CS 468

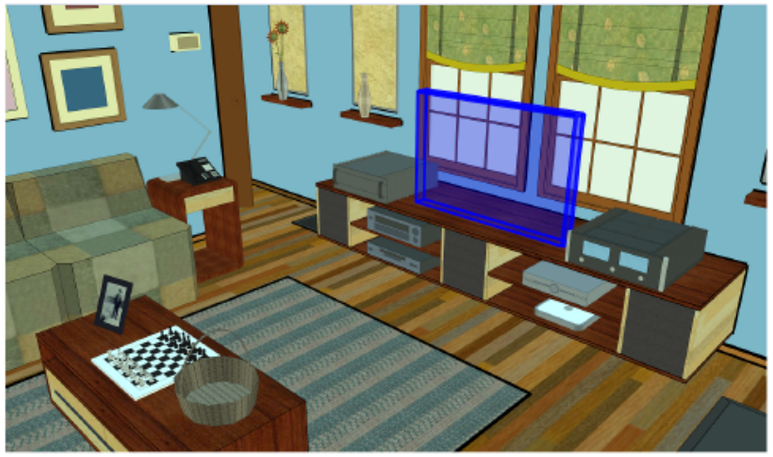
Data-driven Shape Analysis

Scene Analysis and Synthesis

May 27, 2014

Motivation

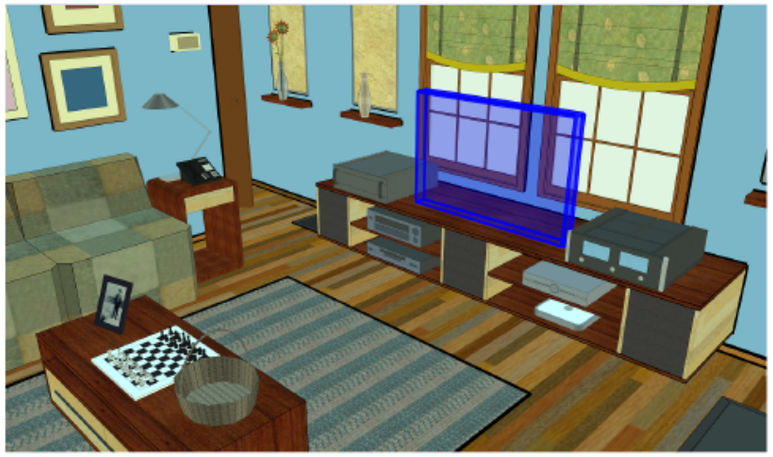
Motivation



Retrieval

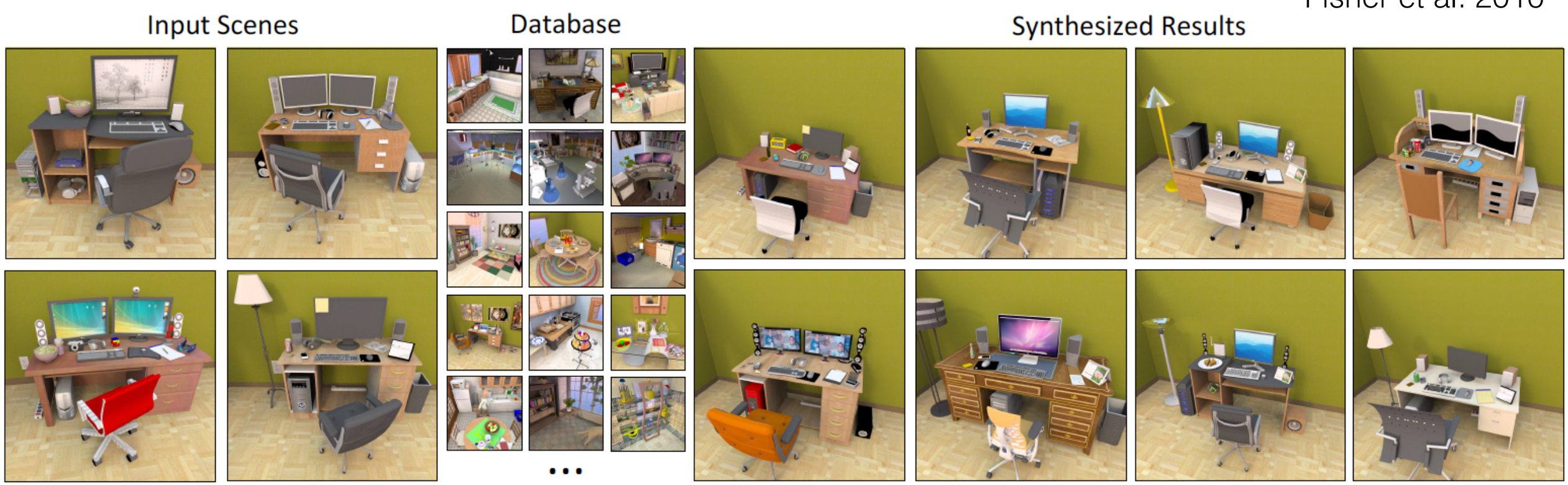
Fisher et al. 2010

Motivation



Retrieval

Fisher et al. 2010



Synthesis

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

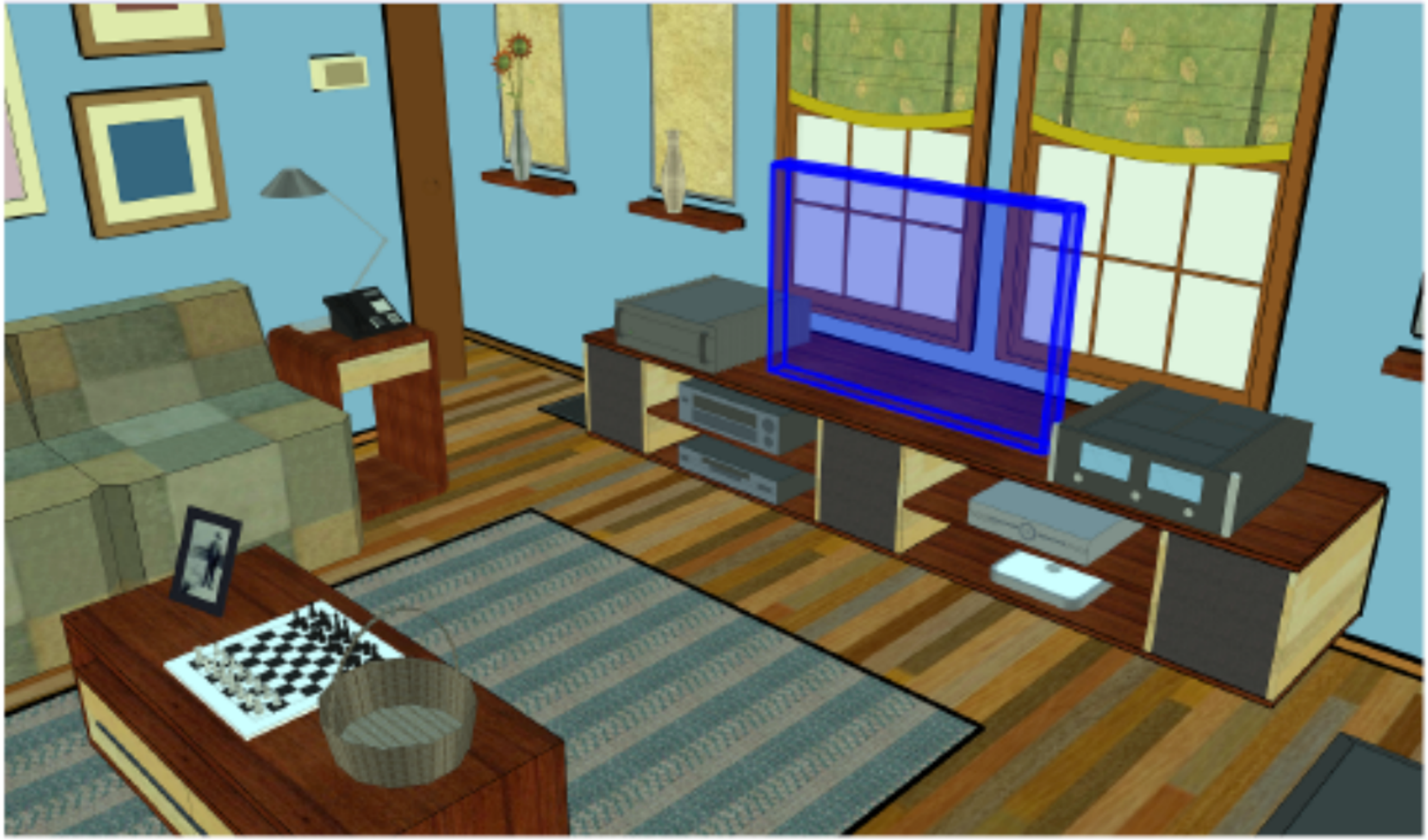
Similarity

- 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?



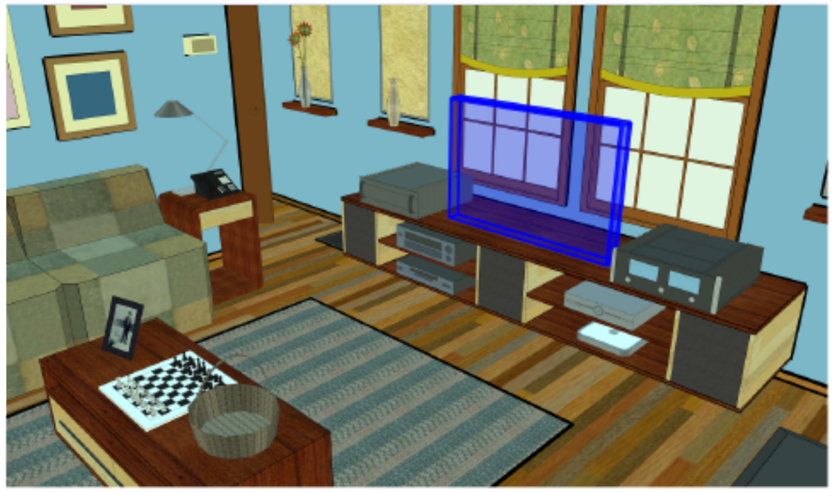
Context is very important for scenes!

Fisher et al. 2010

Shape retrieval in scenes

Probability of an object *a*:

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



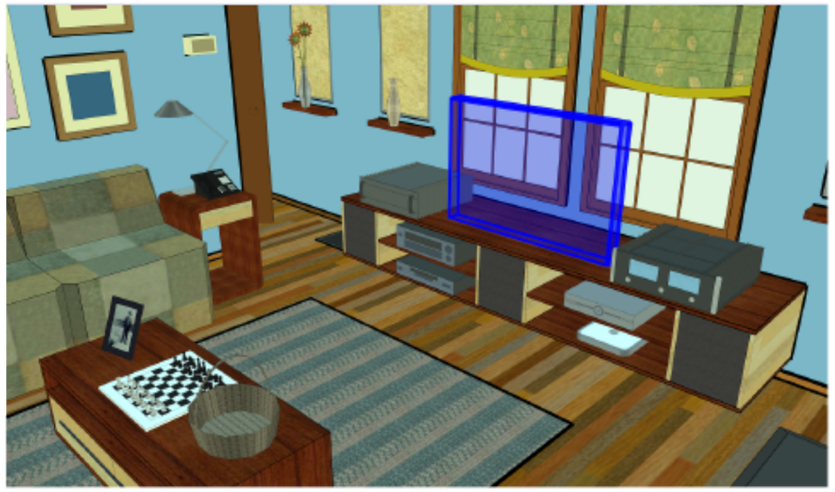
Fisher et al. 2010

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



Fisher et al. 2010

Shape retrieval in scenes

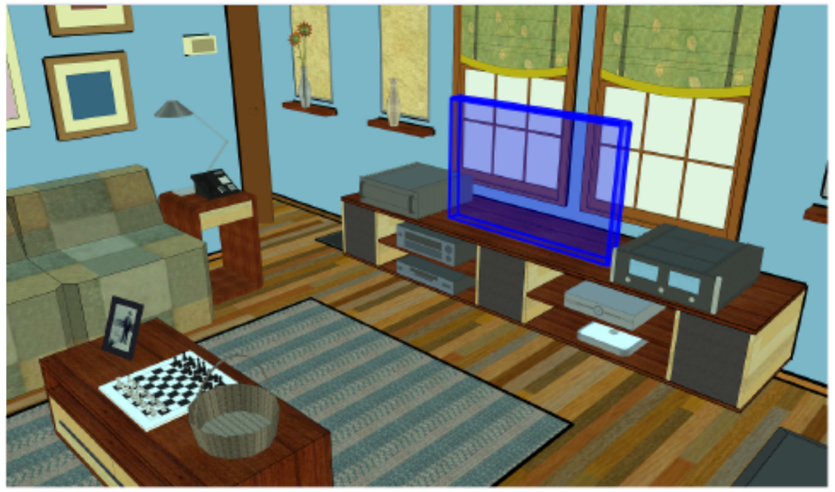
Probability of an object *a*:

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

an object in a database

the object in the scene

relationship between a new object and existing object



Fisher et al. 2010

Shape retrieval in scenes

Probability of an object *a*:

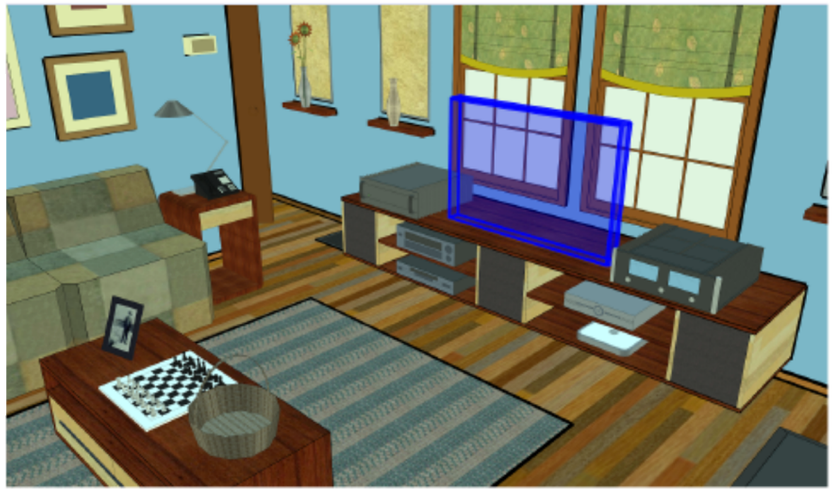
$$p(a|b, f_{ab}) \propto \sum_{c \in M} S_{cb}(\sigma_0)\Psi(a, c, f_{ab})$$

an object in a database

the object in the scene

relationship between a new object and existing object

for every object *c* that's similar to *b*

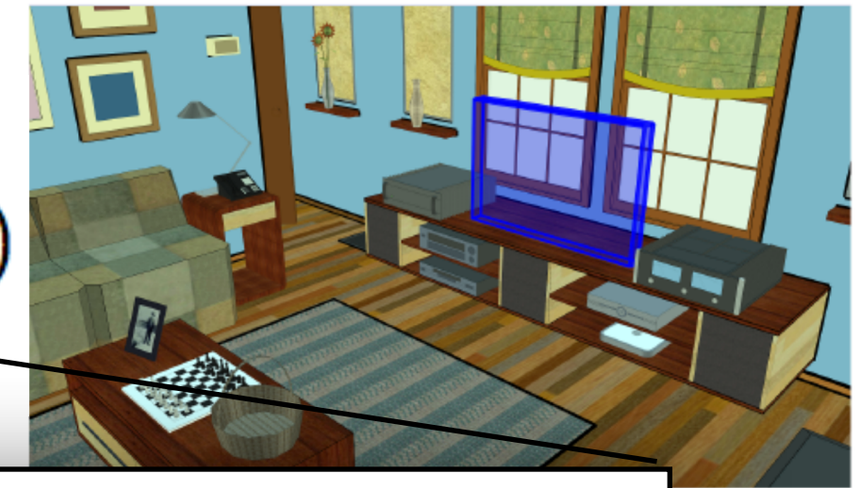


Fisher et al. 2010

Shape retrieval in scenes

Probability of an object *a*:

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



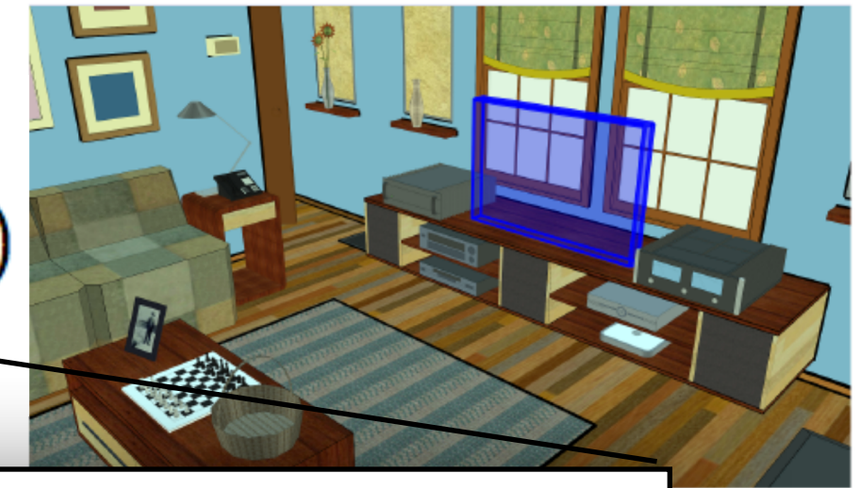
$$S_{st}(\sigma_{\textbf{size}}) = \underline{G(\mathrm{Size}(s), \mathrm{Size}(t), \sigma_{\textbf{size}})} K_{\textbf{model}}(s, t)$$

Similarity in model scales
Gaussian Kernel

Shape retrieval in scenes

Probability of an object *a*:

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



$$S_{st}(\sigma_{\textbf{size}}) = \underbrace{G(\mathrm{Size}(s), \mathrm{Size}(t), \sigma_{\textbf{size}})}_{\text{Similarity in model scales}} \underbrace{K_{\textbf{model}}(s, t)}_{\text{Similarity after scaling}}$$

Gaussian Kernel

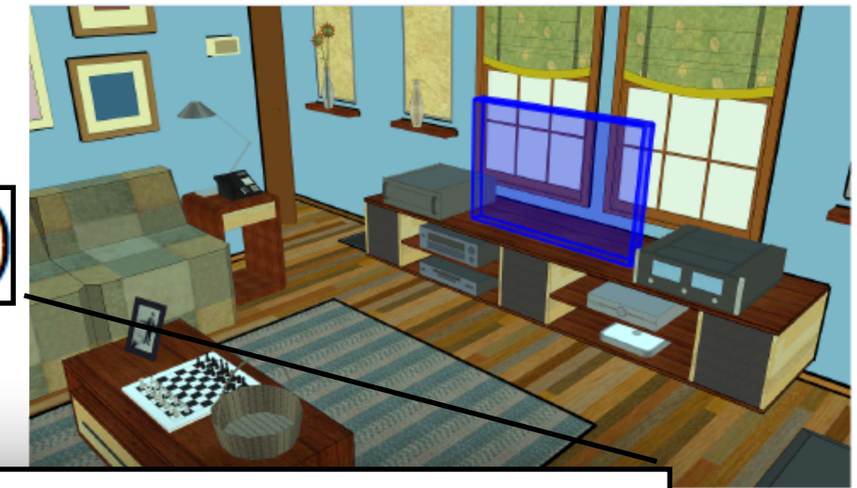
$$K_{\textbf{model}}(s, t) = 0.1\delta_{st} + 0.6K_{\textbf{tag}}(T_s, T_t) + 0.3K_{\textbf{geo}}(s, t)$$

exactly same model **same text tag** **same shape descriptors**

Shape retrieval in scenes

Probability of an object *a*:

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



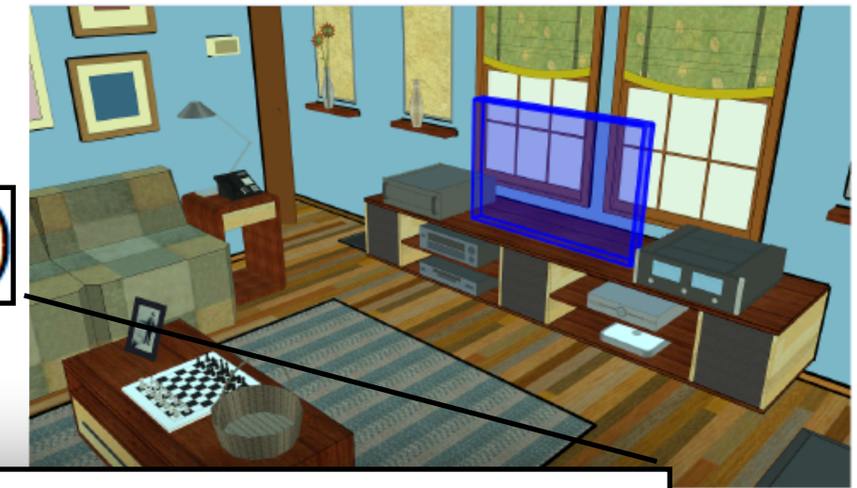
For all (*u,v*) that are similar to (*a,c*) check if their spatial arrangement is similar (i.e. context similarity)

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

Shape retrieval in scenes

Probability of an object *a*:

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



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

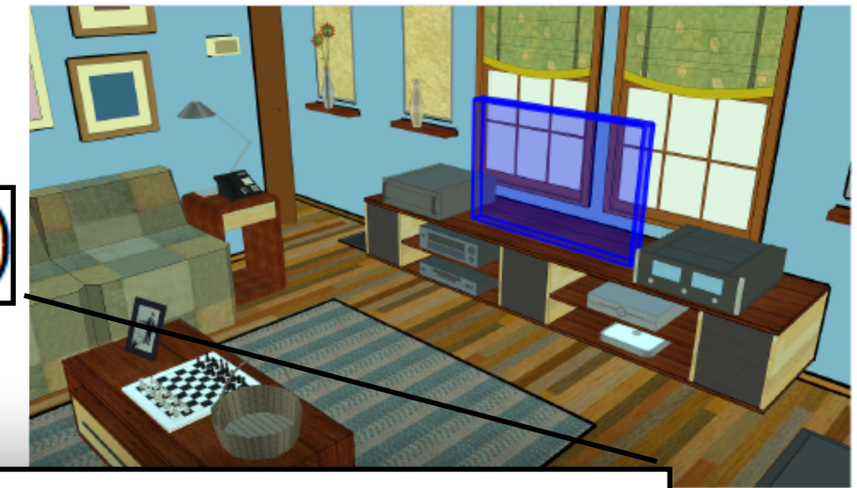
model-to-model similarity

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

Shape retrieval in scenes

Probability of an object *a*:

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



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

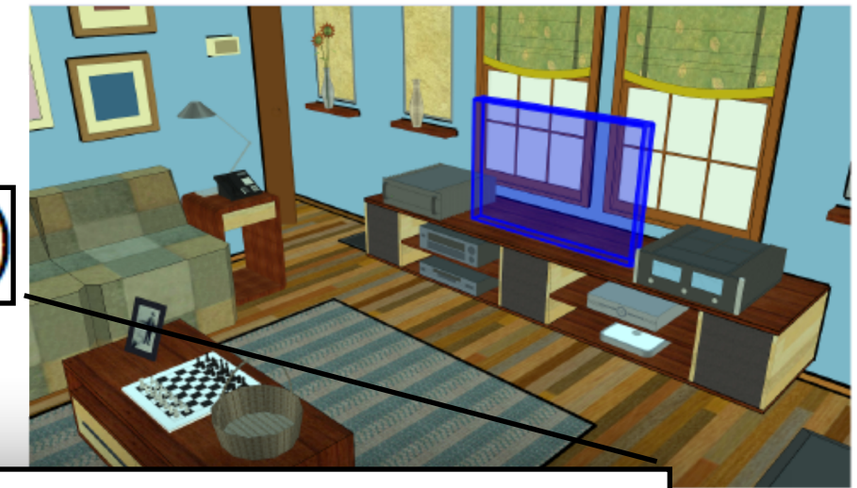
$$K_{\text{spatial}}(f_{st}, f_{uv}) = G(z_{st}, z_{uv}, \sigma_z) G(r_{st}, r_{uv}, \sigma_r)$$

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

Shape retrieval in scenes

Probability of an object *a*:

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



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

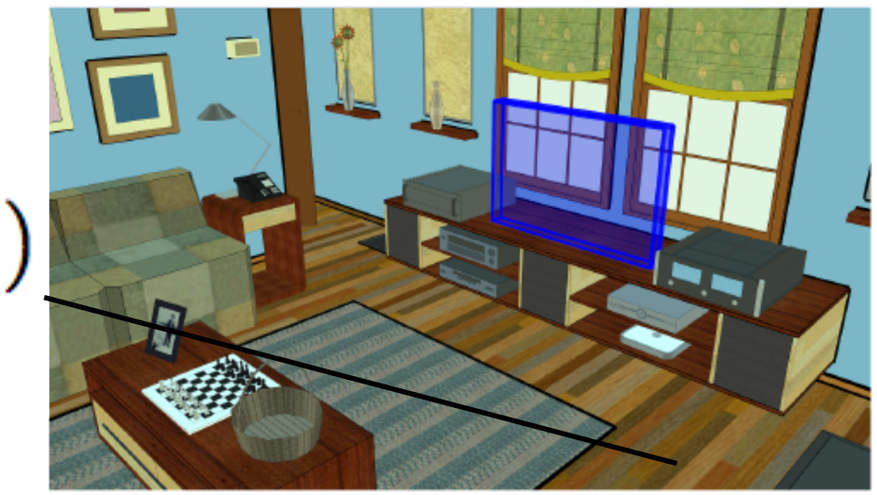
$$\Psi(a, c, f_{ab}) = \frac{\sum_{(u,v) \in O} S_{au}(\sigma_1) S_{cv}(\sigma_2) K(f_{ab}, f_{uv})}{\boxed{\sum_{(u,v) \in O} S_{au}(\sigma_1) S_{cv}(\sigma_2)}}$$

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

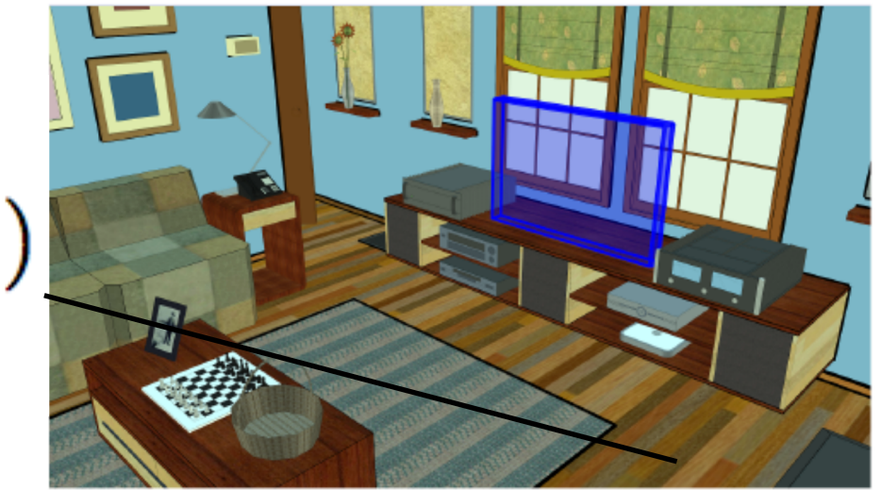
$$p(a|b_1, ...b_K, f_{ab_1}, ...f_{ab_K}) \propto \prod_{i=1}^{q} P_i$$

Fisher et al. 2010

Training scenes

Probability of an object *a*:

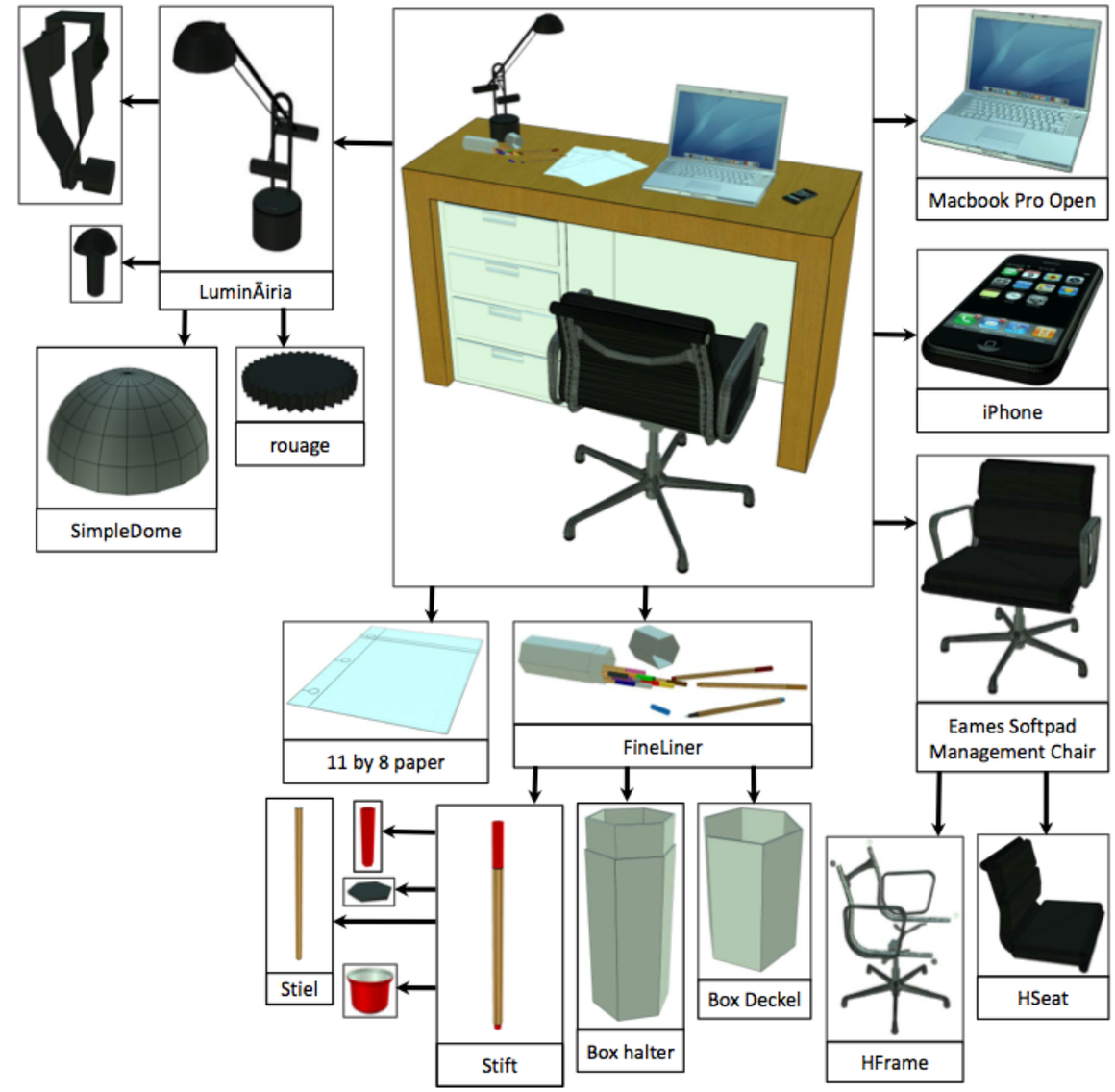
$$p(a|b, f_{ab}) \propto \sum_{c \in M} S_{cb}(\sigma_0)\Psi(a, c, f_{ab})$$



Training Examples?

Fisher et al. 2010

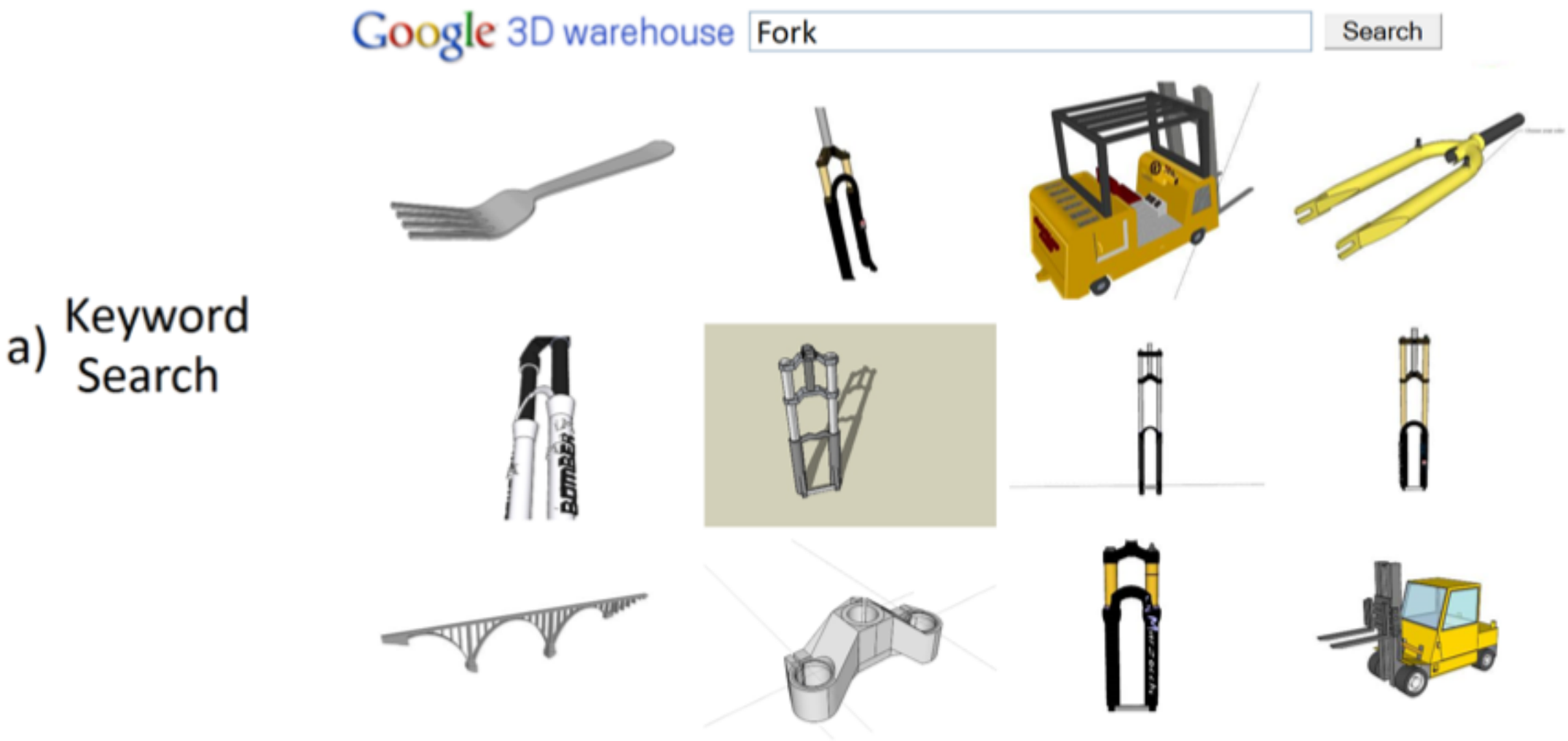
Training scenes



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

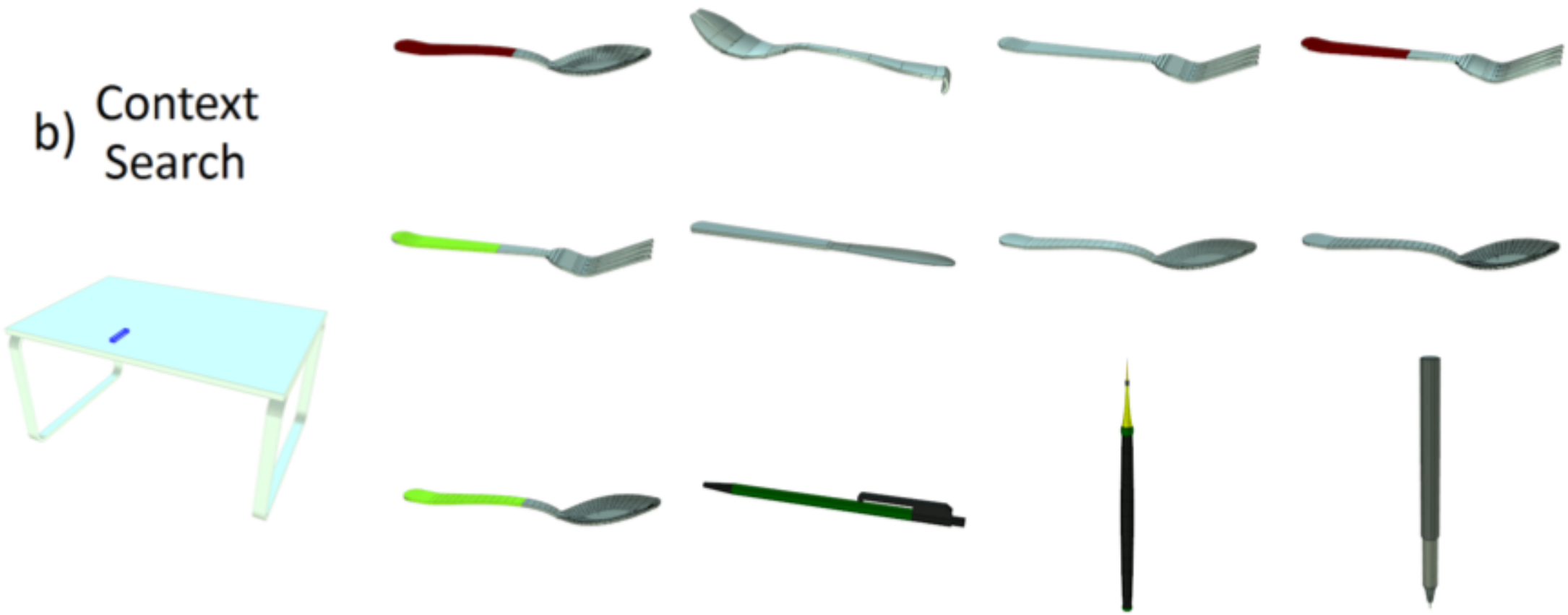
Leverage tags, scene graphs, and material similarity from Google Warehouse

Retrieval results



Fisher et al. 2010

Retrieval results



Fisher et al. 2010

Retrieval results

c) Keyword and Context Search

Fisher et al. 2010

Retrieval results

Fisher et al. 2010

Retrieval results

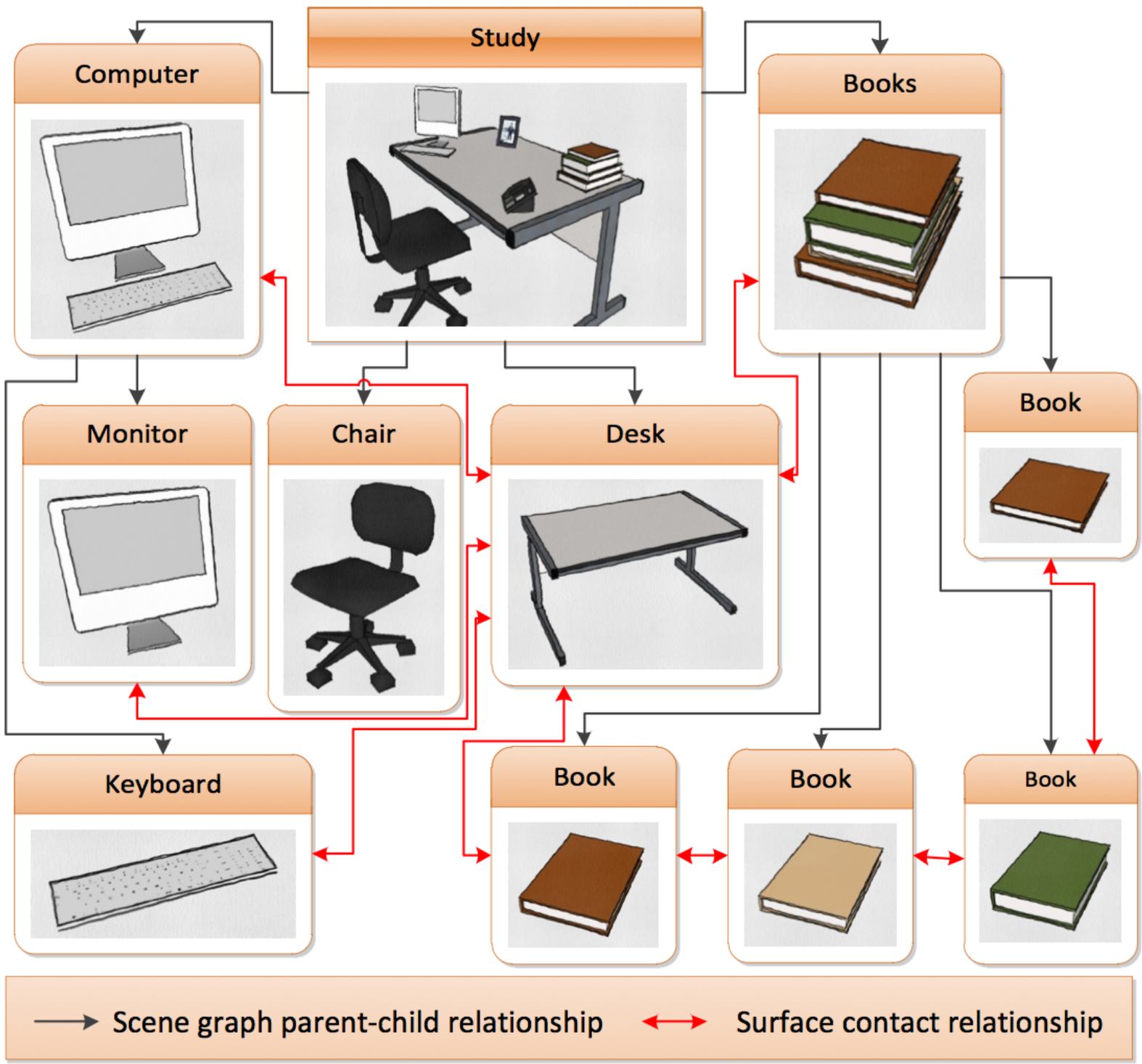
Fisher et al. 2010

Retrieval results

Fisher et al. 2010

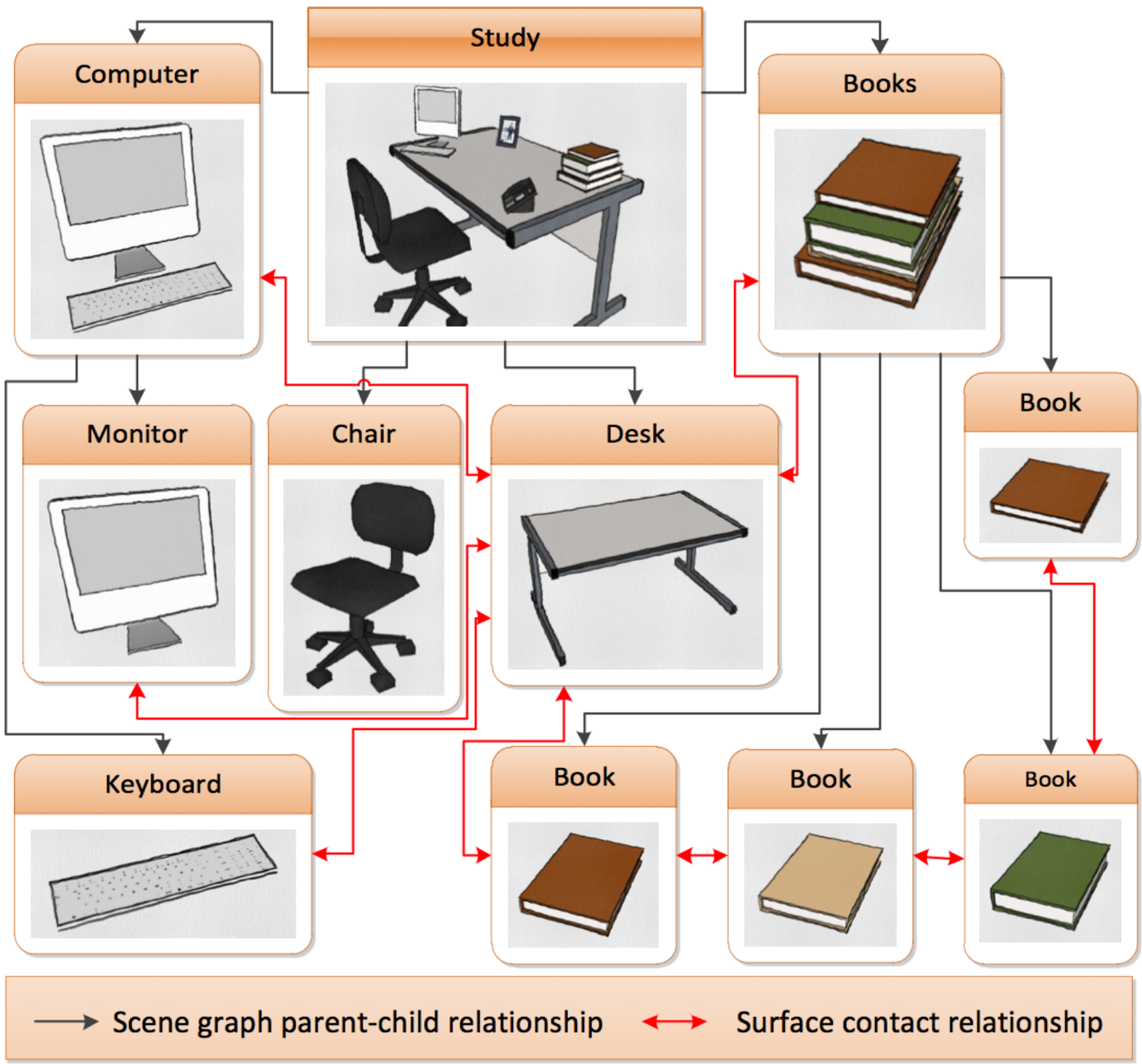
How do we represent the entire scene?

Scenes as graphs



Fisher et al. 2011

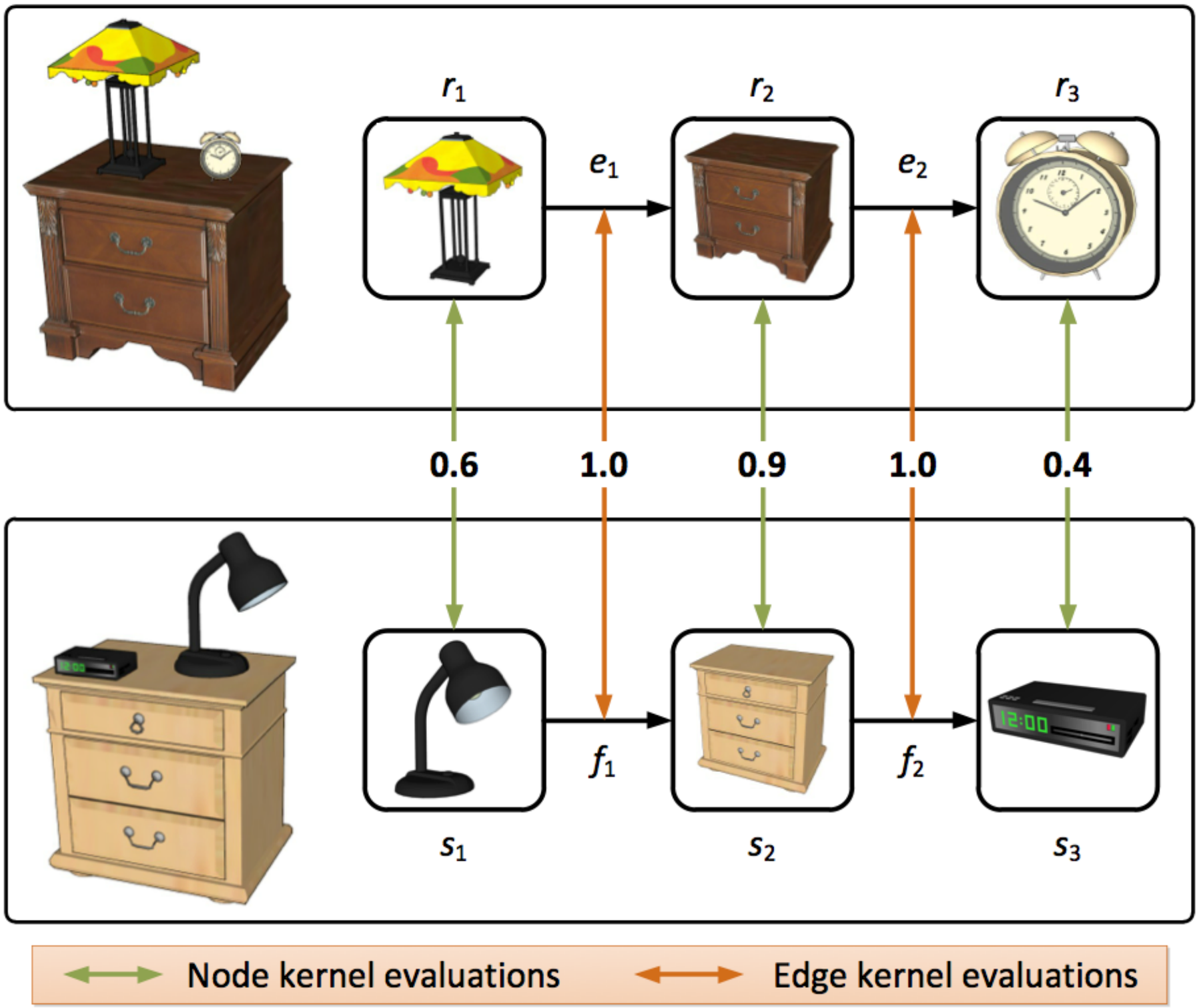
Scenes as graphs



→ **containment**
contact
attachment
piercing
above
→ **horizontal support**
viewport equality
encircled
→ **surface contact**
adhesion
impaled
below
front
circlement
support
hanging
proximity
vertical equality
behind

Fisher et al. 2011

Scenes as graphs



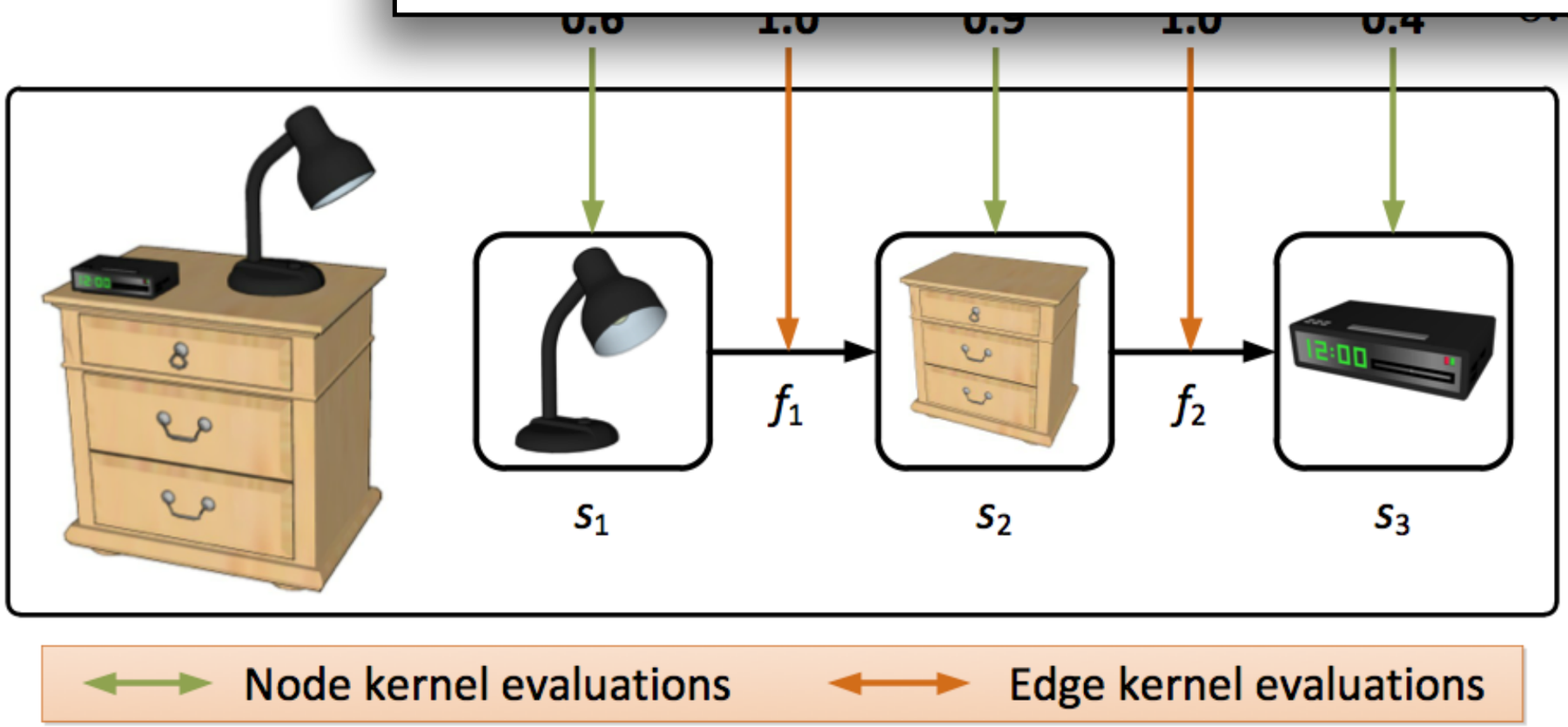
Similarity of graphs as similarity of walks

Fisher et al. 2011

Scenes as graphs

$$k_R^p(G_a, G_b, r, s) = \sum_{\substack{(r_1, e_1, \ldots, e_{p-1}, r_p) \in W_{G_a}^p(r) \\ (s_1, f_1, \ldots, f_{p-1}, s_p) \in W_{G_b}^p(s)}} k_{\text{node}}(r_p, s_p) \prod_{i=1}^{p-1} k_{\text{node}}(r_i, s_i) k_{\text{edge}}(e_i, f_i)$$

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

$$k_R^p(G_a, G_b, r, s) = \sum_{\substack{(r_1, e_1, \ldots, e_{p-1}, r_p) \in W_{G_a}^p(r) \\ (s_1, f_1, \ldots, f_{p-1}, s_p) \in W_{G_b}^p(s)}} k_{\text{node}}(r_p, s_p) \prod_{i=1}^{p-1} k_{\text{node}}(r_i, s_i) k_{\text{edge}}(e_i, f_i)$$

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

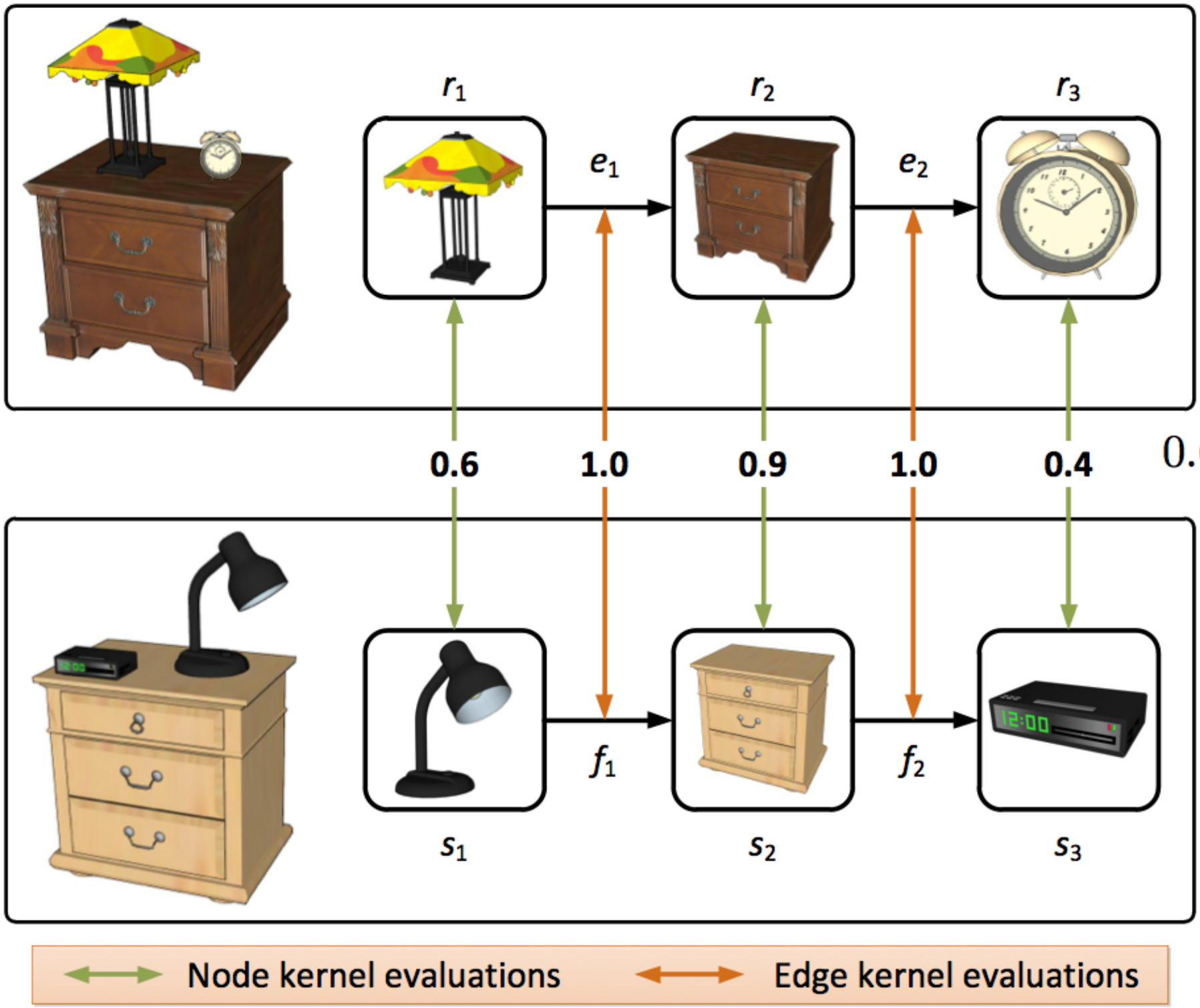
Efficient dynamic programming algorithm to fill values:

$$k_R^p(G_a, G_b, r, s) = k_{\text{node}}(r, s) \times \sum_{\substack{r' \in N_{G_a}(r) \\ s' \in N_{G_b}(s)}} k_{\text{edge}}(e, f) k_R^{p-1}(G_a, G_b, r', s')$$

Node kernel evaluations Edge kernel evaluations

Fisher et al. 2011

Scenes as graphs

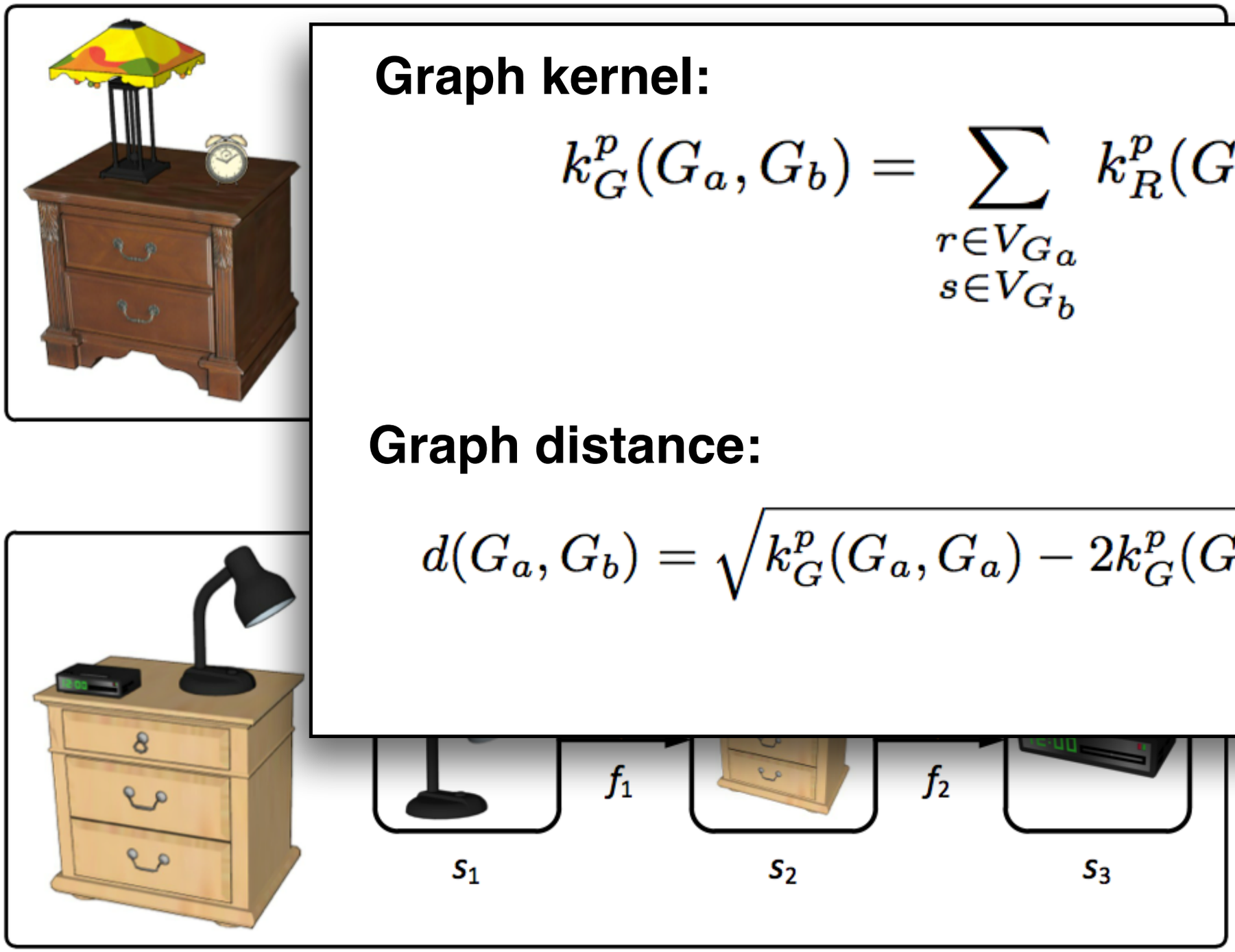


E.g. walk similarity:

$$0.6 * 1 * 0.9 * 1 * 0.4 = 0.22$$

Fisher et al. 2011

Scenes as graphs



Graph kernel:

$$k_G^p(G_a, G_b) = \sum_{\substack{r \in V_{G_a} \\ s \in V_{G_b}}} k_R^p(G_a, G_b, r, s)$$

Graph distance:

$$d(G_a, G_b) = \sqrt{k_G^p(G_a, G_a) - 2k_G^p(G_a, G_b) + k_G^p(G_b, G_b)}$$

Node kernel evaluations Edge kernel evaluations

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

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

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

Scene synthesis

Contextual categories

- Objects A, B that appear in the same context

Same support role

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

Scene synthesis

Contextual categories

- Objects A, B that appear in the same context

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

Same size of objects

Scene synthesis

Contextual categories

- Objects A, B that appear in the same context

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

Neighborhood similarity

Scene synthesis

Contextual categories

- Objects A, B that appear in the same context

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

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

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

Neighborhood similarity

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

Scene synthesis

Contextual categories

- Objects A, B that appear in the same context

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

Neighborhood similarity

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

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

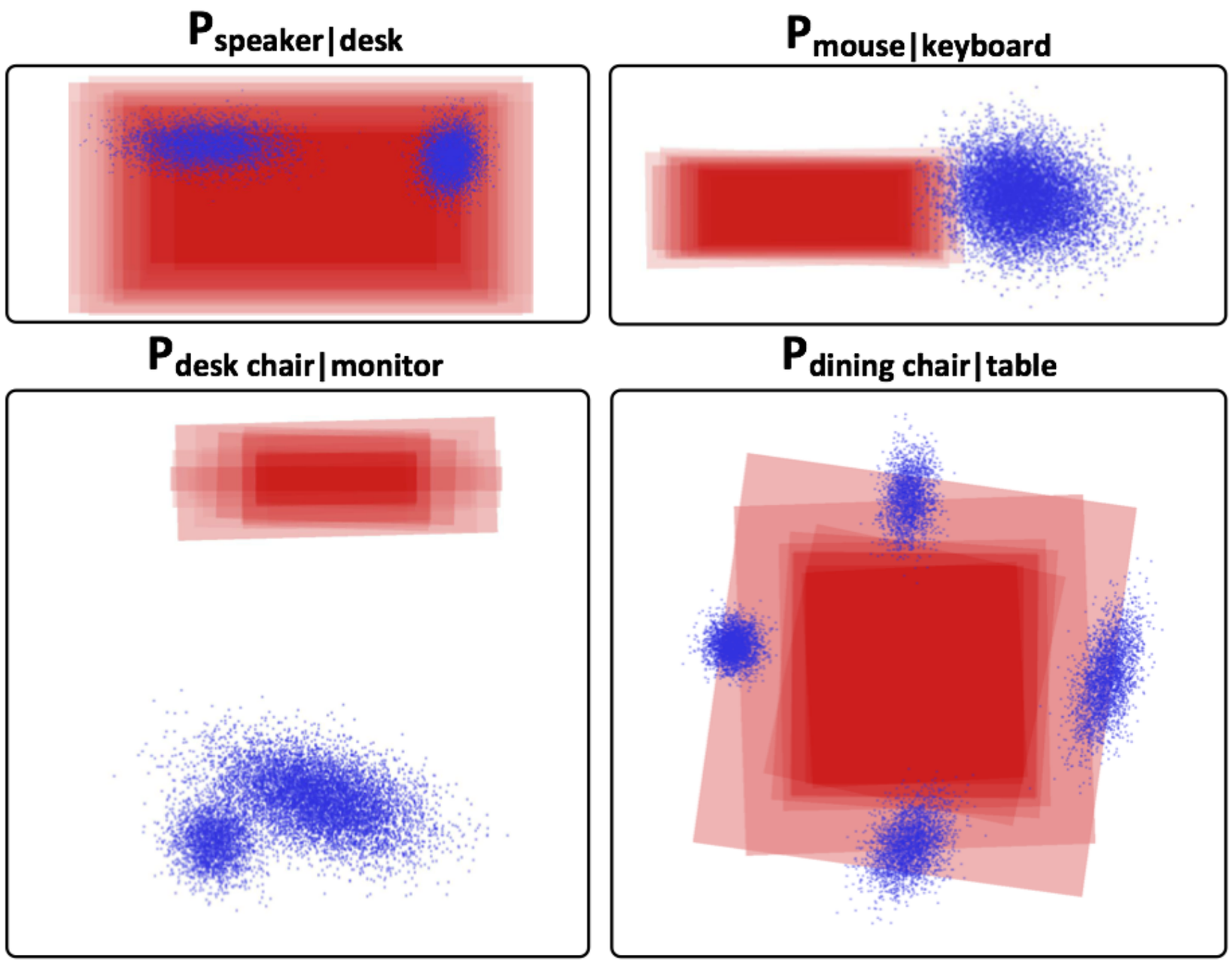
Scene synthesis

Input | Basic categories | Contextual categories

Fisher et al. 2012

Object Placement

Learn pairwise spatial distribution



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

Input

Output

Learn:

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

Yu et al. 2011

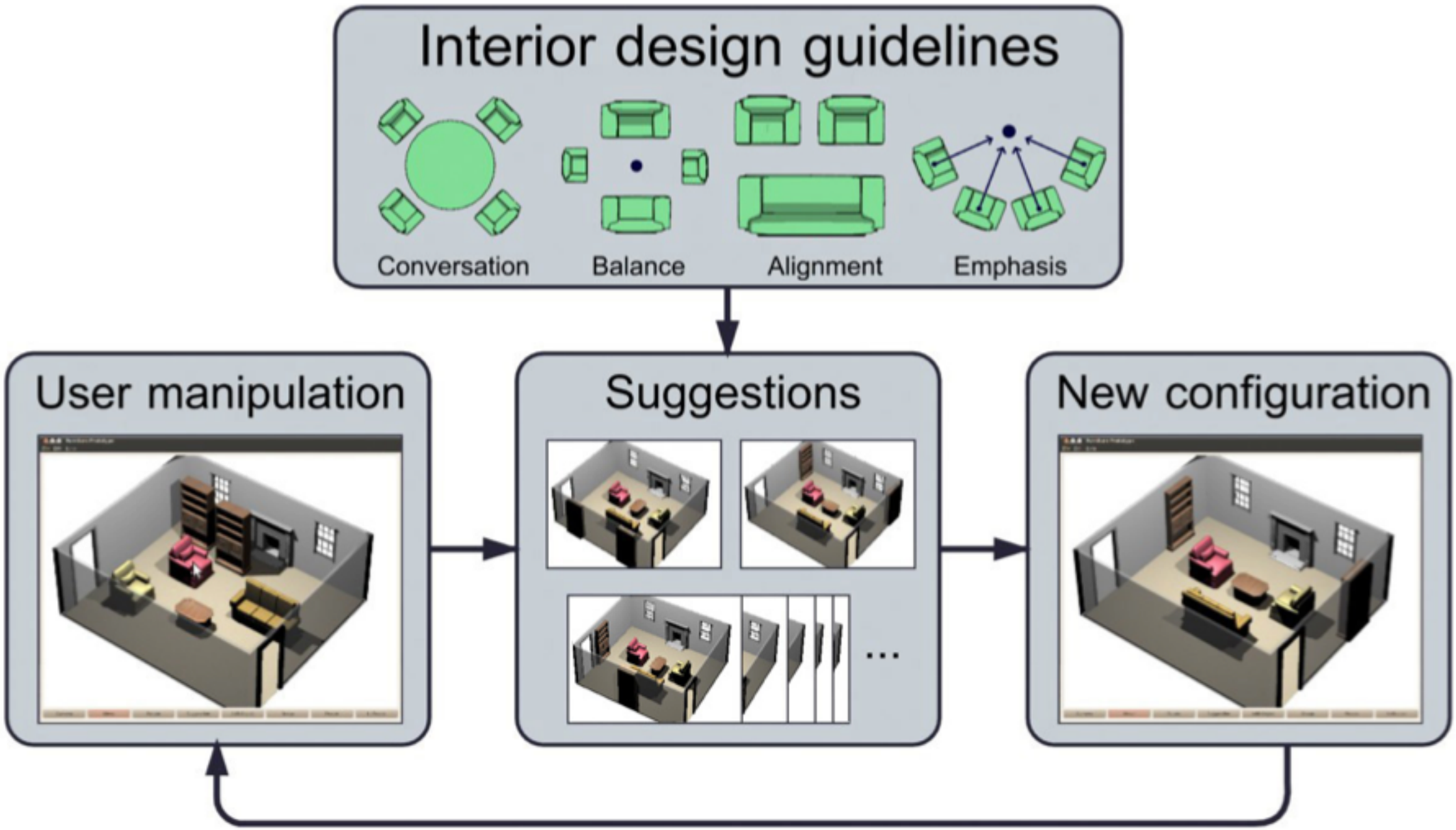
Make it Home: Automatic Optimization of Furniture Arrangement

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



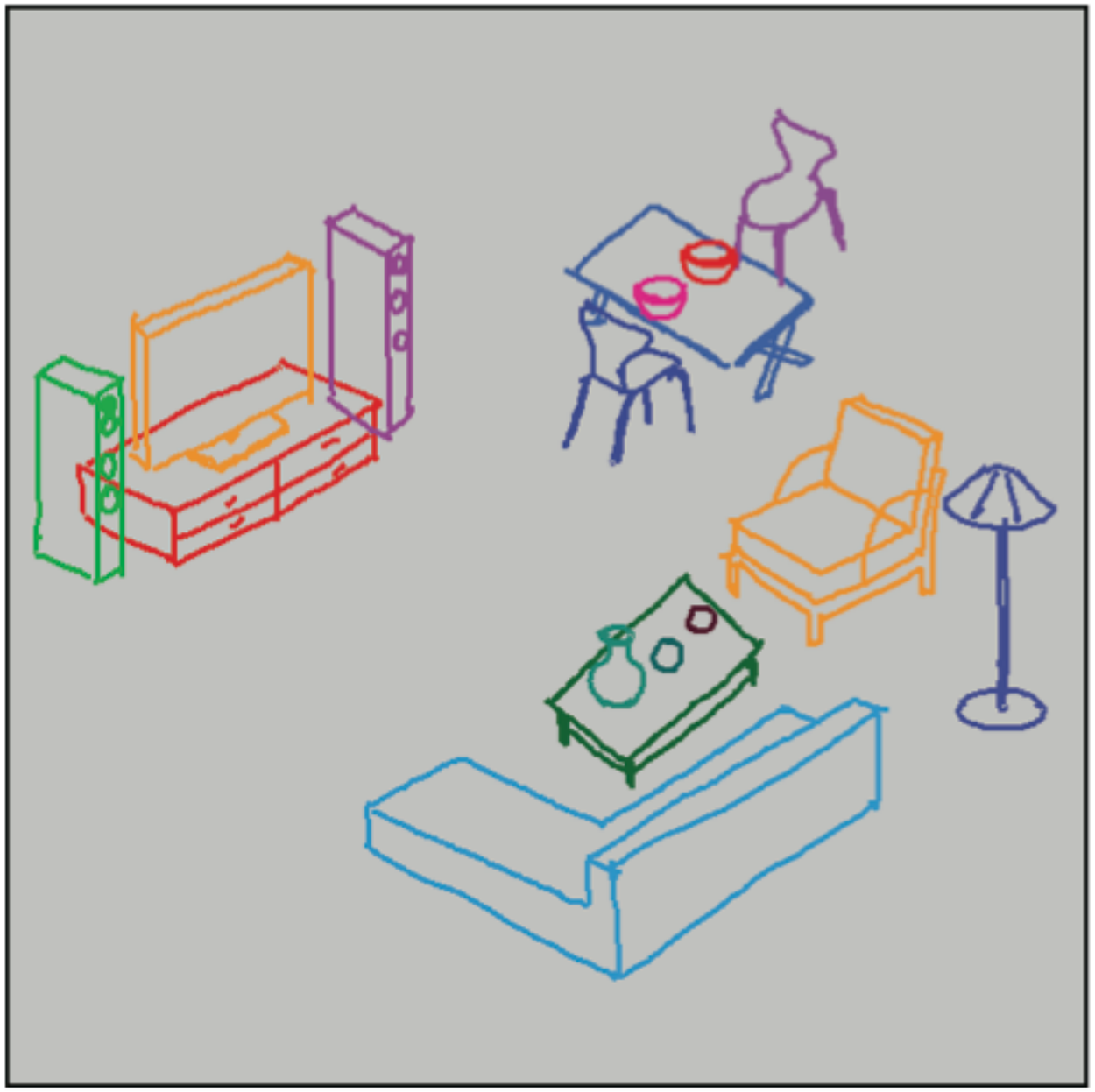
Merell et al. 2011

Interactive Synthesis

VIDEO

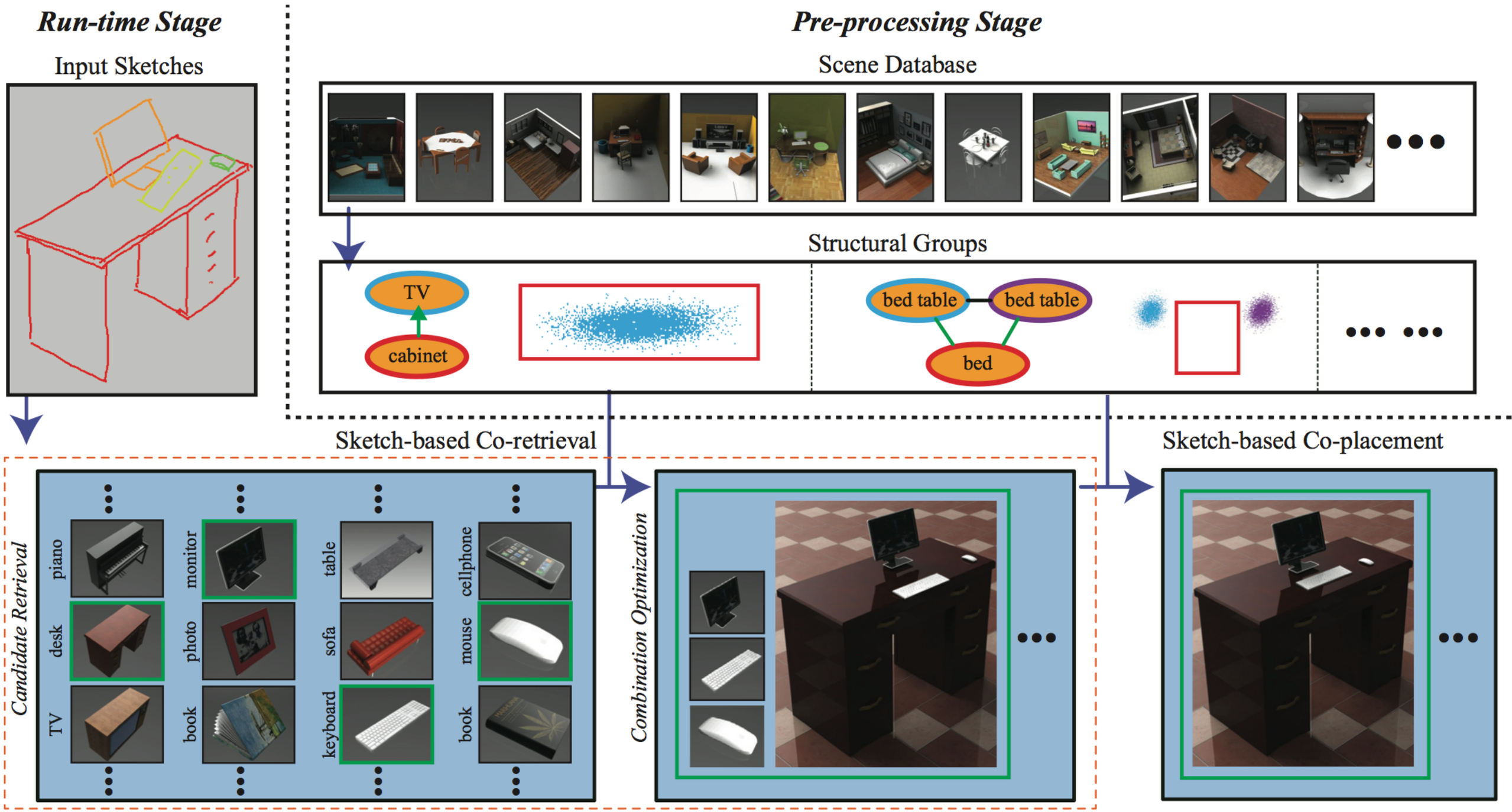
Merell et al. 2011

Sketch-based Scene Synthesis



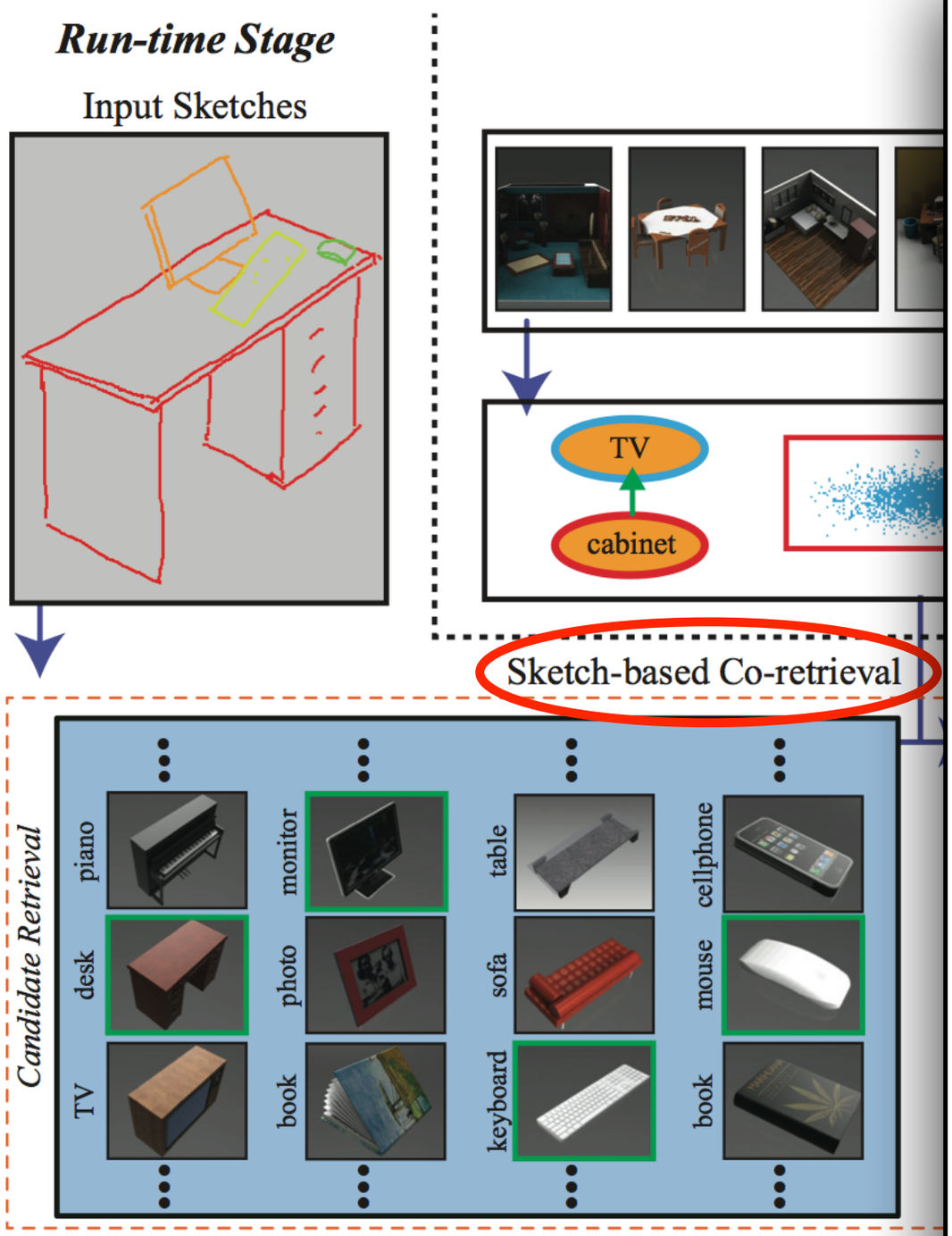
Xu et al. 2013

Sketch-based Scene Synthesis



Xu et al. 2013

Sketch-based Scene Synthesis



Retrieved set *C*

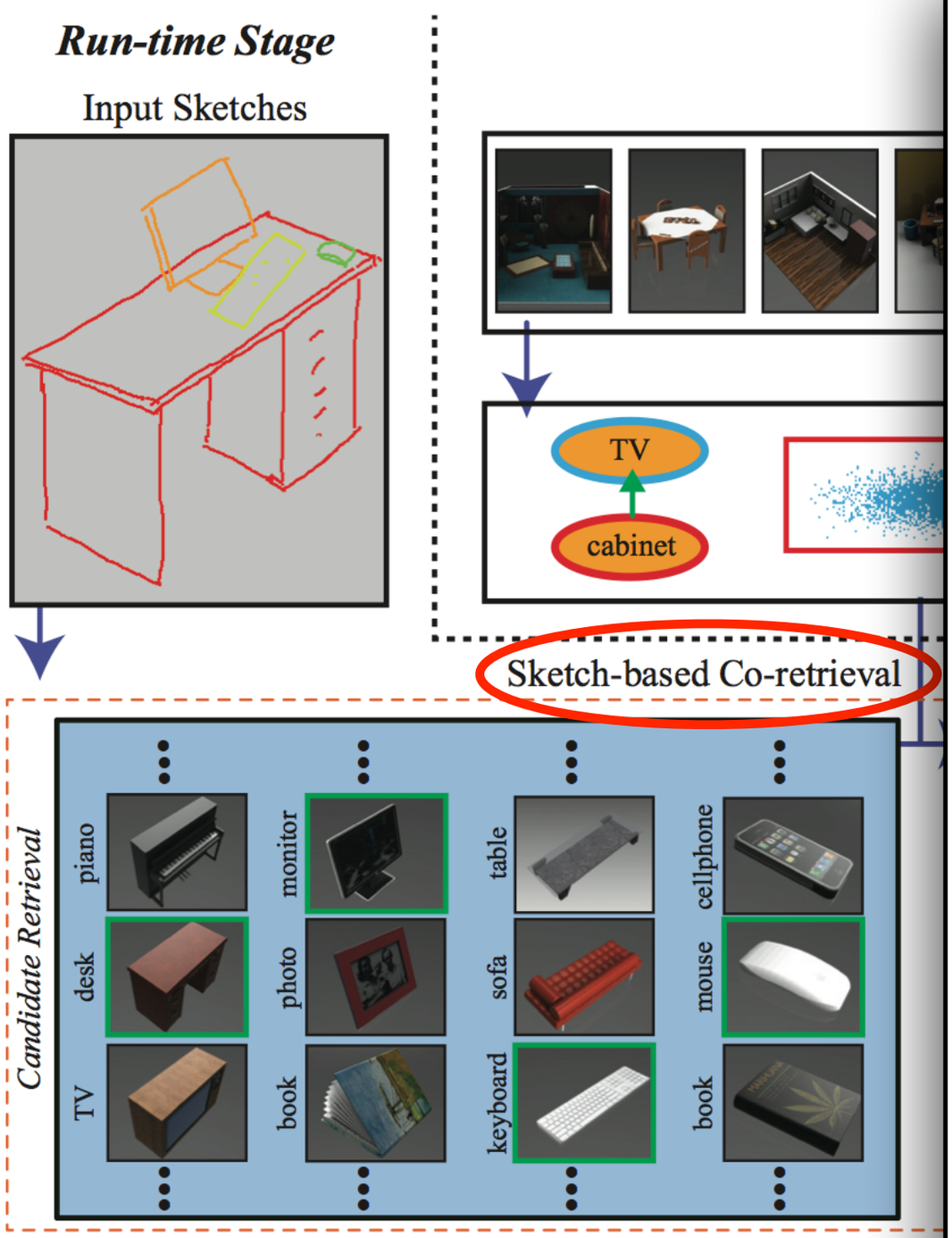
$$s(C) = \boxed{P(C)} \cdot \sum_{\forall U \subset C, 2 \le |U| \le 5} a(|U|) f(G(U)) \cdot s_{\text{sub}}(U)$$

Scale consistency

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

Xu et al. 2013

Sketch-based Scene Synthesis



Retrieved set *C*

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

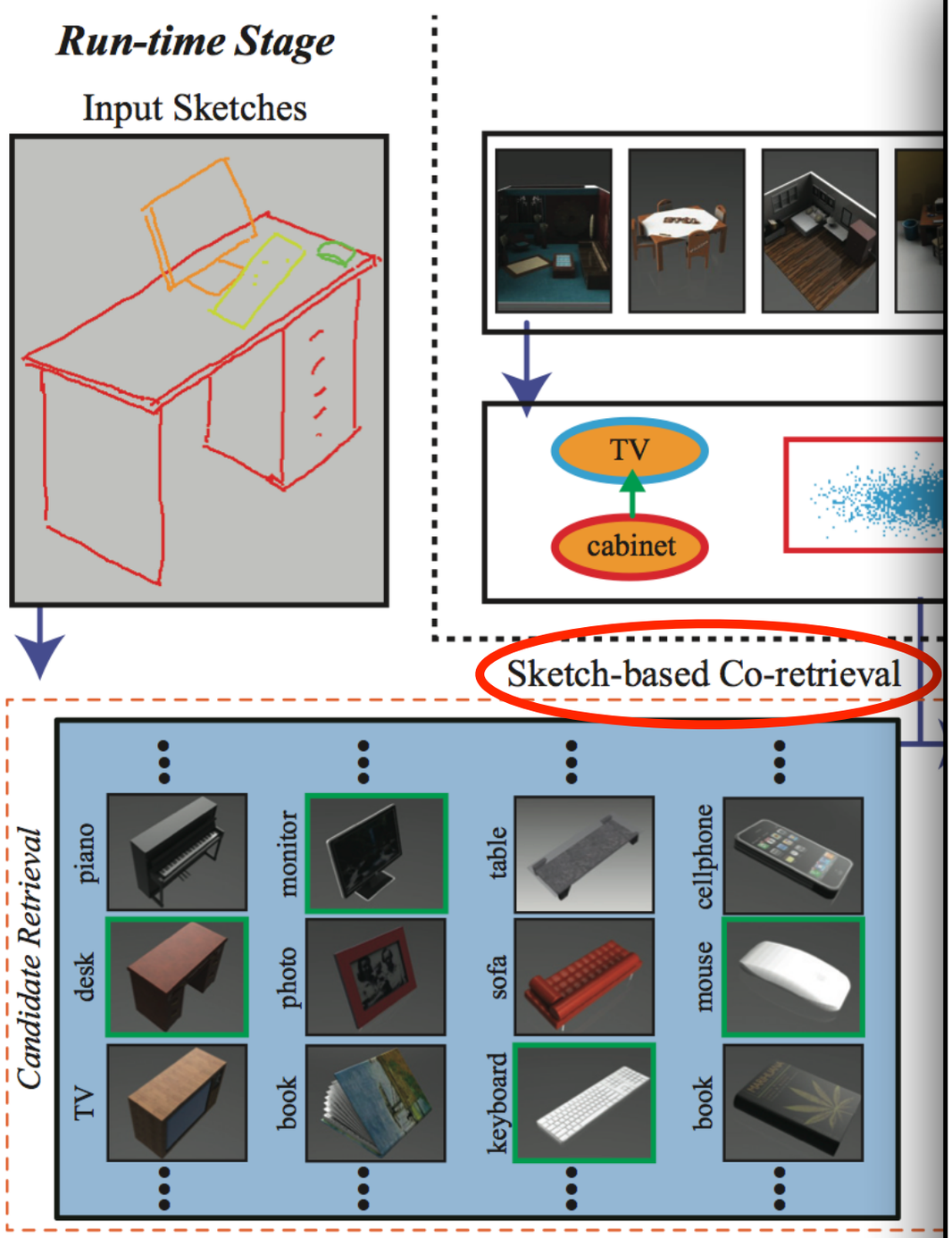
Scale consistency: similar scale for all obj

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

scale of each object to fit to sketch

Xu et al. 2013

Sketch-based Scene Synthesis



Retrieved set *C*

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

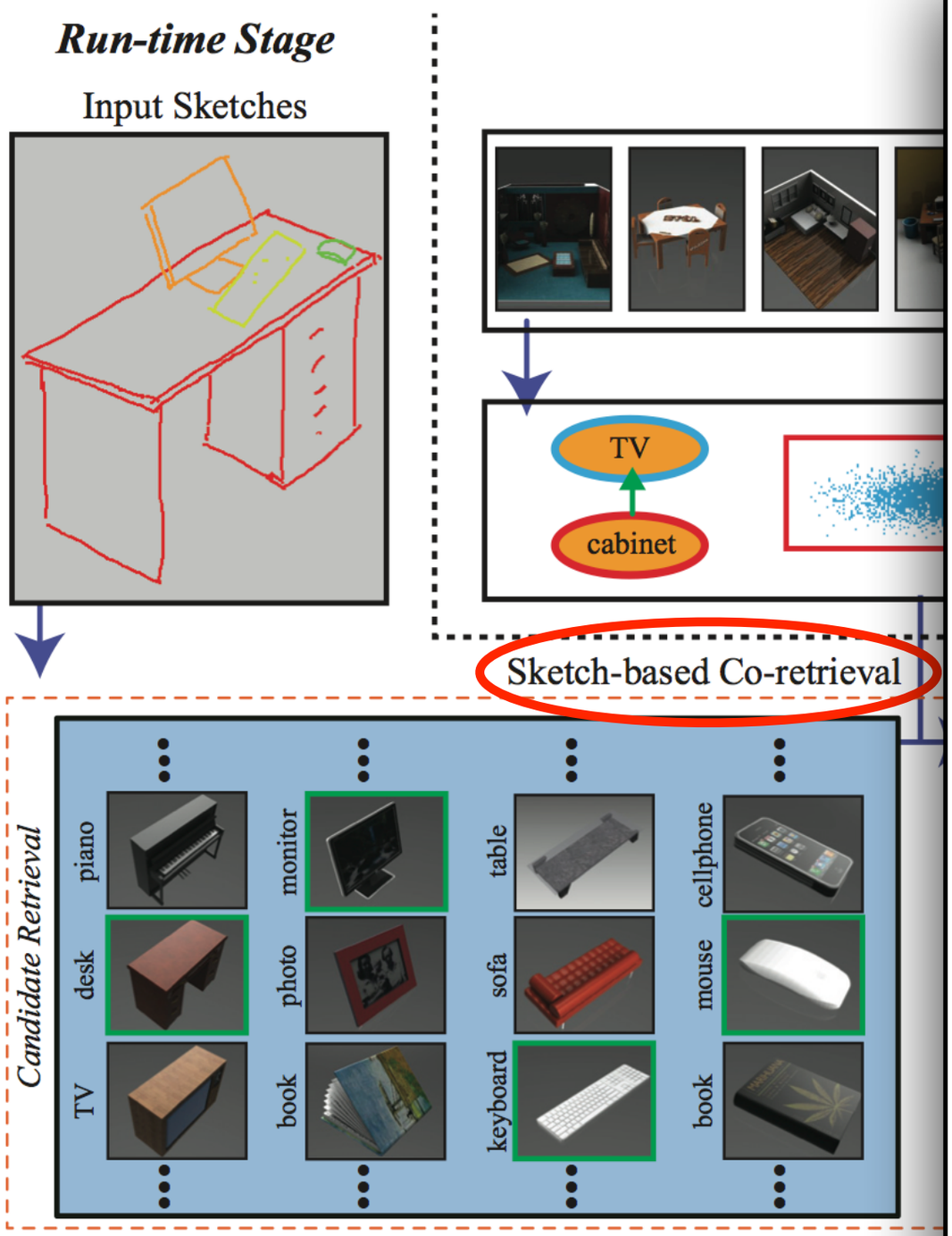
Scale consistency: similar scale for all obj

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

scale of each object to fit to sketch
DISCLAIMER: ignores perspective

Xu et al. 2013

Sketch-based Scene Synthesis



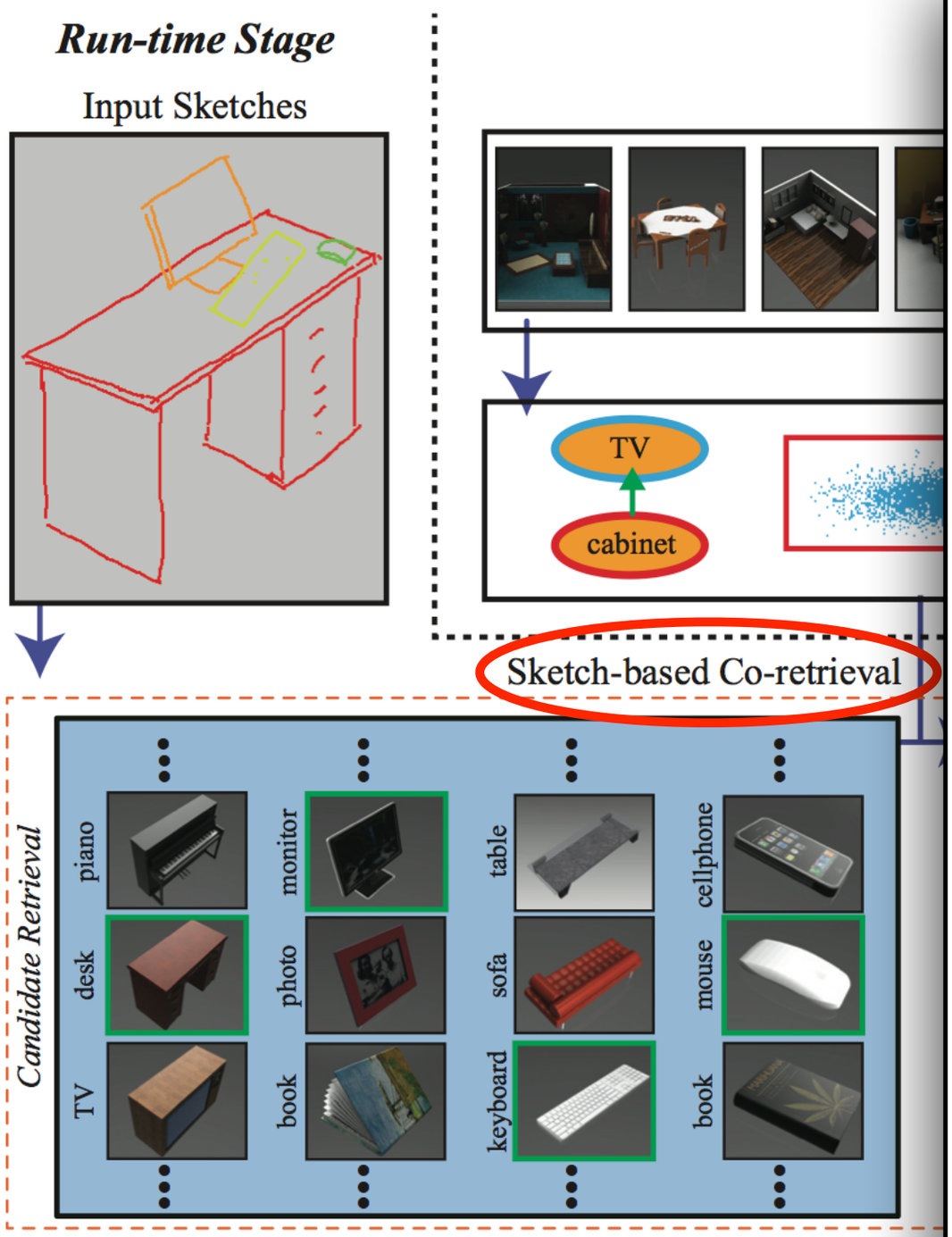
Retrieved set *C*

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

Check different subsets

Xu et al. 2013

Sketch-based Scene Synthesis



Retrieved set *C*

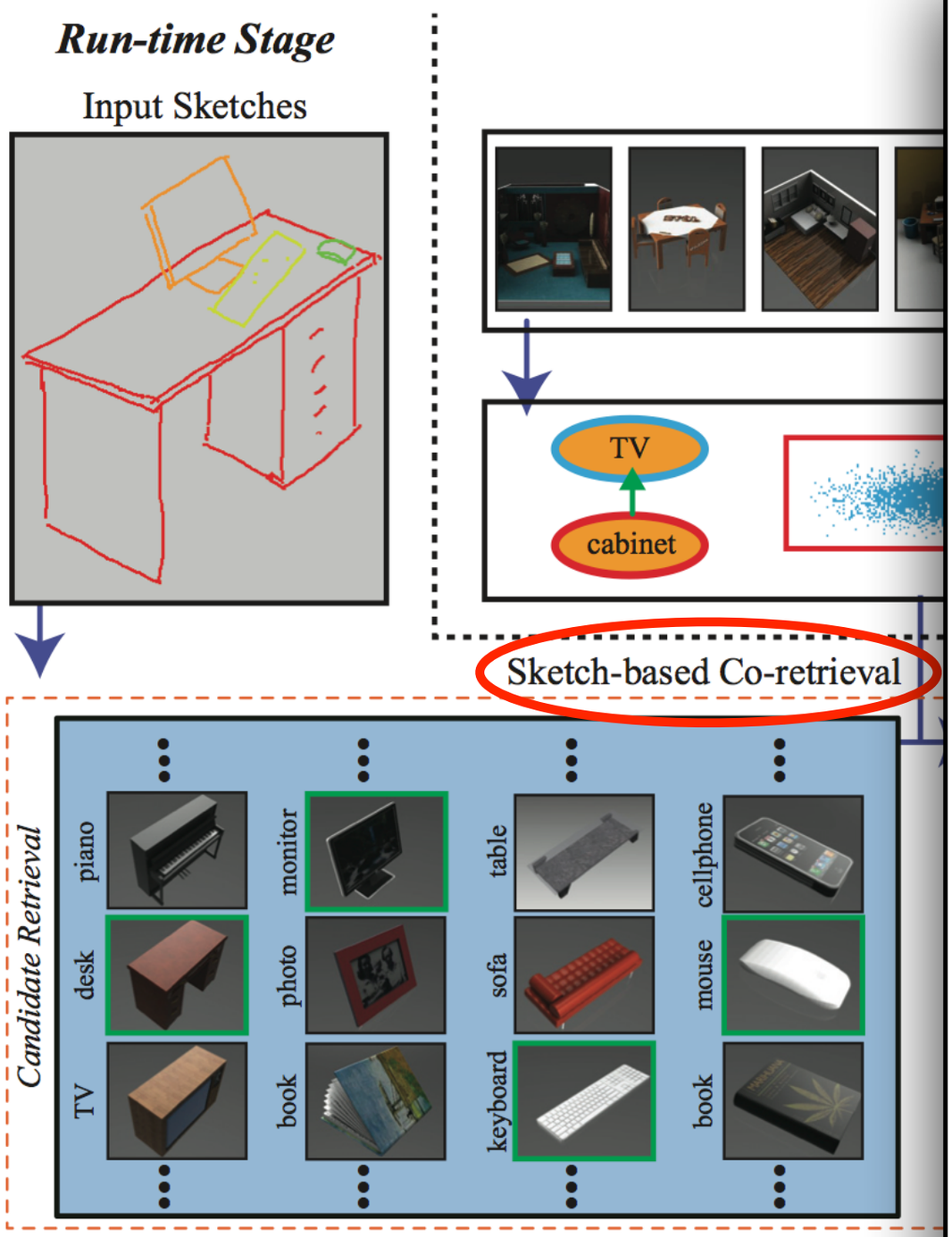
$$s(C) = P(C) \cdot \sum_{\forall U \subset C, 2 \leq |U| \leq 5} a(|U|) f(G(U)) \cdot \boxed{s_{\text{sub}}(U)}$$

Same context

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

Xu et al. 2013

Sketch-based Scene Synthesis



Retrieved set *C*

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

Same context

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

Sketch-to-object similarity

Spatial arrangement similarity

Xu et al. 2013

Sketch-based Scene Synthesis

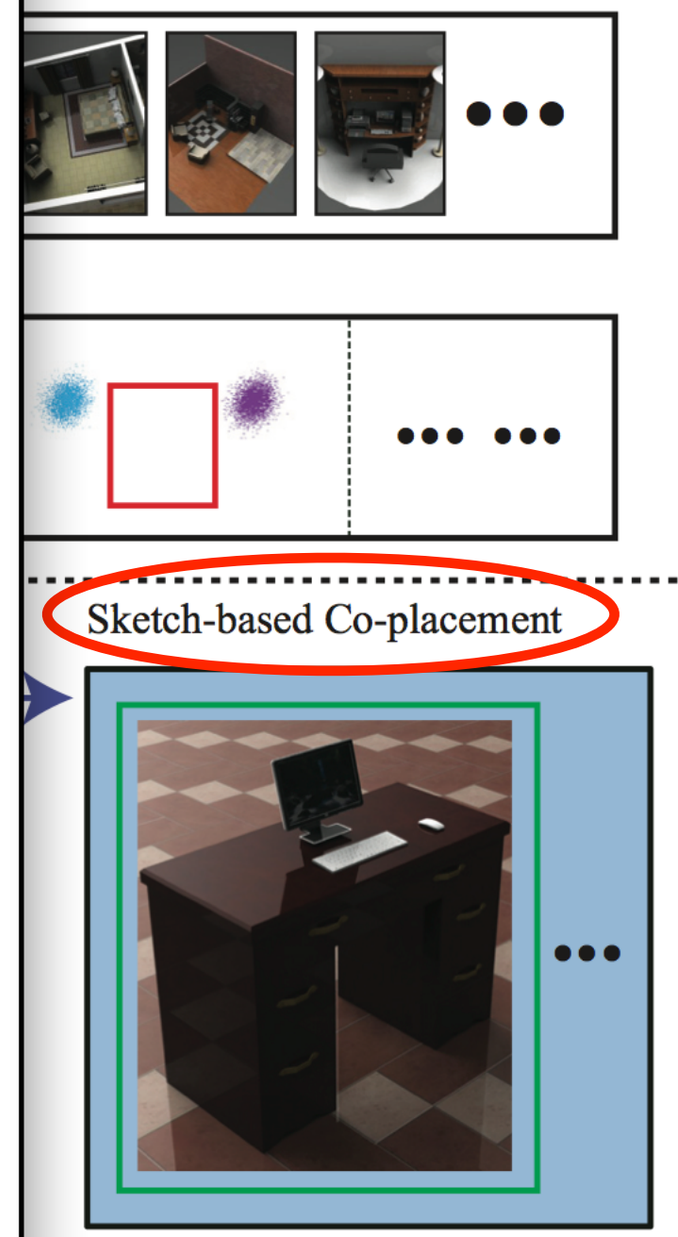
Place objects based on learned relationships

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

Sketch-to-object similarity

Spatial arrangement similarity

Old and new arrangement similarity



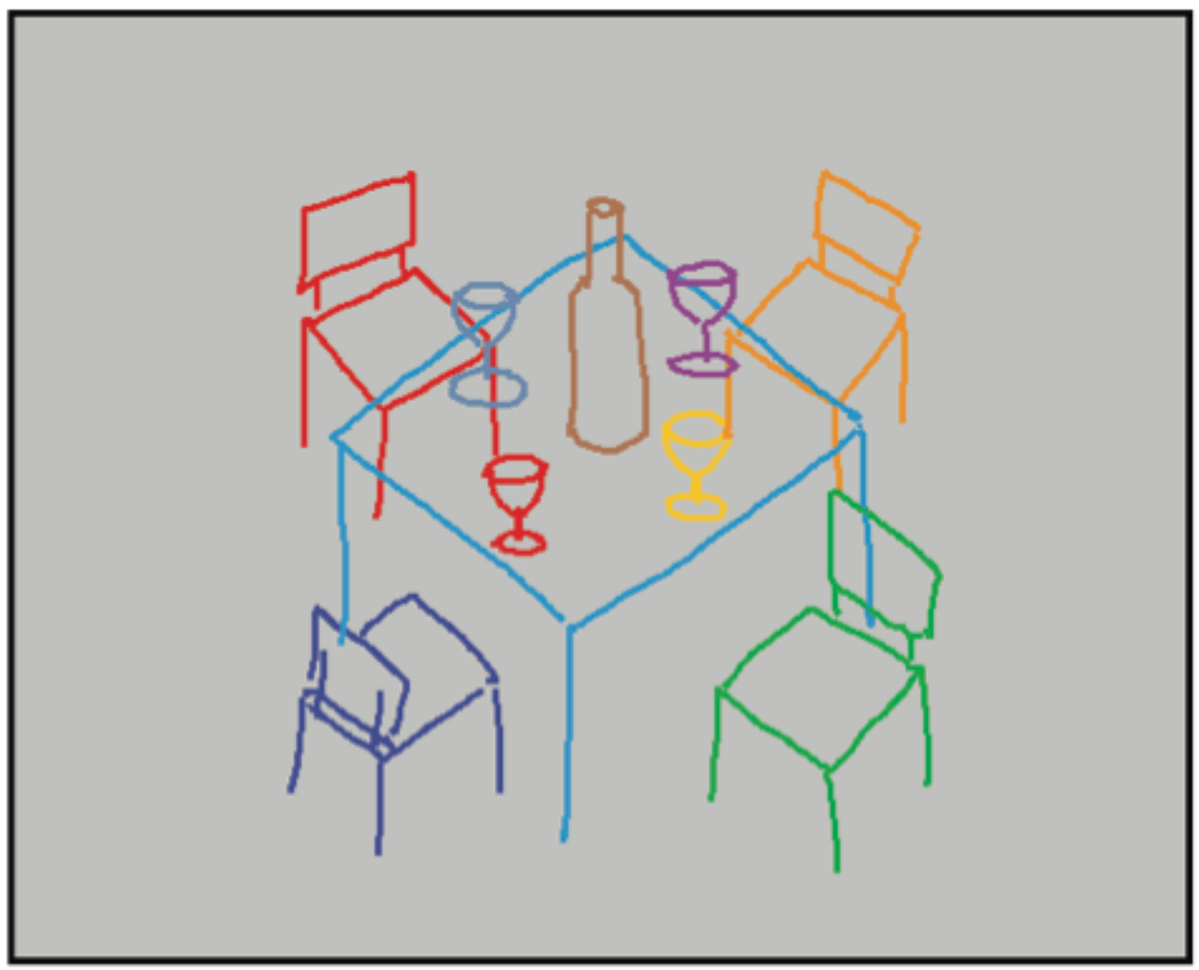
Xu et al. 2013

Sketch-based Scene Synthesis

Place objects based on learned relationships

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

$w_a = 1$



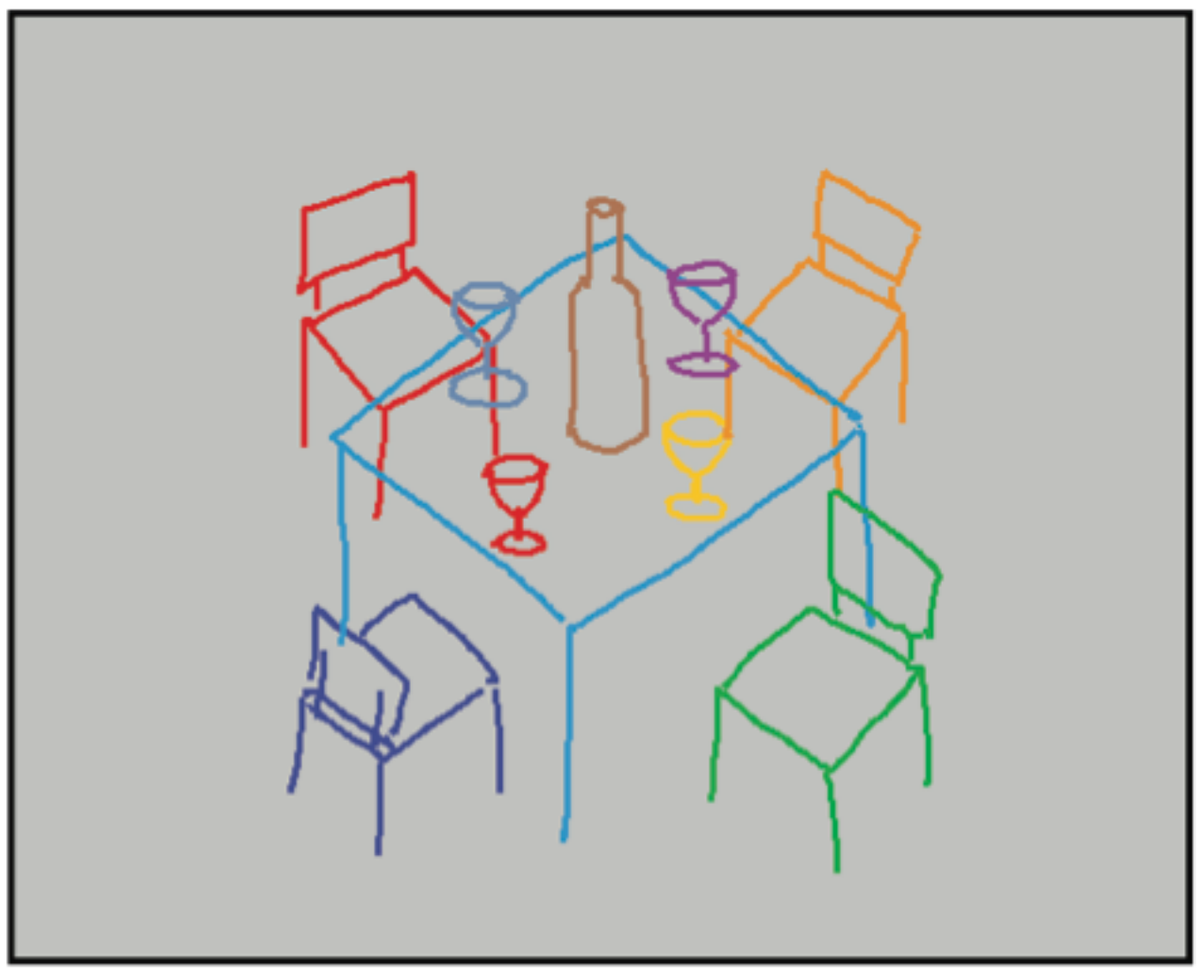
Xu et al. 2013

Sketch-based Scene Synthesis

Place objects based on learned relationships

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

$w_a = 0$



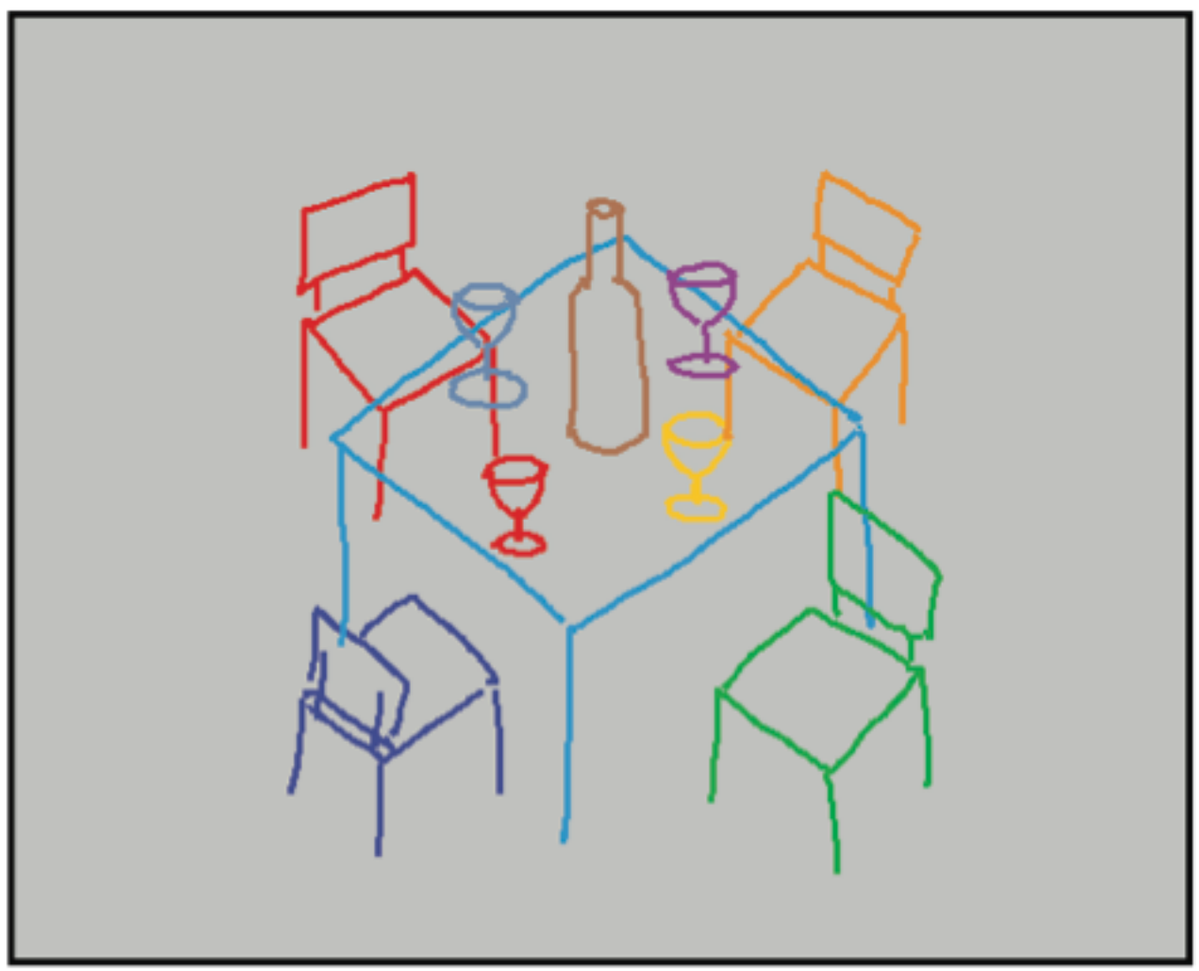
Xu et al. 2013

Sketch-based Scene Synthesis

Place objects based on learned relationships

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

$w_a = .1$



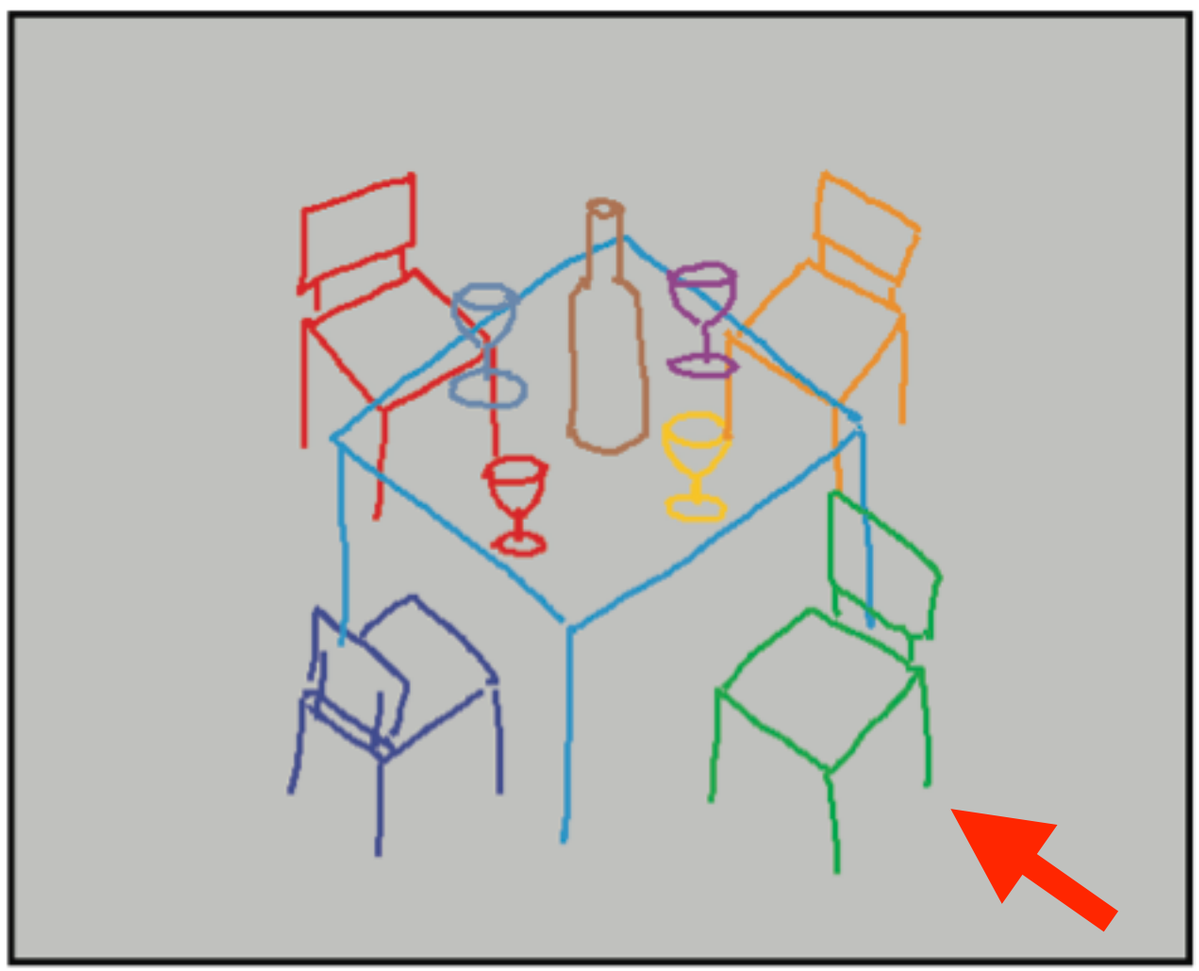
Xu et al. 2013

Sketch-based Scene Synthesis

Place objects based on learned relationships

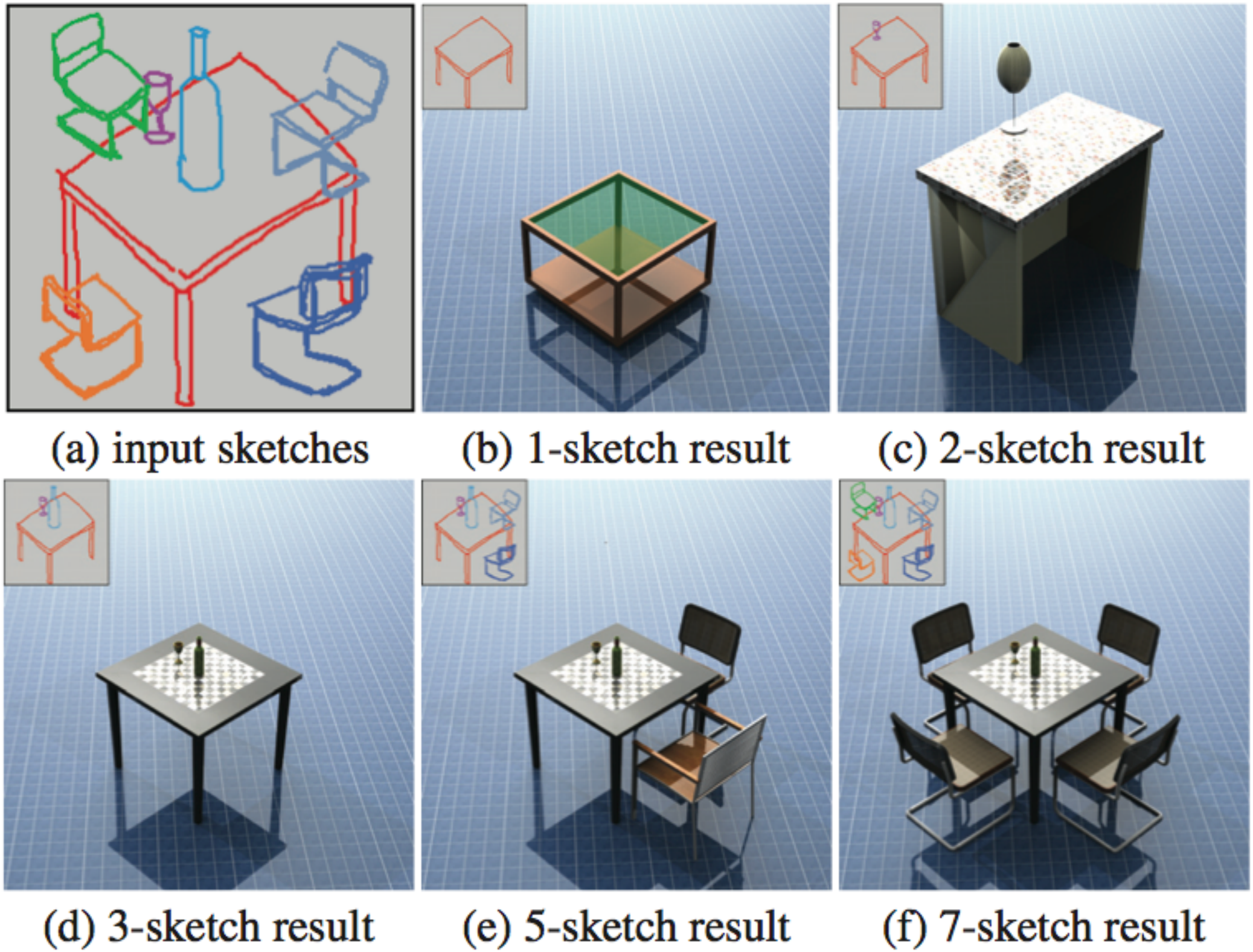
$$s_{\text{sub}}(U) = \prod_{1 \le i \le p} m_S(u_i) \cdot \prod_{1 \le i,j \le p} ((1 - w_a) g_{i,j}(u_i, u_j) + w_a t(u_i, u_j))^{\lambda_c}$$

$w_a = .1$



Xu et al. 2013

Value of context

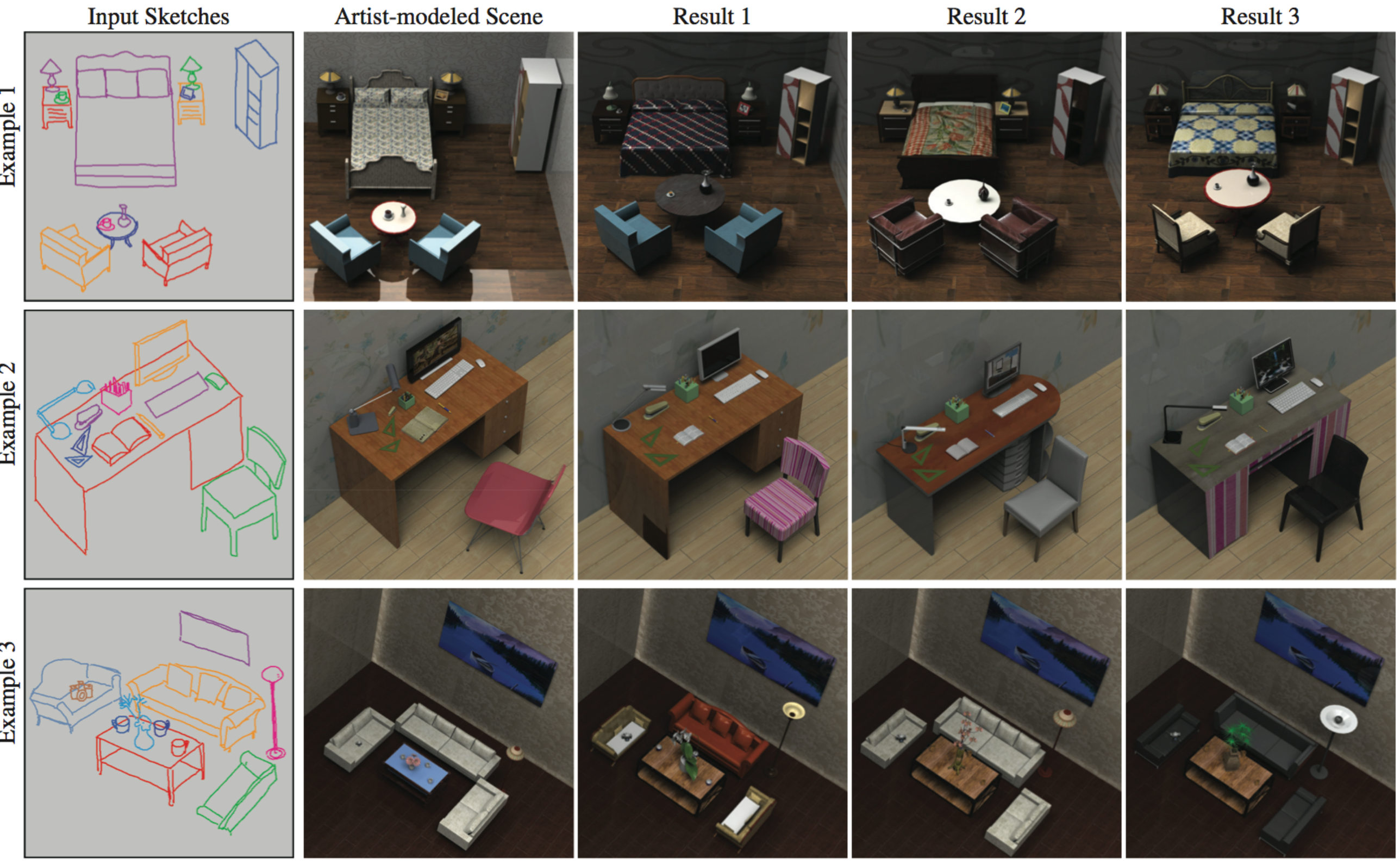


(a) input sketches (b) 1-sketch result (c) 2-sketch result

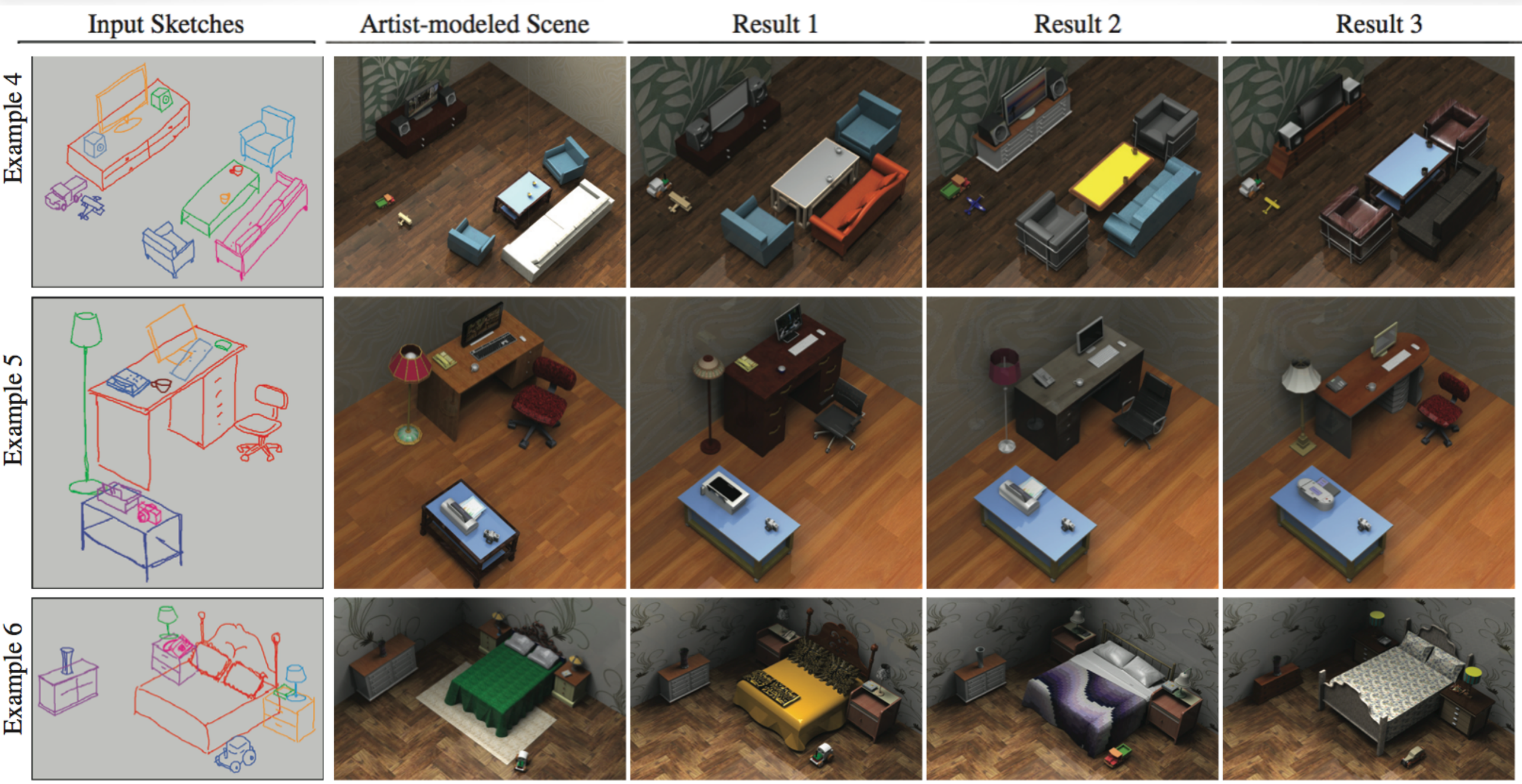
(d) 3-sketch result (e) 5-sketch result (f) 7-sketch result

Xu et al. 2013

Sketch2Scene results



Sketch2Scene results



Xu et al. 2013

Sketch2Scene Interaction

VIDEO

Xu et al. 2013

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
- Characterizing Structural Relationships in Scenes Using Graph Kernels.
  M. Fisher, M. Savva, P. Hanrahan. SIGGRAPH 2011
- Example-based Synthesis of 3D Object Arrangements.
  M. Fisher, D. Ritchie, M. Savva, T. Funkhouser, P. Hanrahan. SIGGRAPH Asia 2012
- Sketch2Scene: Sketch-based Co-retrieval and Co-placement of 3D Models
  K. Xu, K. Chen, H. Fu, W.-L. Sun, S.-M. Hu. SIGGRAPH 2013
- Interactive Furniture Layout Using Interior Design Guidelines.
  P. Merrell, E. Schkufza, Z. Li, M. Agrawala, V. Koltun. SIGGRAPH 2011
- Make it Home: Automatic Optimization of Furniture Arrangement.
  L.-F. Yu, S.-K. Yeung, C.-K. Tang, D. Terzopoulos, T. F. Chan, S. Osher.)