Automatic Scene Inference for 3D Object Compositing

Kevin Karsch[1] Kalyan Sunkavalli[2] Sunil Hadap[2] Nathan Carr[2]
Hailin Jin[2] Rafael Fonte[1] Michael Sittig[1] David Forsyth[1]


[1]University of Illinois

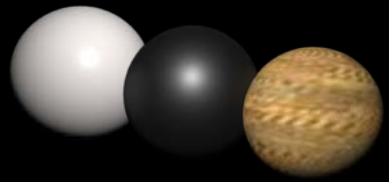

[2]Adobe Research

[Kee et al. ‘14]

Camera parameters

Geometry

Materials

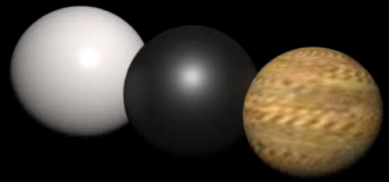


Light sources

Camera parameters

Geometry

Materials



Light sources

Single RGB image
(no depth or any other aids)

Automatic 3D scene inference

Camera parameters
(pinhole projection)

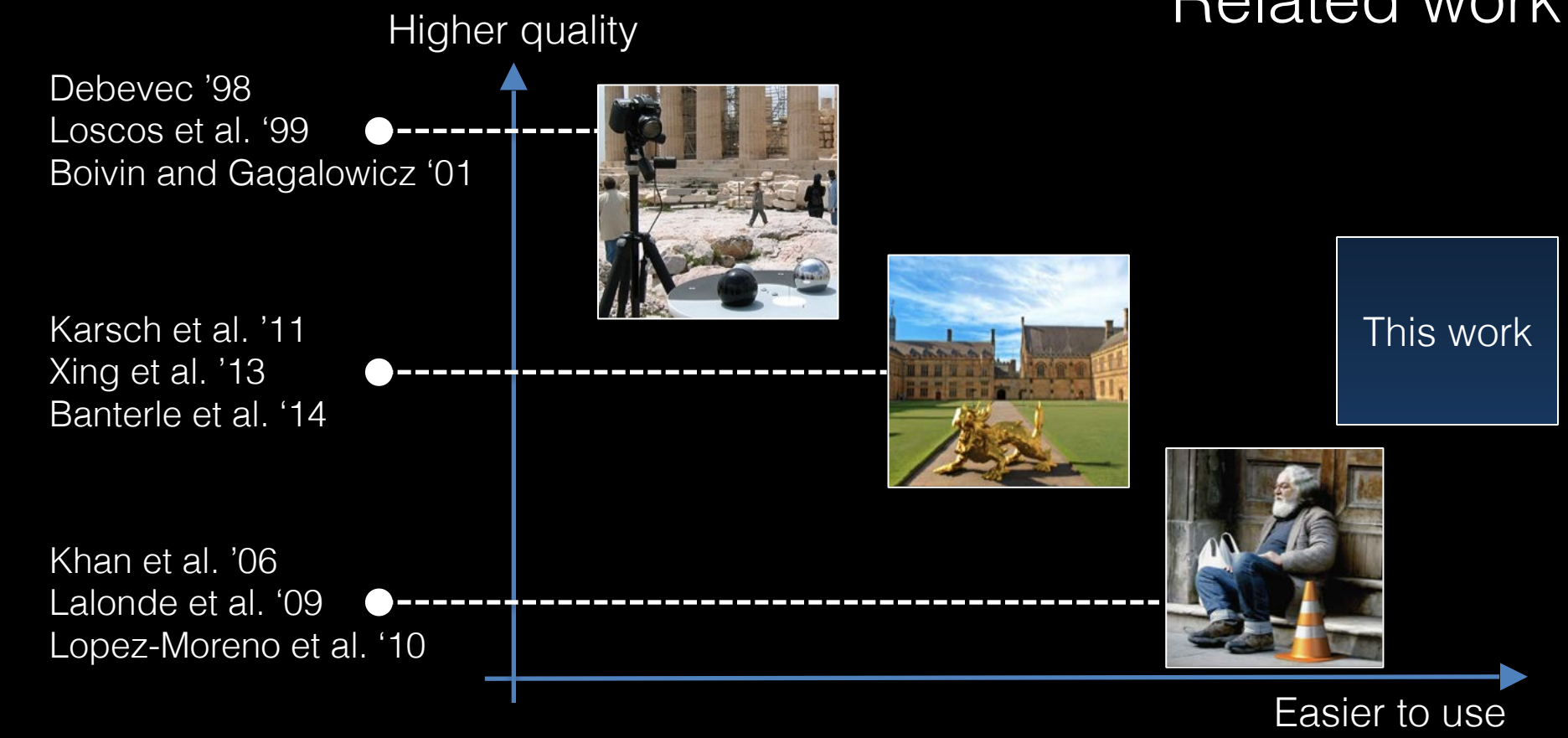
$$P = K[R, t]$$

Geometry
(dense depth map)

Materials
(diffuse reflectance)

Light sources
(IBL + area lights)

Related work



Overview

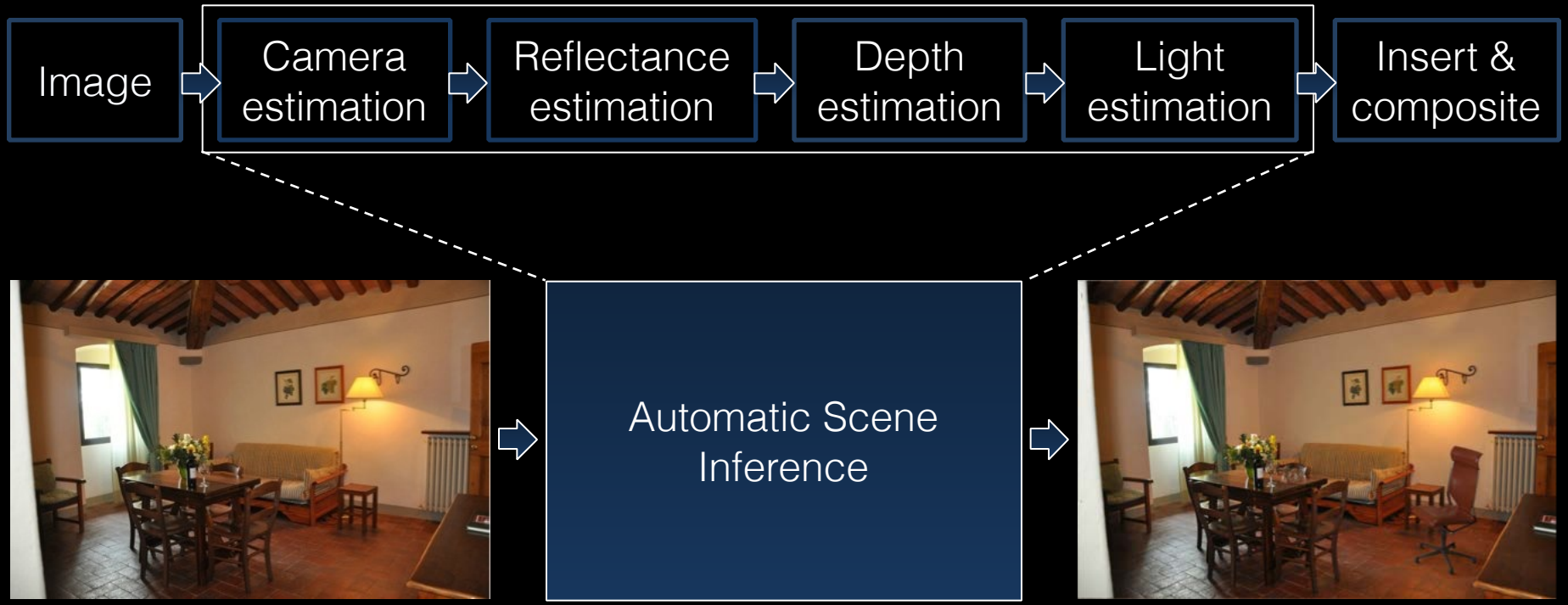


Image
Camera estimation
Reflectance estimation
Depth estimation
Light estimation
Insert & composite
Automatic Scene Inference
Input image
Insertion result

Overview

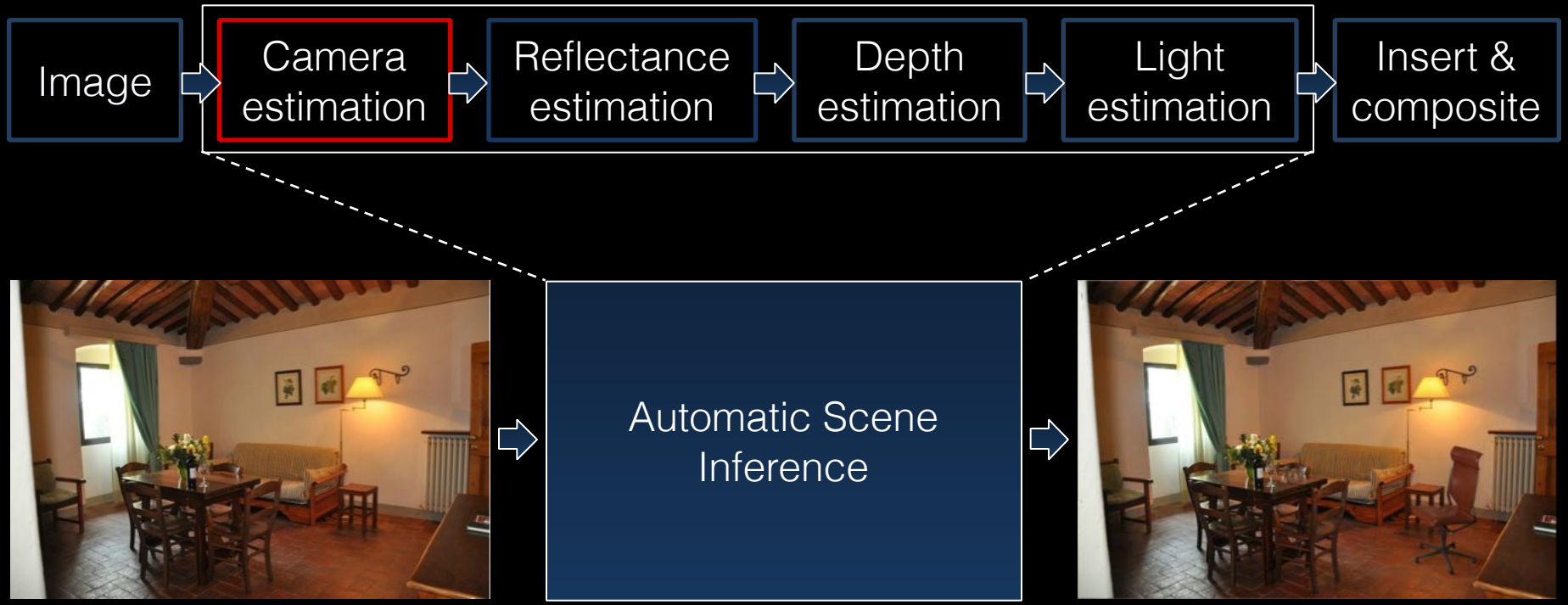
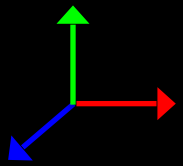


Image
Camera estimation
Reflectance estimation
Depth estimation
Light estimation
Insert & composite
Automatic Scene Inference
Input image
Insertion result

Camera estimation



Camera estimation

[Rother 2000]

Pinhole projection matrix

$$\begin{bmatrix} f & 0 & c_x \\ 0 & f & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} R, & \begin{matrix} 0 \\ 0 \\ 0 \end{matrix} \end{bmatrix}$$

intrinsics extrinsics

Overview

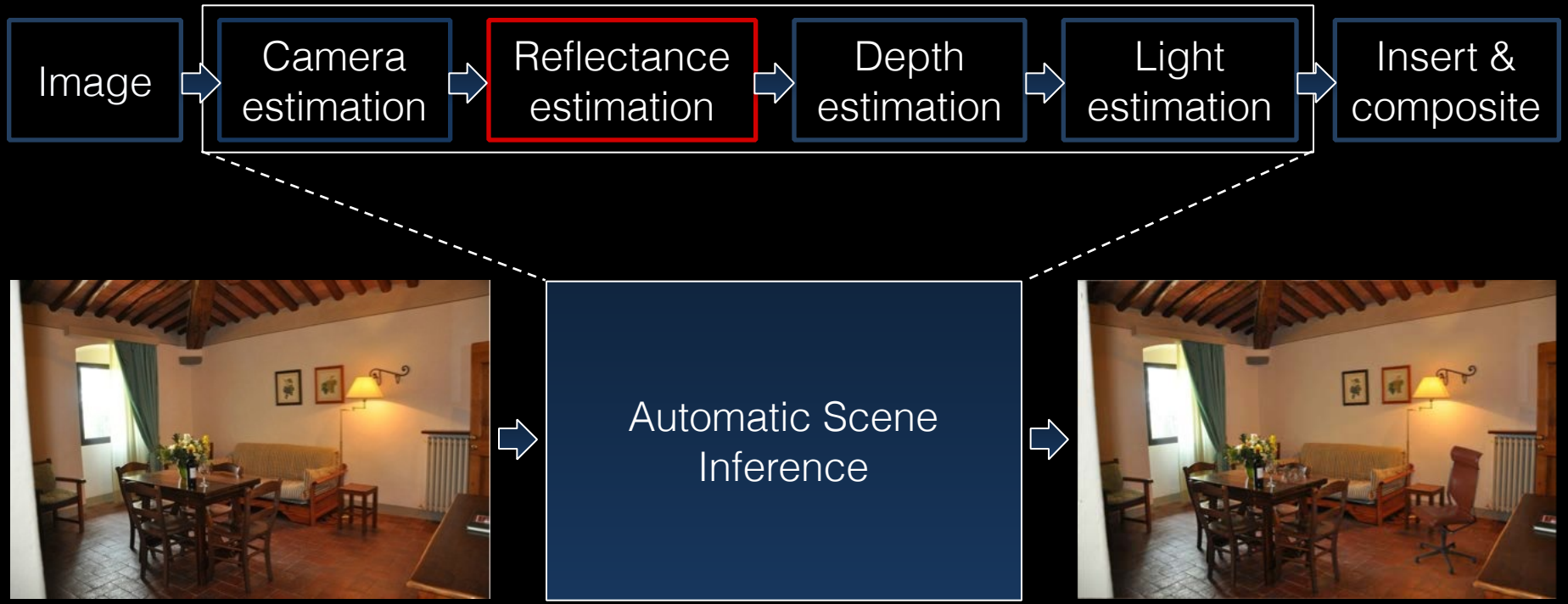


Image
Camera estimation
Reflectance estimation
Depth estimation
Light estimation
Insert & composite
Automatic Scene Inference
Input image
Insertion result

Diffuse material estimation

Diffuse material estimation

Diffuse reflectance

[Grosse et al. 2009; Land 1971]

Overview

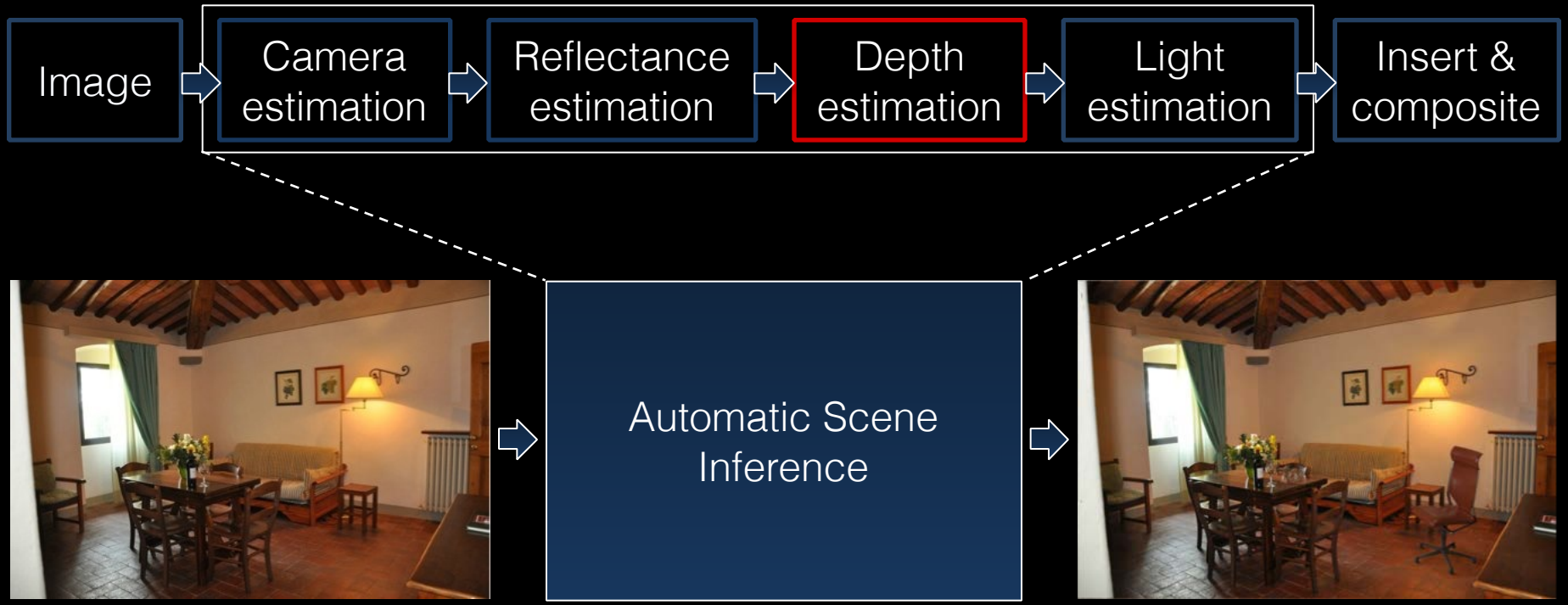
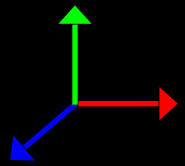


Image
Camera estimation
Reflectance estimation
Depth estimation
Light estimation
Insert & composite
Automatic Scene Inference
Input image
Insertion result

Depth estimation

Depth Transfer
[Karsch et al. 2012]



Depth estimation

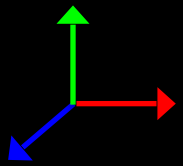
Orientation maps
[Lee et al. 2009]

Our result
(Flat, smooth surfaces better for insertion)

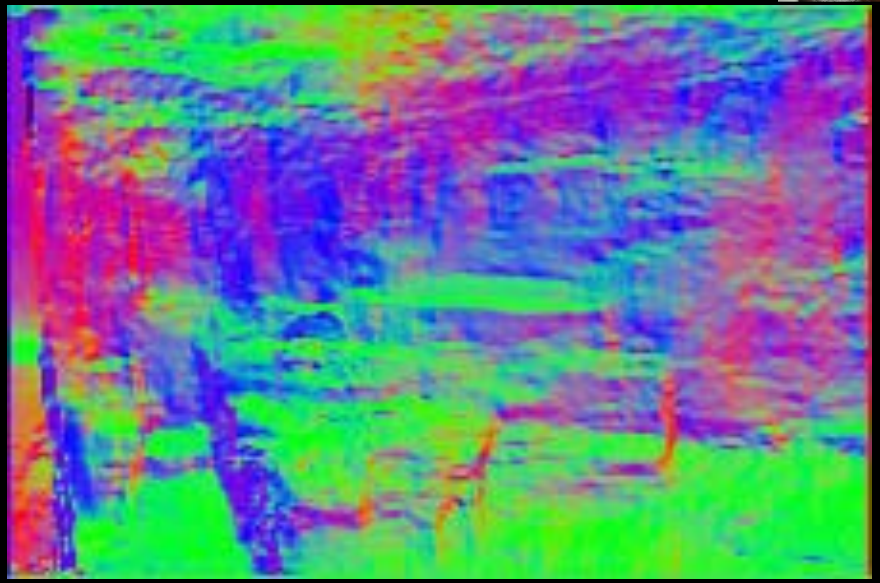
Depth estimation

Depth Transfer
[Karsch et al. 2012]

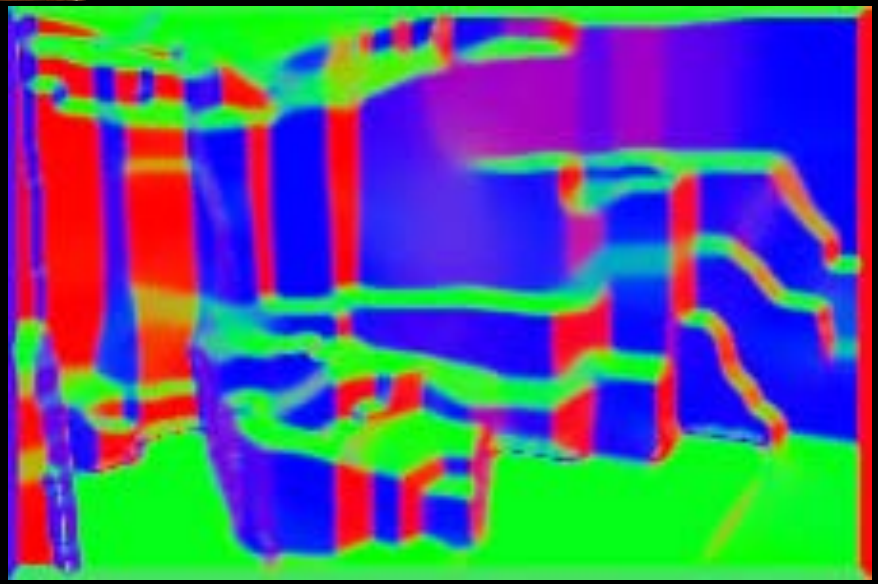
Our result
(Flat, smooth surfaces better for insertion)



Depth estimation



Depth Transfer
[Karsch et al. 2012]



Our result
(Flat, smooth surfaces better for insertion)

Overview

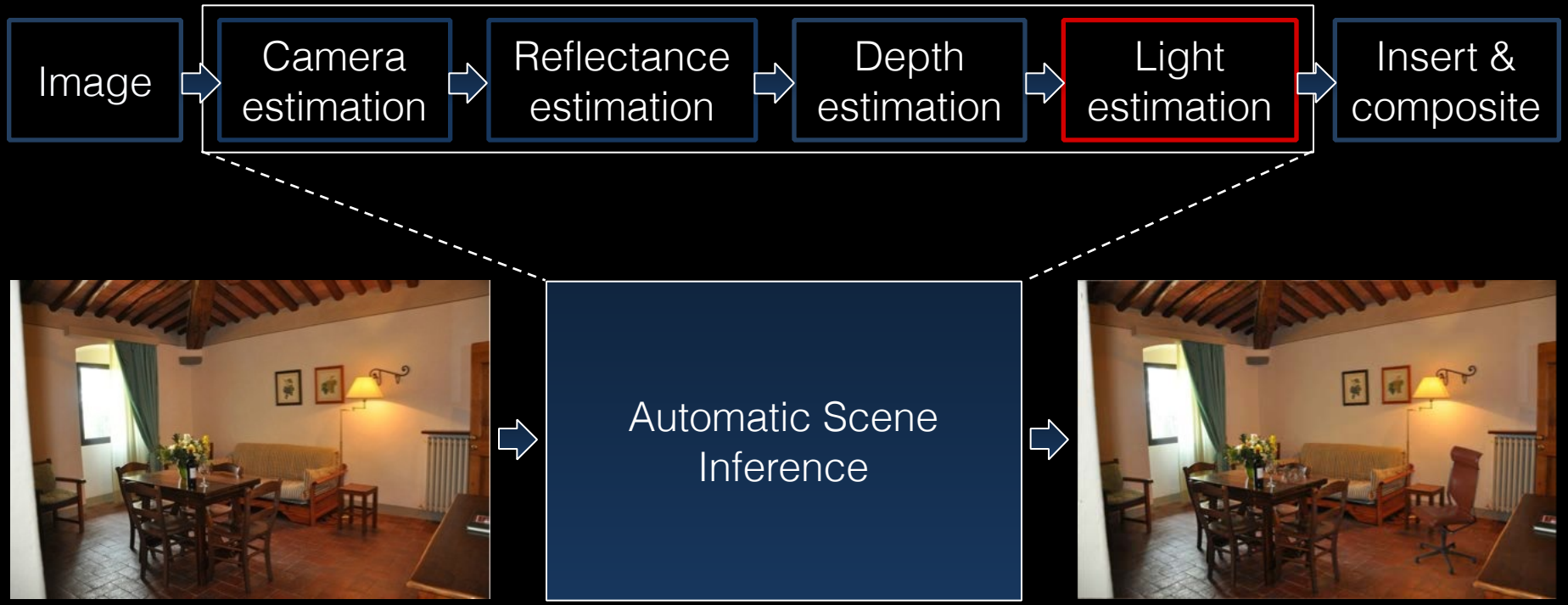


Image
Camera estimation
Reflectance estimation
Depth estimation
Light estimation
Insert & composite
Automatic Scene Inference
Input image
Insertion result

Light estimation overview

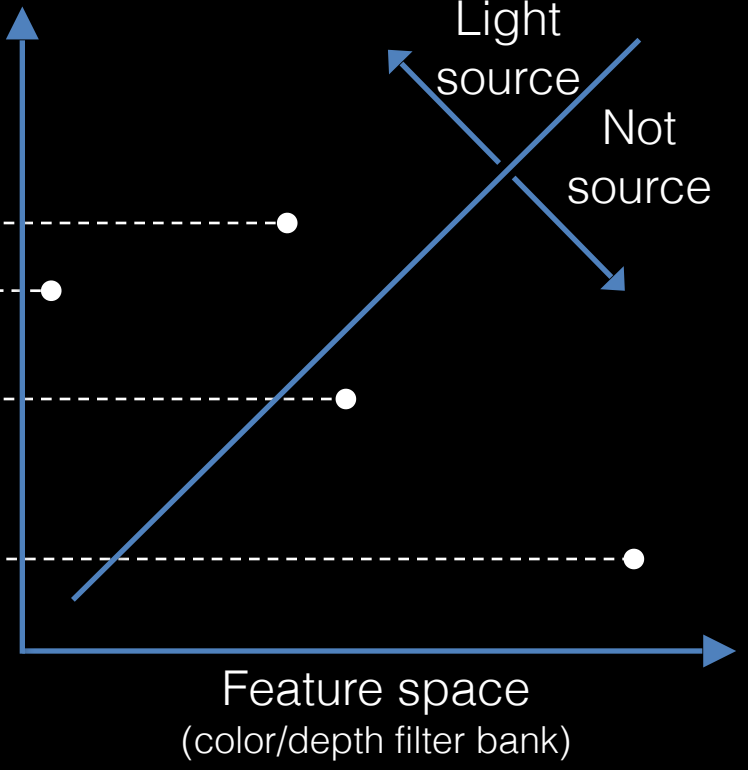
1. “In-view” lighting
2. “Out-of-view” lighting

Light estimation overview

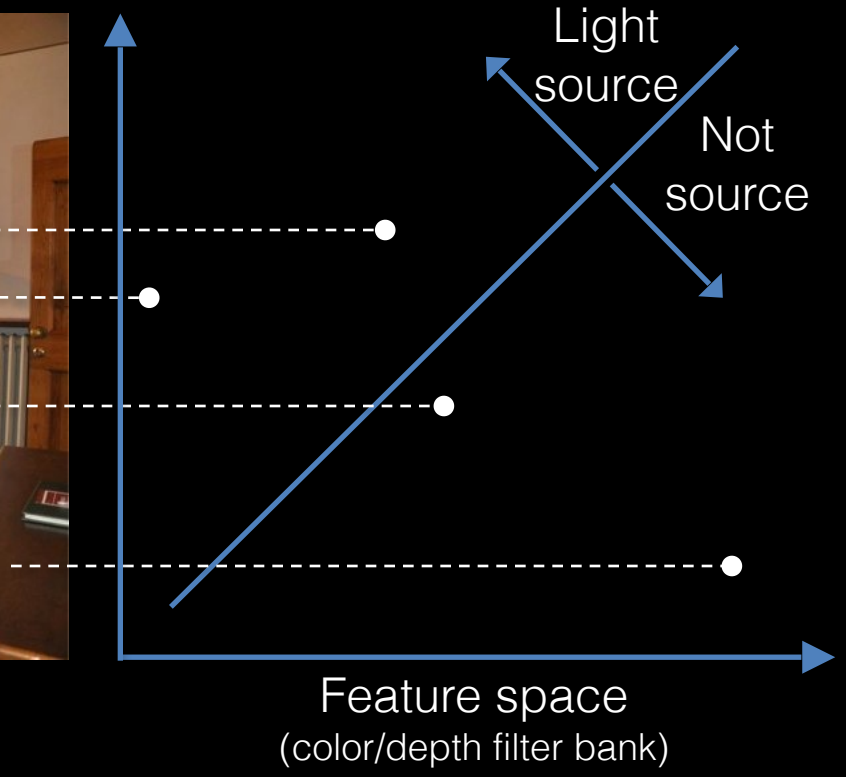
1. “In-view” lighting
2. “Out-of-view” lighting
3. Intensity optimization

Estimating “in-view” sources

Estimating “in-view” sources



Estimating “in-view” sources



Estimating “out-of-view” sources

Assumption: if two pictures look similar,
their lighting environments are also similar

Image 1  ≈  Image 2

Environment 1  ≈  Environment 2

Estimating “out-of-view” sources

SUN360 dataset [Xiao et al. 2012]

Estimating “out-of-view” sources

…

Estimating “out-of-view” sources

0.13

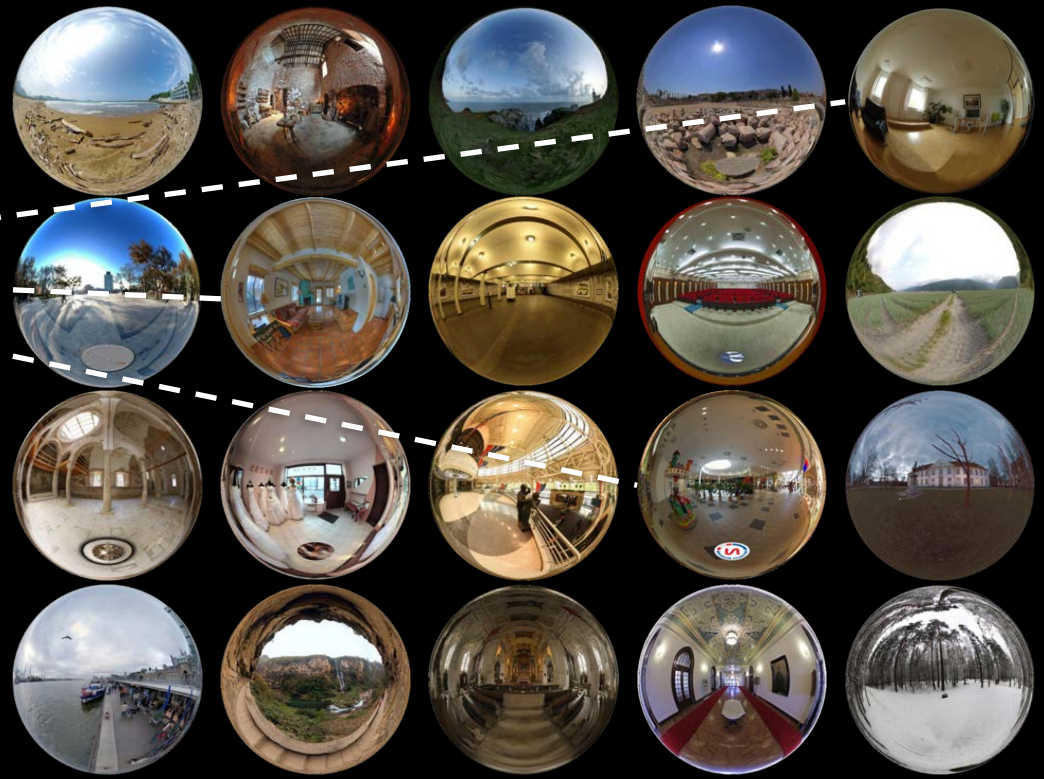
0.08

0.21

…

0.92

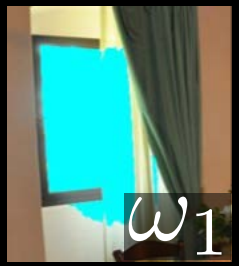
Estimating “out-of-view” sources



Intensity optimization

Input image

Initial rendered scene



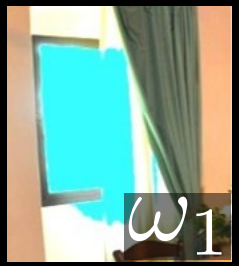
In-view sources

Out-of-view sources

Intensity optimization

Input image

Optimized rendered scene



In-view sources

Out-of-view sources

Intensity optimization

Input image

Optimized rendered scene

$$\underset{\omega,\gamma}{\operatorname{argmin}} \; ||I - R(\omega)^\gamma|| + P(\omega) + Q(\gamma)$$

Rendered image should look like input

Intensities snap to zero; not too bright

Gamma should stay near 1/2.2

Overview

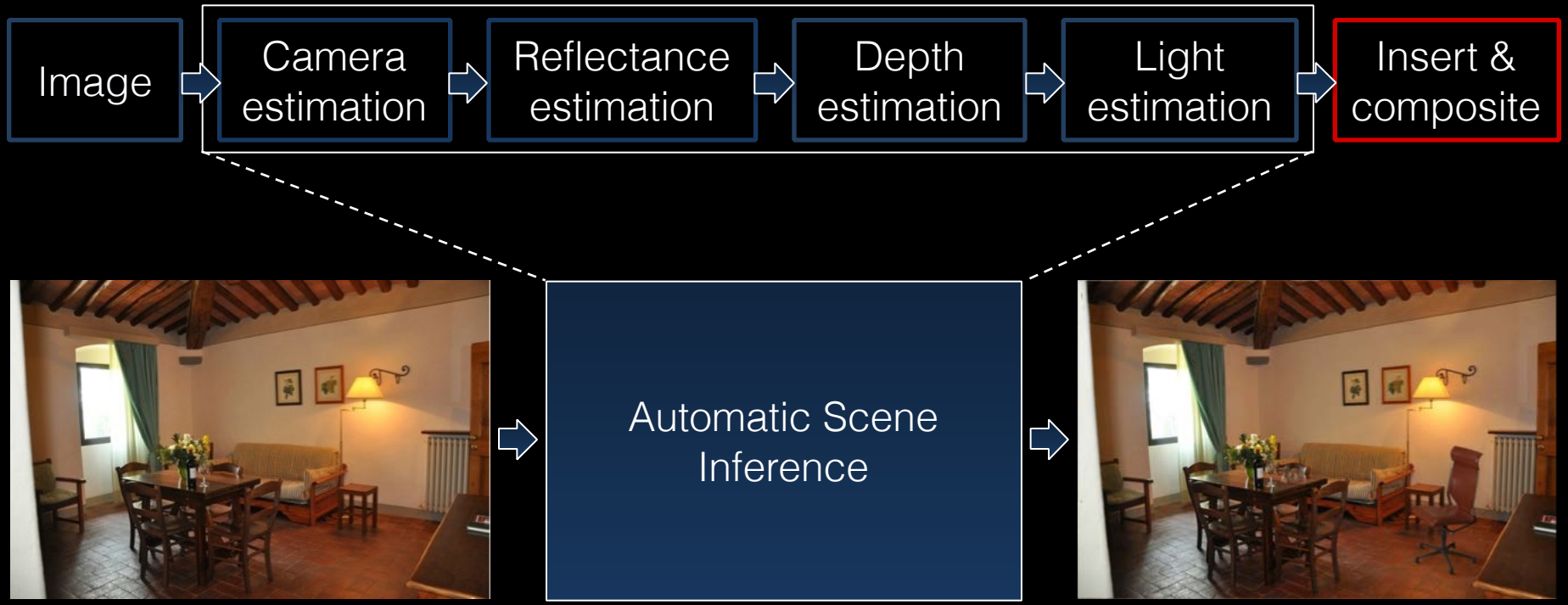
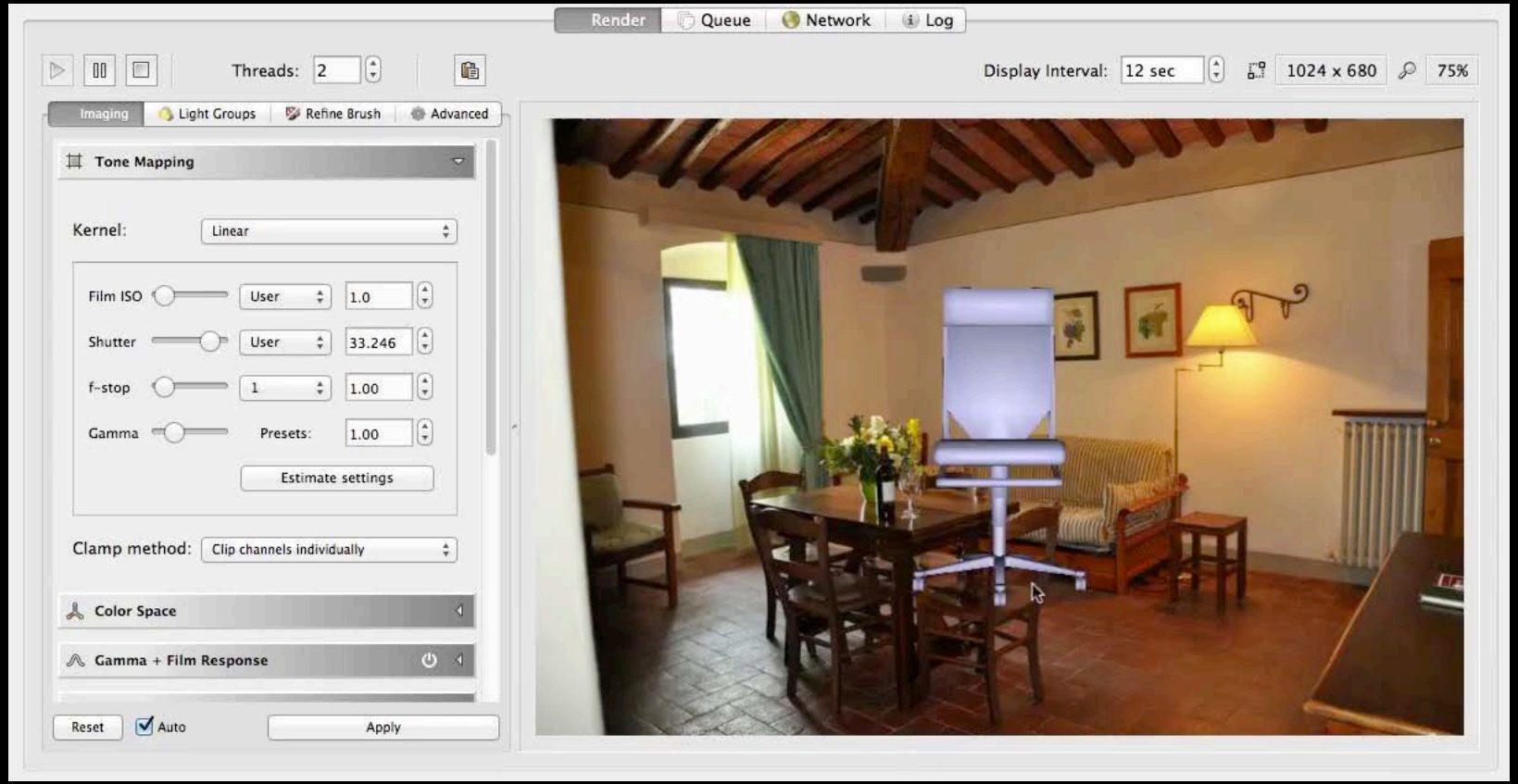


Image
Camera estimation
Reflectance estimation
Depth estimation
Light estimation
Insert & composite
Automatic Scene Inference
Input image
Insertion result

3D object insertion



User study

2AFC test: “Which image looks more realistic?”

Synthetic (yet realistic) scenes

- Ground truth/unedited: insert object + render
- Our results: auto-estimate scene, insert + composite

Ground truth

Our result

Our result

Ground truth

Study results

- Subjects chose our result over ground truth in 35.5% of trials (N=1776)
  - *[Karsch et al. ’11]: 34.0% (*similar study, different subjects/images)

- Conditions:
  - Experts: 33.3% | non-experts: 36.1%
  - Failed perspective/shading test: 48.2%
  - Indoor scenes: 37.7% | Outdoor scenes: 28.8%
  - Many more in paper/supplemental

Our result

Ground truth

Summary

- First fully automatic scene (depth+lighting) inference technique
- Produces rough scene estimates suitable for 3D object insertion
- Results in human confusion rates of >35%
- Future work: improved outdoor lighting, specular materials