Scene synthesis

Contextual categories

- Objects A, B that appear in the same context

\[ n(A, B) = 1\{\text{isLeaf}(A) = \text{isLeaf}(B)\} \cdot G\left(\frac{|A| - |B|}{\min(|A|, |B|)}, \sigma_s\right) \cdot \sum_{(a,b) \in M} k(a, b) \]

Same size of objects
Scene synthesis

Contextual categories

- Objects A, B that appear in the same context

\[ n(A, B) = 1 \{ \text{isLeaf}(A) = \text{isLeaf}(B) \} \cdot \]

\[ G \left( \frac{|A| - |B|}{\min(|A|, |B|)}, \sigma_s \right) \cdot \sum_{(a,b) \in M} k(a, b) \]

Neighborhood similarity
Scene synthesis

Contextual categories

- Objects A, B that appear in the same context

\[ n(A, B) = 1 \{ \text{isLeaf}(A) = \text{isLeaf}(B) \} \cdot G\left( \frac{|A| - |B|}{\min(|A|, |B|)}, \sigma_s \right) \cdot \sum_{(a, b) \in M} k(a, b) \]

Neighborhood similarity
(use Hungarian algorithm to decide which neighboring pairs to match)
Scene synthesis

Contextual categories
  - Objects A, B that appear in the same context

\[ n(A, B) = 1 \{ \text{isLeaf}(A) = \text{isLeaf}(B) \} \cdot G \left( \frac{|A| - |B|}{\min(|A|, |B|)}, \sigma_s \right) \cdot \sum_{(a,b) \in M} k(a, b) \]

Neighborhood similarity

\[ k(a, b) = 1 \{ L^a = L^b \} \cdot G(|\vec{a} - \vec{b}|, \sigma_d) \cdot G(\min(|\vec{a}|, |\vec{b}|), \sigma_n) \]
Scene synthesis

Contextual categories

○ Objects $A$, $B$ that appear in the same context

$$n(A, B) = 1\{\text{isLeaf}(A) = \text{isLeaf}(B)\} \cdot \frac{|A| - |B|}{\min(|A|, |B|)} \cdot G\left(\frac{|\vec{a} - \vec{b}|}{\sigma_d}\right) \cdot \sum_{(a, b) \in M} k(a, b)$$

$\text{Neighborhood similarity}$

$$k(a, b) = 1\{L^a = L^b\} \cdot G(|\vec{a} - \vec{b}|, \sigma_d) \cdot G(\min(|\vec{a}|, |\vec{b}|), \sigma_n)$$

- Same label
- Objects are close
- Distance to reference objects $A$ and $B$
Scene synthesis

Contextual categories

- Objects A, B that appear in the same context
  - Agglomerative clustering to decide on equivalent contextual groups
Scene synthesis

Input

Basic categories

Contextual categories

Objects A, B that appear in same context

Agglomerative clustering to decide on equivalent contextual groups

Fisher et al. 2012
Object Placement

Learn pairwise spatial distribution

$P_{\text{speaker|desk}}$

$P_{\text{mouse|keyboard}}$

$P_{\text{desk chair|monitor}}$

$P_{\text{dining chair|table}}$

Fisher et al. 2012
Scene Synthesis Results

Input Scenes

Synthesized Results

Fisher et al. 2012
Scene Synthesis Results

Input Scenes

Synthesized Results

Fisher et al. 2012
Interactive Synthesis

Input

Output

Yu et al. 2011
Interactive Synthesis

Learn:

- Object-to-wall relations
- Support relations
- Object-to-object relations
- Pathways connecting doors

Yu et al. 2011
Make it Home: Automatic Optimization of Furniture Arrangement

Lap-Fai Yu*  Sai-Kit Yeung*
Chi-Keung Tang**  Demetri Terzopoulos*
Tony F. Chan**  Stanley J. Osher*

*University of California, Los Angeles
**Hong Kong University of Science and Technology
Interactive Synthesis

Hardcode design rules

Merell et al. 2011
Interactive Synthesis

VIDEO
Sketch-based Scene Synthesis

Xu et al. 2013
Sketch-based Scene Synthesis

\[ s(C') = P(C) \cdot \sum_{\forall U \subset C, 2 \leq |U| \leq 5} \alpha(|U|) f(G(U)) \cdot s_{\text{sub}}(U) \]

Scale consistency

\[ P(C) = \exp(-\lambda_p \cdot \max_{1 \leq i \leq M} |\log d(o_i) - \frac{1}{M} \sum_{1 \leq i \leq M} \log d(o_i)|) \]
Sketch-based Scene Synthesis

Retrieved set $C$

$$s(C') = \frac{P(C)}{\max U \subset C, 2 \leq |U| \leq 5} \cdot \sum a(|U|) f(G(U)) \cdot s_{\text{sub}}(U)$$

Scale consistency: similar scale for all objects

$$P(C') = \exp(-\lambda_p \cdot \max_i \log d(o_i) - \frac{1}{M} \sum_{1 \leq i \leq M} \log d(o_i))$$

Scale of each object to fit to sketch
Sketch-based Scene Synthesis

Run-time Stage
Input Sketches

Sketch-based Co-retrieval

Candidate Retrieval
TV desk piano monitor
photo mouse cellphone

Retrieved set $C$

$$s(C') = P(C) \cdot \sum_{\forall U \subset C, 2 \leq |U| \leq 5} a(|U|) f(G(U)) \cdot s_{sub}(U)$$

Scale consistency: similar scale for all obj

$$P(C) = \exp(-\lambda_p \cdot \max_{1 \leq i \leq M} |\log d(o_i) - \frac{1}{M} \sum_{1 \leq i \leq M} \log d(o_i)|)$$

scale of each object to fit to sketch

DISCLAIMER: ignores perspective

Xu et al. 2013
Sketch-based Scene Synthesis

**Retrieved set** $C$

$$s(C') = P(C) \cdot \sum_{\forall U \subset C, 2 \leq |U| \leq 5} a(|U|) f(G(U)) \cdot s_{sub}(U)$$

Check different subsets
Sketch-based Scene Synthesis

Retrieved set $C$

$$s(C') = P(C) \cdot \sum_{\forall U \subseteq C, 2 \leq |U| \leq 5} a(|U|) f(G(U)) \cdot s_{sub}(U)$$

Same context

$$s_{sub}(U) = \prod_{1 \leq i \leq p} m_S(u_i) \cdot \prod_{1 \leq i,j \leq p} (g_{i,j}(u_i, u_j))^\lambda_c$$
Sketch-based Scene Synthesis

Retrieved set $C$

$$s(C') = P(C) \cdot \sum_{\forall U \subset C, 2 \leq |U| \leq 5} a(|U|) f(G(U)) \cdot s_{sub}(U)$$

- Same context
- Sketch-to-object similarity
- Spatial arrangement similarity
Sketch-based Scene Synthesis

Place objects based on learned relationships

\[ s_{sub}(U) = \prod_{1 \leq i \leq p} m_S(u_i) \cdot \prod_{1 \leq i, j \leq p} ((1 - w_a)g_{i,j}(u_i, u_j) + w_a t(u_i, u_j))^{\lambda_c} \]

Sketch-to-object similarity

Spatial arrangement similarity

Old and new arrangement similarity

Xu et al. 2013
Sketch-based Scene Synthesis

Place objects based on learned relationships

\[ s_{\text{sub}}(U) = \prod_{1 \leq i \leq p} m_S(u_i) \cdot \prod_{1 \leq i, j \leq p} ((1 - w_a)g_{i,j}(u_i, u_j) + w_a t(u_i, u_j))^{\lambda_c} \]

\( w_a = 1 \)
Sketch-based Scene Synthesis

Place objects based on learned relationships

\[ s_{\text{sub}}(U) = \prod_{1 \leq i \leq p} m_S(u_i) \cdot \prod_{1 \leq i, j \leq p} ((1 - w_a)g_{i,j}(u_i, u_j) + w_a t(u_i, u_j))^\lambda_c \]

\[ w_a = 0 \]
Sketch-based Scene Synthesis

Place objects based on learned relationships

\[ s_{\text{sub}}(U) = \prod_{1 \leq i \leq p} m_S(u_i) \cdot \prod_{1 \leq i, j \leq p} ((1 - w_a) g_{i,j}(u_i, u_j) + w_at(u_i, u_j))^{\lambda_c} \]

\[ W_a = .1 \]
Sketch-based Scene Synthesis

Place objects based on learned relationships

\[ s_{\text{sub}}(U) = \prod_{1 \leq i \leq p} m_S(u_i) \cdot \prod_{1 \leq i, j \leq p} ((1 - w_a)g_{i,j}(u_i, u_j) + w_a t(u_i, u_j))^\lambda_c \]

\( w_a = 0.1 \)
Value of context

(a) input sketches  (b) 1-sketch result  (c) 2-sketch result

(d) 3-sketch result  (e) 5-sketch result  (f) 7-sketch result

Xu et al. 2013
Sketch2Scene results

Example 1
Input Sketches: [Sketch Image]
Artist-modeled Scene: [Scene Image]
Result 1: [Result Image]
Result 2: [Result Image]
Result 3: [Result Image]

Example 2
Input Sketches: [Sketch Image]
Artist-modeled Scene: [Scene Image]
Result 1: [Result Image]
Result 2: [Result Image]
Result 3: [Result Image]

Example 3
Input Sketches: [Sketch Image]
Artist-modeled Scene: [Scene Image]
Result 1: [Result Image]
Result 2: [Result Image]
Result 3: [Result Image]
Sketch2Scene results

Xu et al. 2013
VIDEO
Related Topics

Computer Vision
- Scene understanding: segmentation, parsing
- Depth estimation in images

Computer Graphics
- Scene reconstruction
References

- Beyond categories: The visual memex model for reasoning about object relationships. T. Malisiewicz, A. Efros, NIPS 2009
- Image classification with segmentation graph kernels. Z. Harchaoui, F. Bach, CVPR 2007
- Context-Based Search for 3D Models. M. Fisher, P. Hanrahan. SIGGRAPH Asia 2010