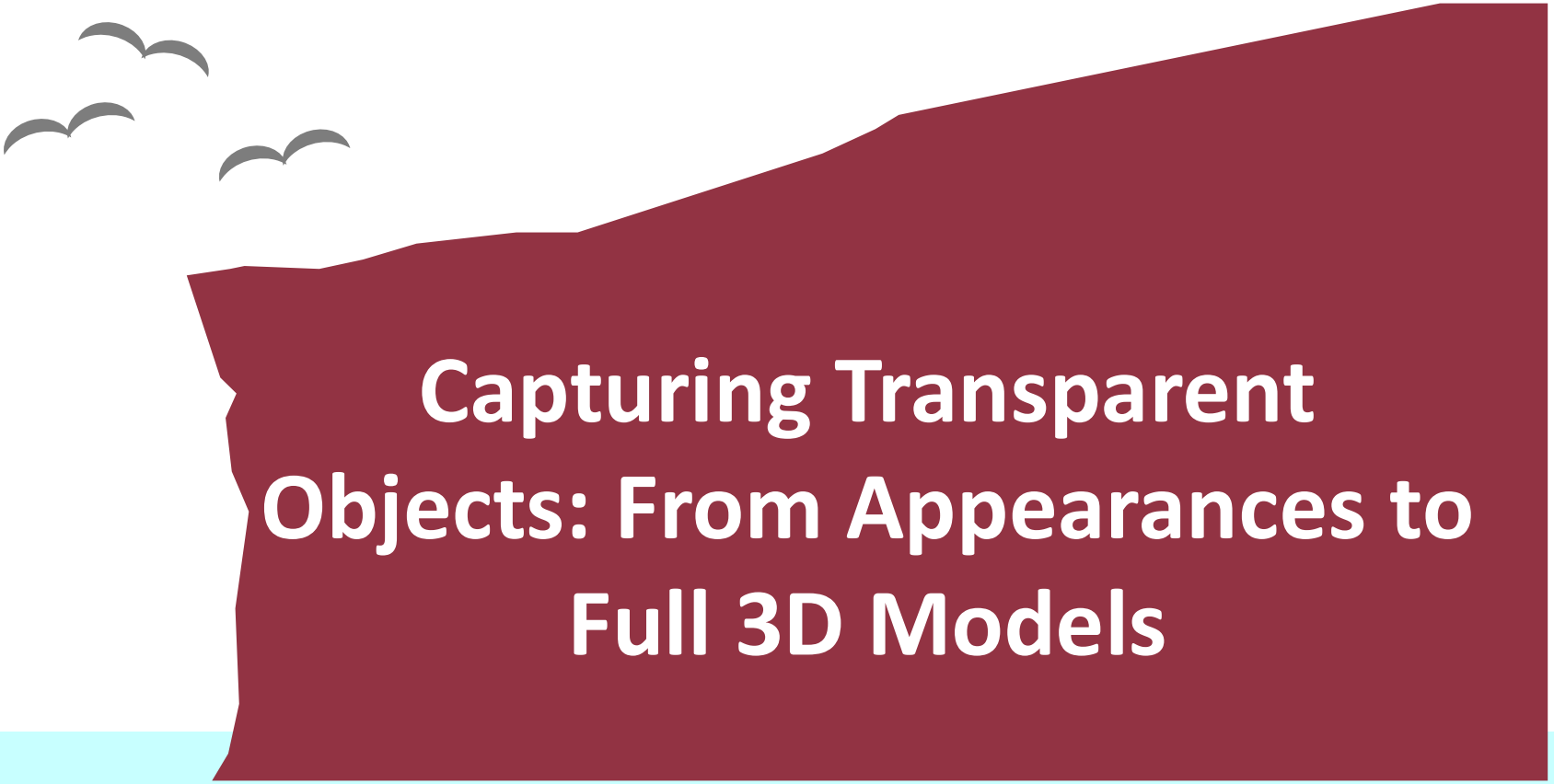


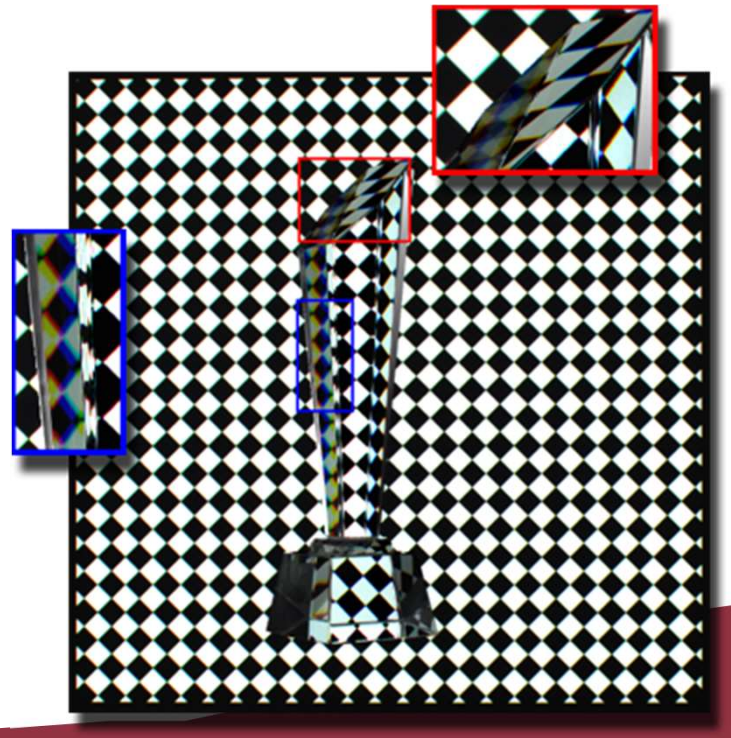
Capturing Transparent Objects: From Appearances to Full 3D Models

Minglun Gong

Dept. of CS, Memorial Univ.

- **Capture transparent object appearance**
  - **Use environmental matting & reduce # of images needed through compressive sensing**
- **Reconstruct transparent surface shape**
  - **Measure how the light refracted & optimize 3D surface positions & normal**
- **Reconstruct full transparent object model**
  - **Consolidate a point set surface to optimize light refraction, silhouette, & smoothness constraints**

Three Research Projects



Part I

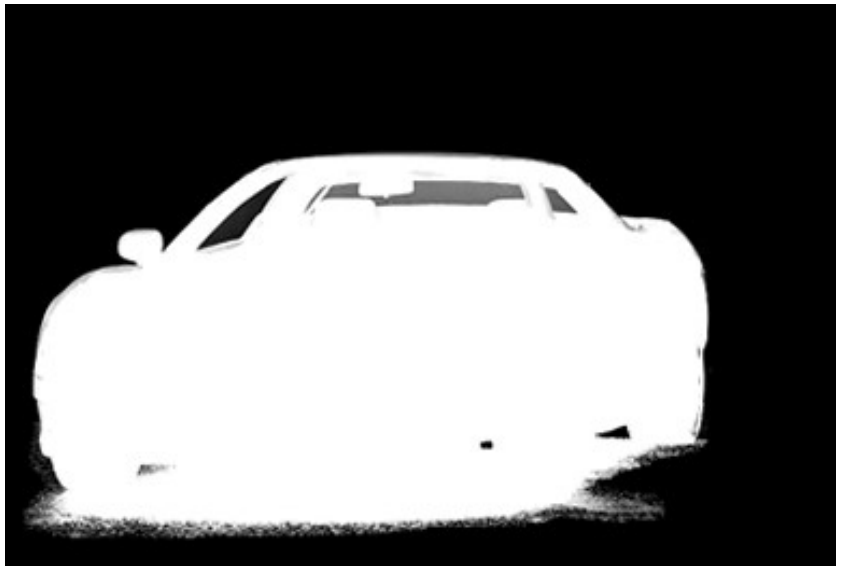
Capture Transparent Object Appearance

Extract Objects from Images

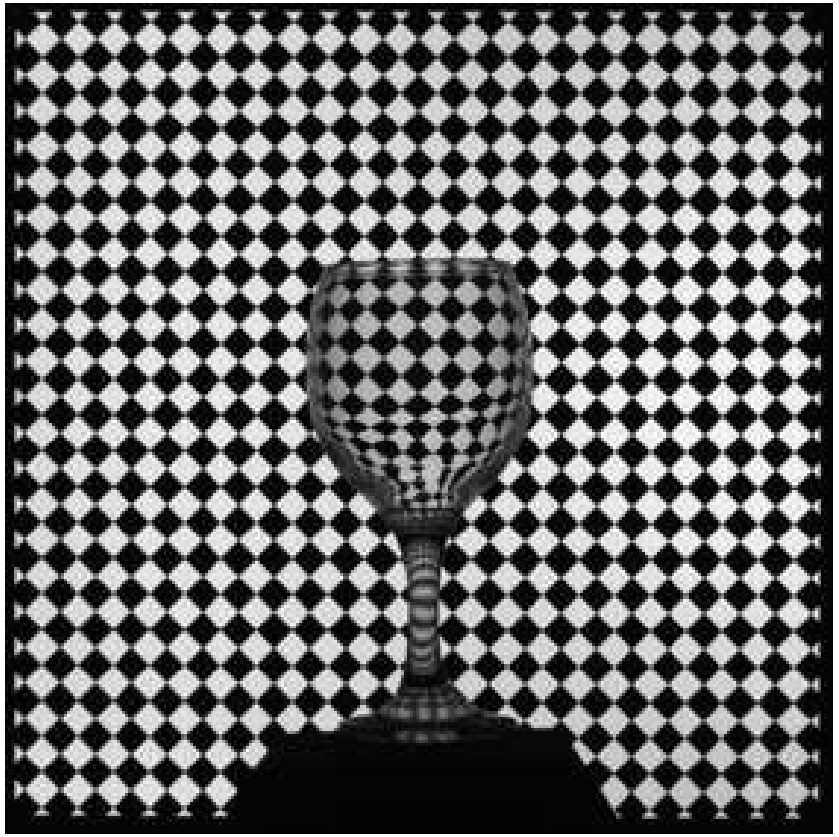
- How to extract an object then insert it to a new scene?
  - Often referred as object cutout
  - Simply using binary mask introduce aliasing artifacts
  - Image matting is used to extract the transparent parts & fuzzy object boundaries

Image Matting Formulation

- $C = \alpha F + (1 - \alpha)B$
  - C: **the observed intensity of an pixel**
  - α : **the percentage of the pixel covered by the foreground**
  - F: **foreground color**
  - B: **background color**
- **Assume that light does not change directions**



- Do not have their own colors but acquire their appearances from the environments
  - Reflect, refract, & scatter environment light
- Require environment matting (EM)



Transparent Objects

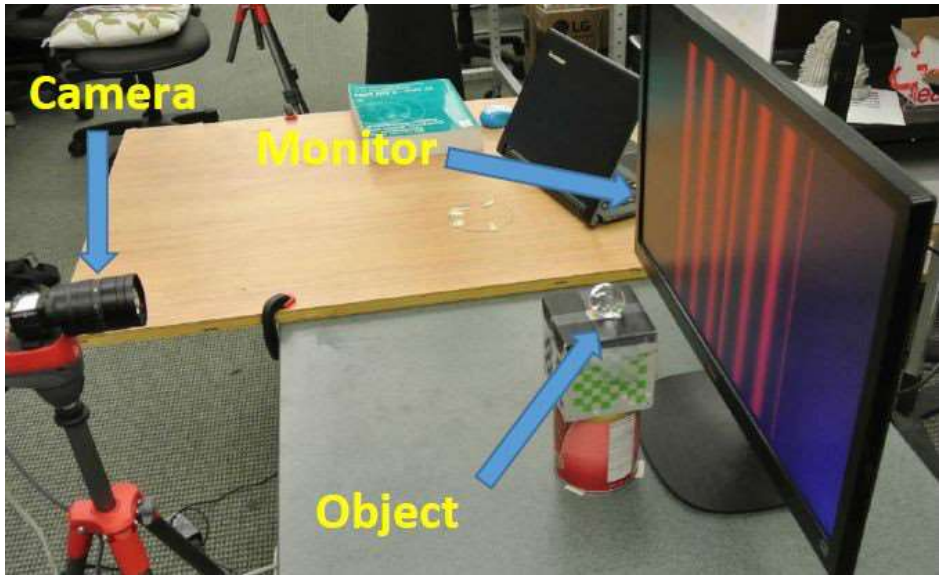
- $C = F + \rho \mathbf{WB}$
  - F: ambient illumination
  - ρ: light attenuation index
  - $\mathbf{B}$: $n^2 \times 1$ background image vector
  - $\mathbf{W}$: $1 \times n^2$ light transport vector
    - Describes the amount of contribution from background
    - $\|\mathbf{W}\|_1 = 1, \mathbf{W}_i \geq 0$

- Under a solid black backdrop:
  - $\mathbf{B} = 0$,
  - $C = F$
- Under a solid white backdrop:
  - $\mathbf{B} = b$,
  - $\|\mathbf{W}\|_1 = 1$,
  - $C = F + \rho b$

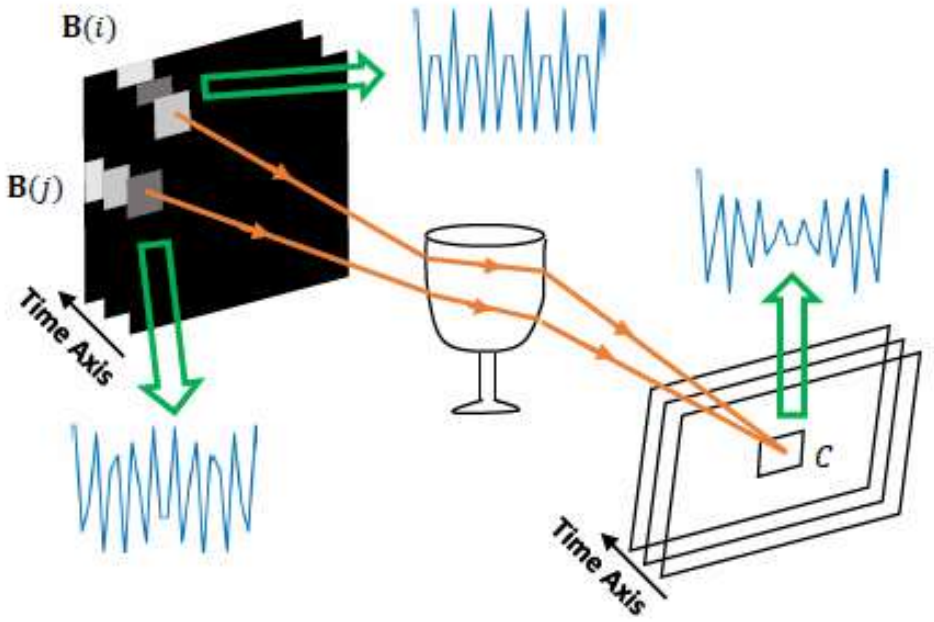
Environment Matting Formulation

Environment Matting (Cont'd)

- **The major task of EM is to recover the light transport vector W**
  - **Many-to-one decomposition**
  - **Need to photograph the object in front of a series of predesigned backdrops**



- **Proposed by Zhu & Yang in PG 2004**
  - **Let each background pixel emit an unique frequency signal**
  - **Find the accurate contributing sources by analyze the frequency of observed pixel**



Frequency-based EM Approach

Data Capture in Frequency-based EM

- Number of images needs to be captured depends on the resolution of the backdrop
  - In theory, $2 \times n^2$ images are needed for backdrop with $n \times n$ pixels
  - Over a million of images for backdrop with 1K resolution
- Assume $\mathbf{W}$ is the element-wise product of a row vector & a column vector
  - Display row-based patterns then column-based patterns
  - Use row/column number to determine contribution source
  - Number of images needed drops to $4 \times n$

- **A framework for reconstructing sparse signals**
  - **A N-dimensional signal x is called an s-sparse signal if x contains at most $s \ll N$ nonzero elements**
- **Uses $M < N$ linear measurements $y = Ax$ for reconstruction**
  - **x can be stably recovered by solving the following problem:**
    - $\min \|x\|_1 , \quad \text{s.t.}\, y = Ax$
    - **with only $M = O\left(s \times log(N/s)\right)$ measurements**

Compressive Sensing

- **Introduce CS to frequency-based EM:**
  - **A foreground pixel is contributed by only a sparse number of background pixels**
  - **The DFT of the recorded signal of an object pixel contains only a small number of frequencies**
- **Augment with phase information to distinguish signal with the same frequency**
  - **Further reduce the measurement cost**
  - **Accelerate the signal reconstruction process in CS**

Our Approach

- **Given the recorded signal C & the computed ambient illumination F:**
  - $C - F = \mathbf{DX}$
    - **D: the inverse of the $N \times N$ discrete Fourier transform matrix**
    - **X: an N-dimensional sparse complex vector representing the frequency information**

- **Randomly generate a M-dim permutation Ω from $\{0, 1, \cdots, N-1\}$ & display frequency patterns with frame ids from Ω**
  - $\min\|\mathbf{X}\|_1$, **s.t.** $C - F = \mathbf{D}(\Omega, :)\mathbf{X}$
  - $\mathbf{W}\big(\mathrm{ind}(r, c)\big) = \bar{\mathbf{W}}_{\mathrm{row}}(\mathrm{r})\bar{\mathbf{W}}_{\mathrm{col}}(\mathrm{c})$

Reconstruction via CS

Augment with Phase Information

- **Use both frequency & phase to encode source location**
  - $B(f, t, \varphi_p) = \xi\left(\cos\left(2\pi f \frac{t}{N} + \varphi_p\right) + 1\right)$
    - φ_p: **per-designed phase value for the pth region**
    - $1 \le p \le k$
- **Reduce the maximal frequency from n to $\frac{n}{k}$**

$1 \le f \le \frac{n}{k},\ \varphi_1$
$1 \le f \le \frac{n}{k},\ \varphi_2$
... ...
$1 \le f \le \frac{n}{k},\ \varphi_k$

Augment with Phase Information (Cont'd)

- **Both frequency search & phase search are now needed to determine the contributing sources**
  - **By displaying row-based patterns with phase info, we use CS to obtain the contributing frequencies**
  - **For a contributing frequency, we compute its phase value to locate the region from which the frequency originates**

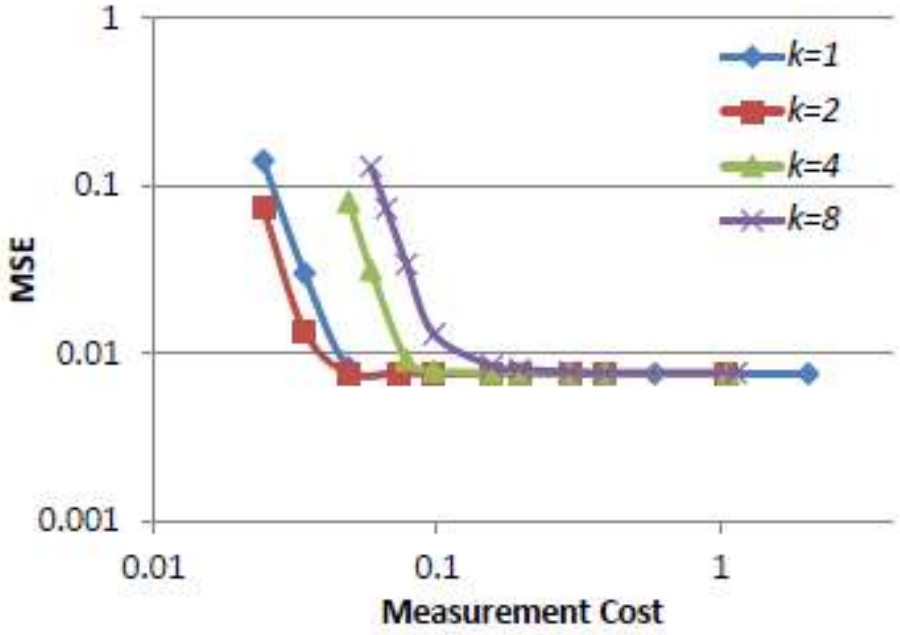
- **Settings:**
  - **Backdrop resolution:** $n = 1024$
  - L_1 **minimization: dynamic group sparsity (DGS)**
- **Implemented in MATLAB R2014b**
  - **The matte extraction at each pixel is independent & are performed in parallel**
- **Run on an 8-core PC with 3.4GHz Intel Core i7 CPU & 24GB RAM**
  - **Processing time varies between 10~100 minutes**

Experiments

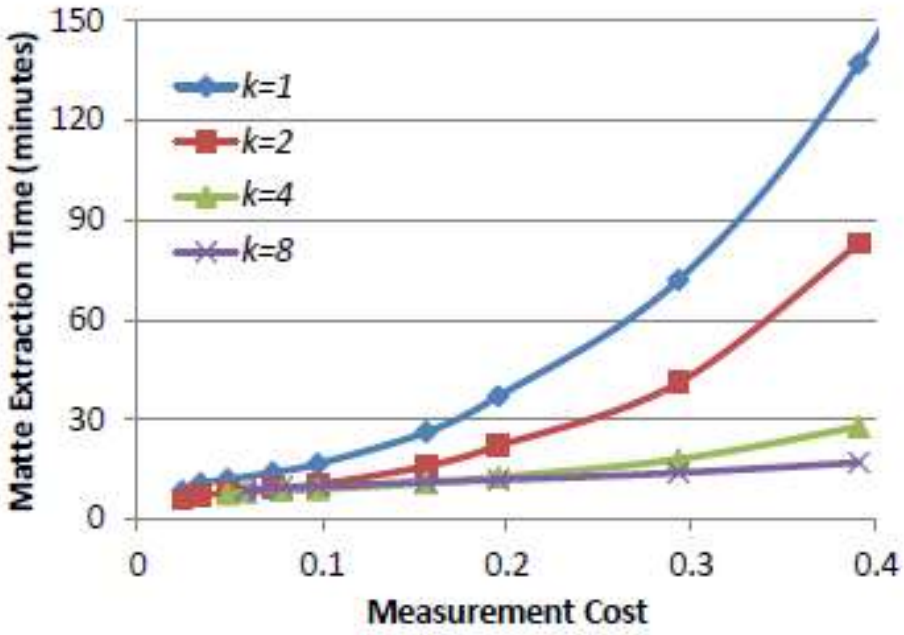
- **The efficiency of CS is usually quantified using measurement cost:**
  - **The ratio between the number of measurements & the number of unknowns**
    - **Need to compute $\bar{\mathbf{W}}_{row}$ & $\bar{\mathbf{W}}_{col}$, a total of $2n$ unknowns**
    - **If the number of images captured is m, then the measurement cost is $\frac{m}{2n}$**
  - **The original frequency-based method has a measurement cost of 2**

Measurement Cost

- **Use POV-Ray tracing library to simulate the data**

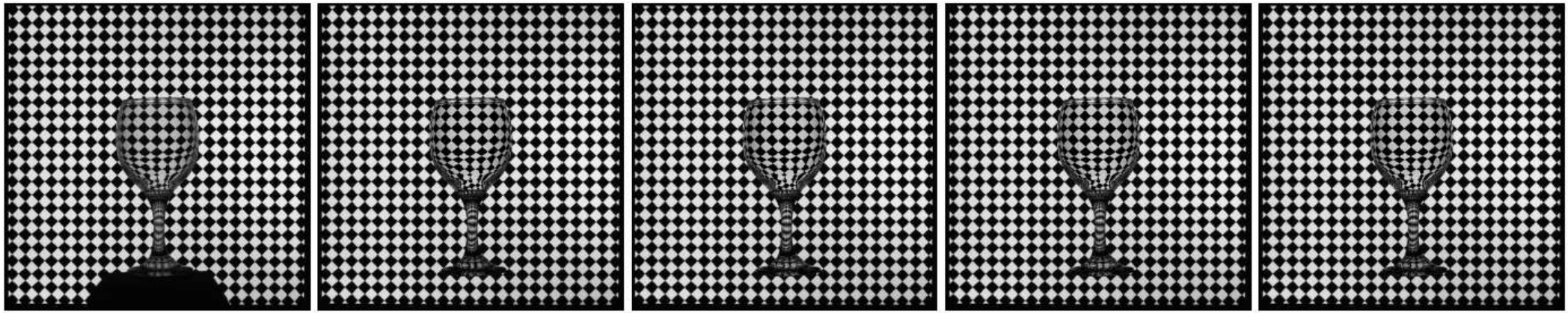


(a) Accuracy

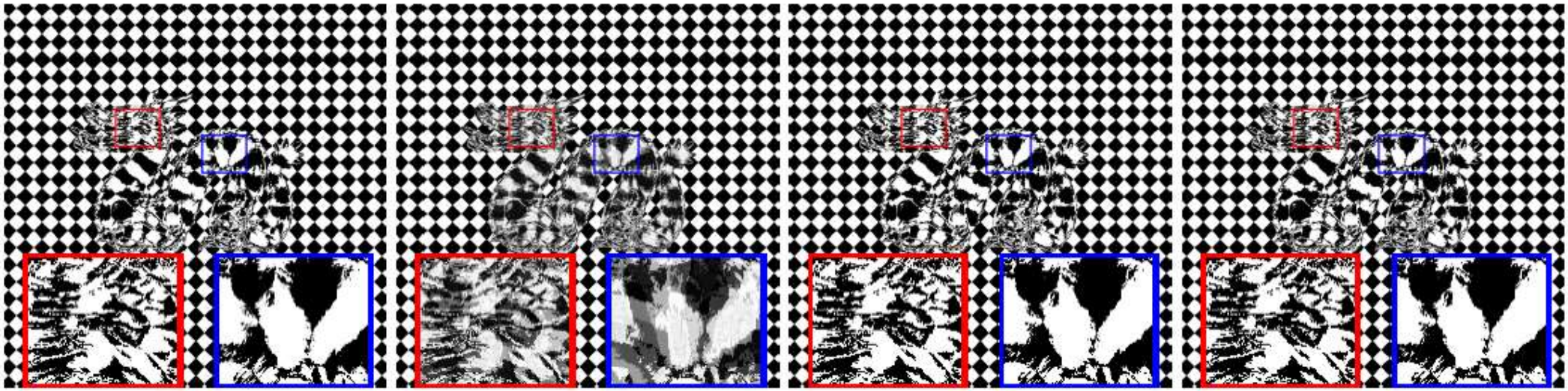


(b) Matte Extraction Time

Quantitative Evaluation on Synthetic Object

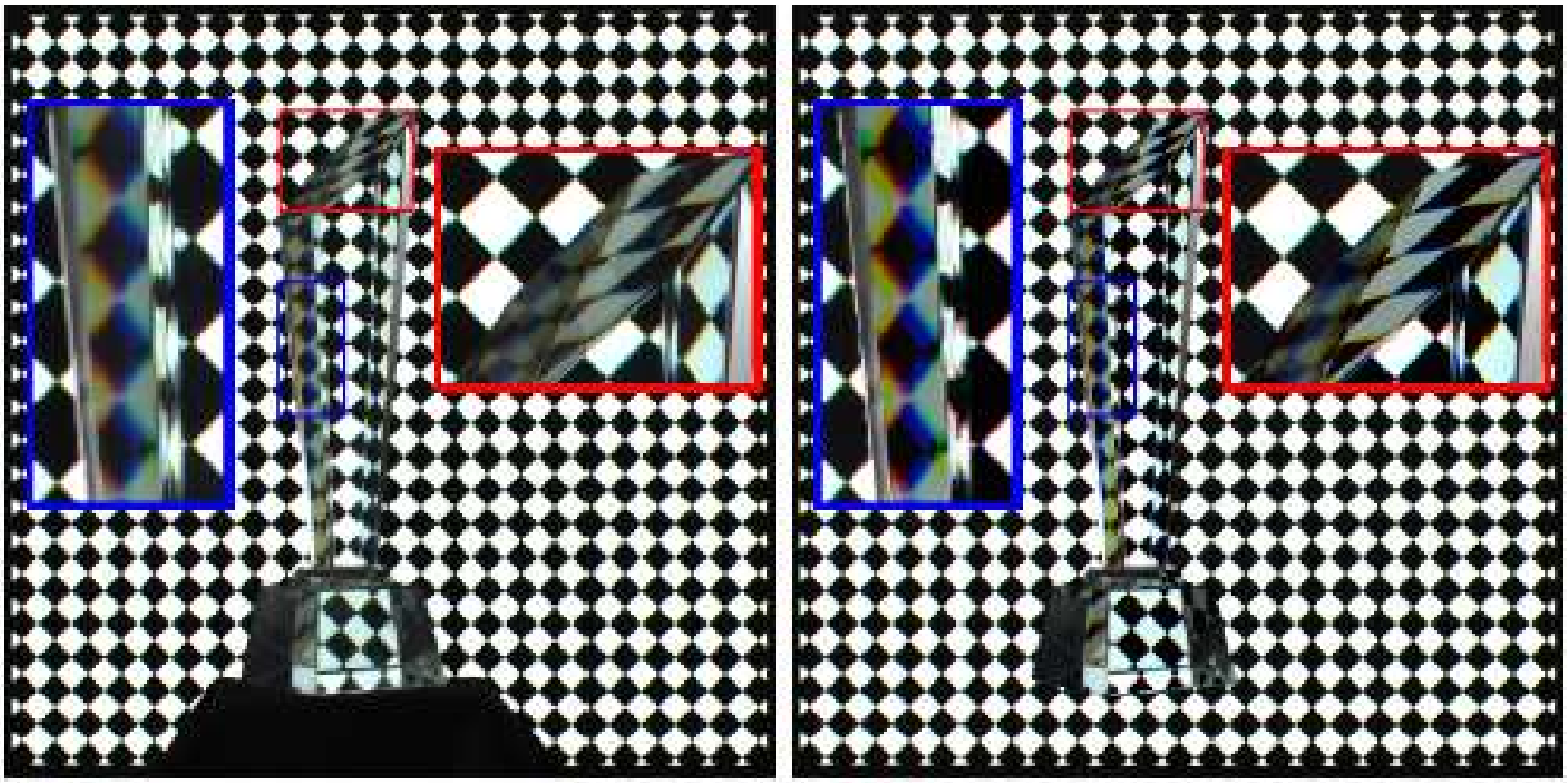


(a) Photograph (b) 200 images by CS (c) 400 images by CS (d) 800 images by CS (e) 2120 images by DFT



(a) Ground Truth (b) Duan *et al.*, 340 images captured (c) Ours without Phase, 320 images captured (d) Ours with Phase, 320 images captured

Real Transparent Objects



(a) Photograph (b) Composite

Dispersion Effect

- **Contributions:**
  - Accurately locate the contributing sources
  - Apply CS to reduce the data acquisition cost
  - Augment phase information to further cut acquisition cost & processing time

- **Limitations:**
  - Assume W can be decomposed into the element-wise product of a row vector & a column vector
  - May lead to artifacts when a foreground pixel has two non-adjacent dominating contributing regions

Summary

Part II

Reconstruct Transparent Surface Shape

- **Existing approaches for 3D reconstruction:**
  - **Stereo triangulation:**
    - **Matching by intensity, followed by triangulation**
  - **Structured light:**
    - **Illuminate with light pattern, followed by triangulation**
  - **Time-of-flight:**
    - **Measure the time between light emission & observed reflection**

Laser scanner

Multi-view stereo

RGBD camera

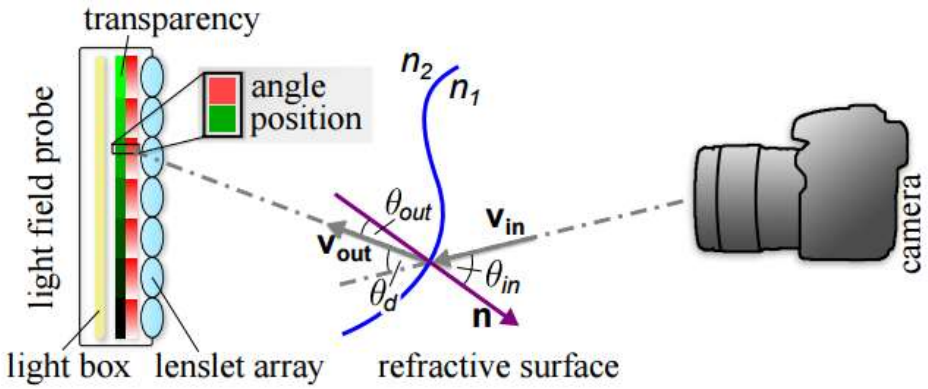
3D Reconstruction for Opaque Object

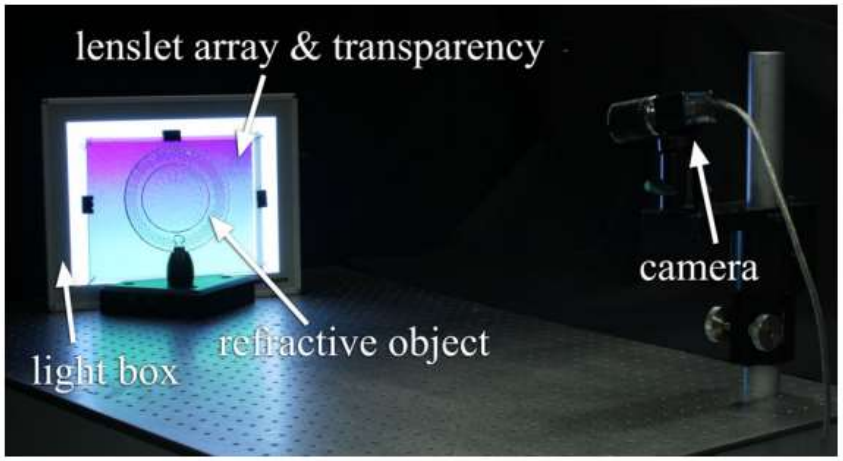
Transparent Objects

- The appearance is mostly determined by refraction
  - The intensity is view dependent
  - Stereo matching does not work
- Does not have strong light reflection
  - Structured light & time-of-flight do not work

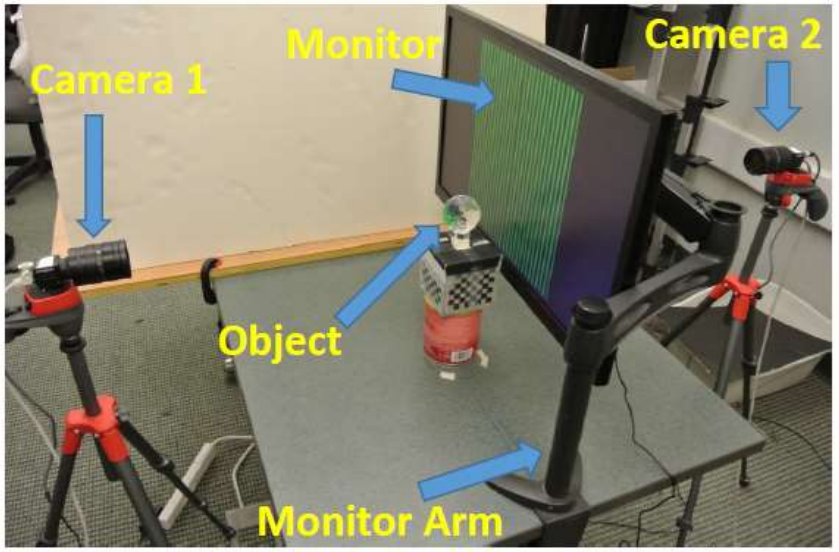
Refraction-based Triangulation

- **Proposed by Wetzstein et al. in ICCV 2011**
  - **Use light field probes to acquire the correspondences between the incident & exit rays**
  - **Assume object is thin & hence light is refracted only once**
  - **Compute refraction positions through triangulation**



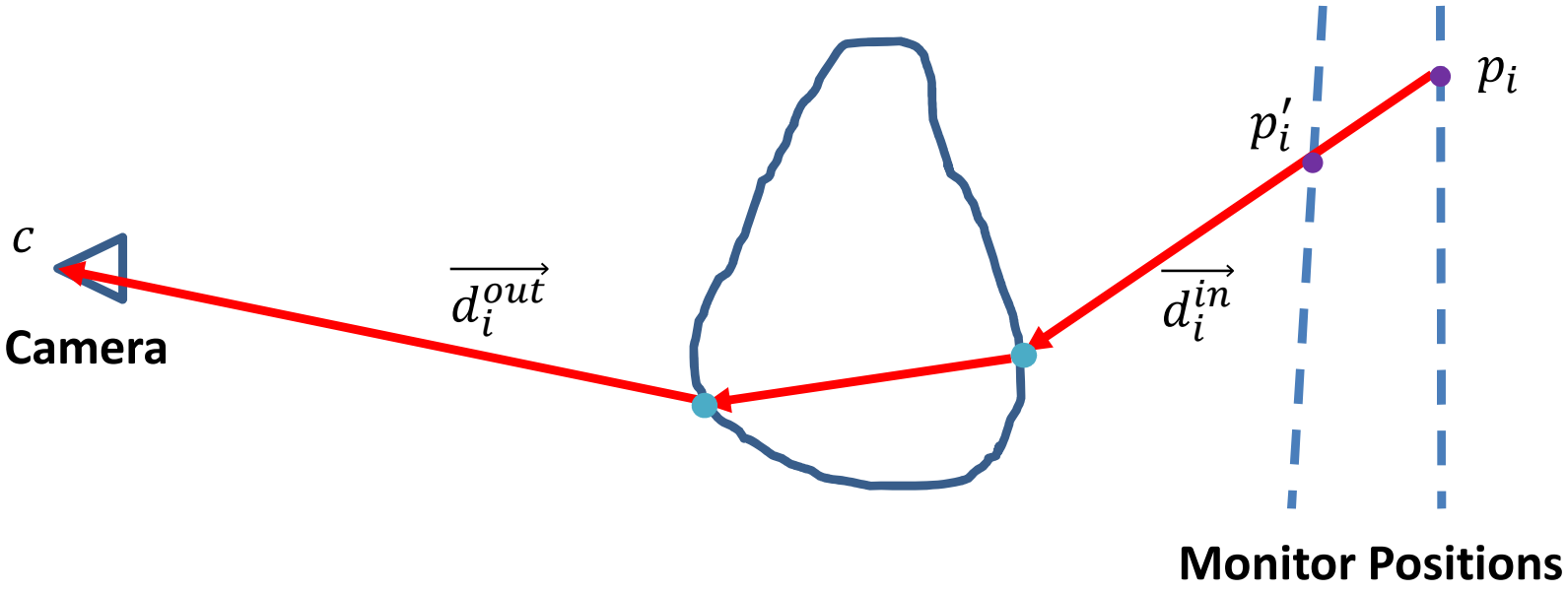


- Use two cameras & a monitor
  - Perform EM at two monitor locations
  - Measure where the incident way comes from for each observed (exit) ray
- Assume two refractions
  - Can handle thick objects



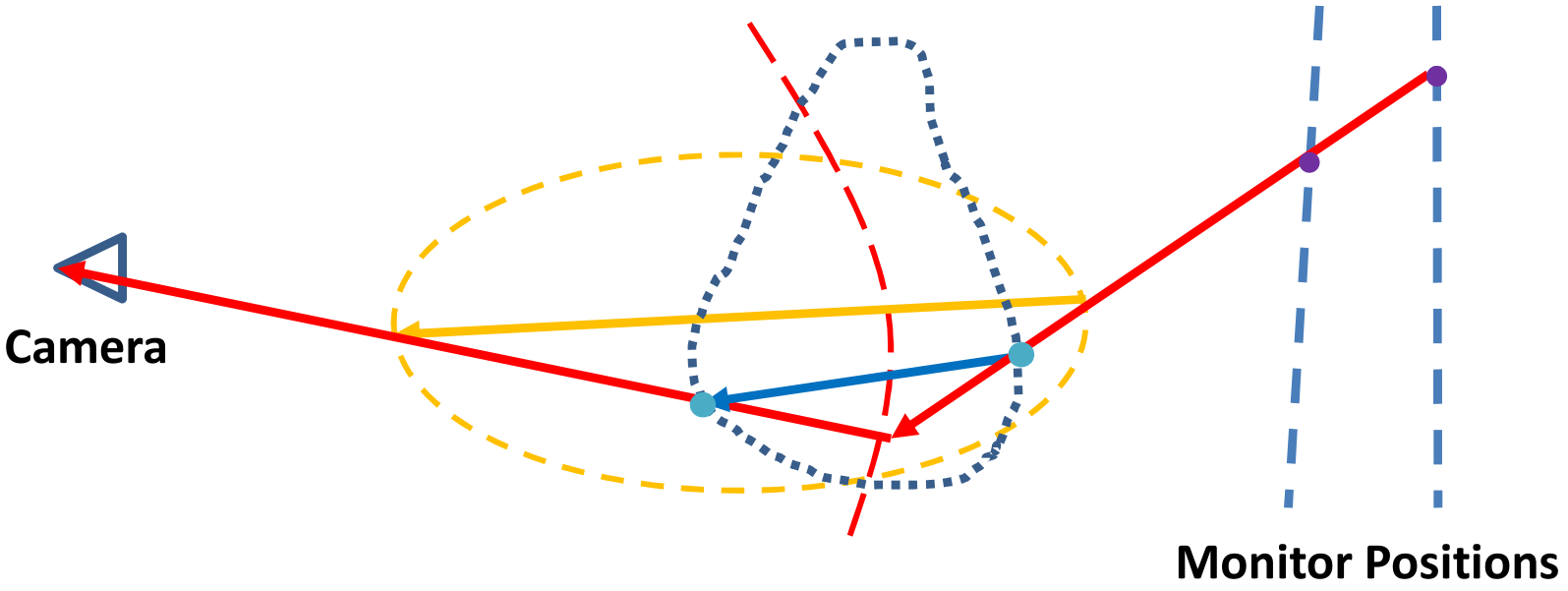
Our Approach

- **Environment matting measures the location of the contribution source, no directional information**
  - **Capturing the ray-ray correspondences $(p, \overrightarrow{d^{in}}) \Leftrightarrow (c, \overrightarrow{d^{out}})$ requires performing EM twice**



Ray-Ray Correspondences Acquisition

- **Thin surfaces**
  - **Refraction location can be computed directly**
- **Thick surfaces**
  - **The light path cannot be determined**

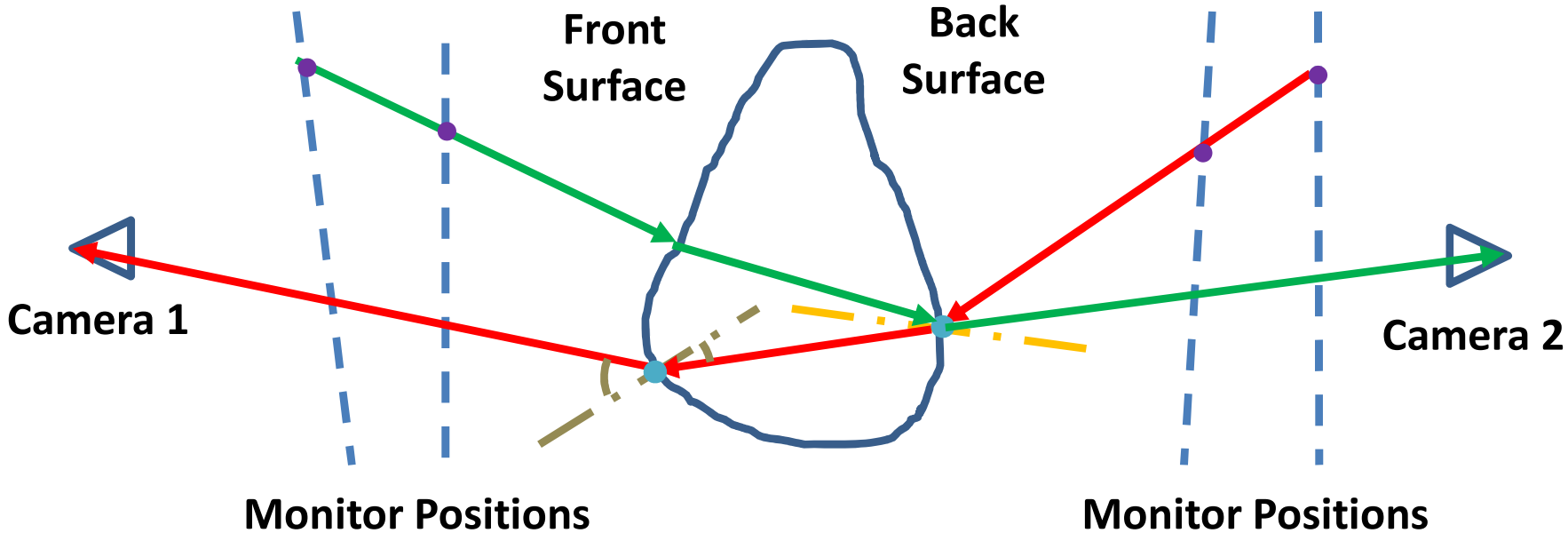


Surface Ambiguities

- **Each 3D surface point can only have one unique normal**
  - **A normal can be estimated from the 3D positions of neighboring points**
    - **PCA normal**
  - **Another normal can be computed for generating the observed light refraction effect**
    - **Snell's law normal**
  - **The two normals shall be consistent at both front & back surfaces**

Position-Normal Consistency (PNC)

- **Enforcing PNC at single refraction location does not provide enough constraints**
  - **Capture ray-ray correspondences from both front & back of the object**
  - **The normal measured from both sides shall be the same**



Enforce PNC at Both Refraction Locations

- **Minimize a position-normal consistency term & a smoothness term for both front & back surfaces:**

$$-\min_{D_f, D_b}\left(\sum_{i\in\Omega} E_{pnc}(i) + \lambda\left(E_{so}\left(D_f\right) + E_{so}(D_b)\right)\right)$$

  - D_f: **depth map of front surface**
  - D_b: **depth map of back surface**
  - Ω: **the set containing all the ray-ray correspondences**

Objective Function

The Two Terms

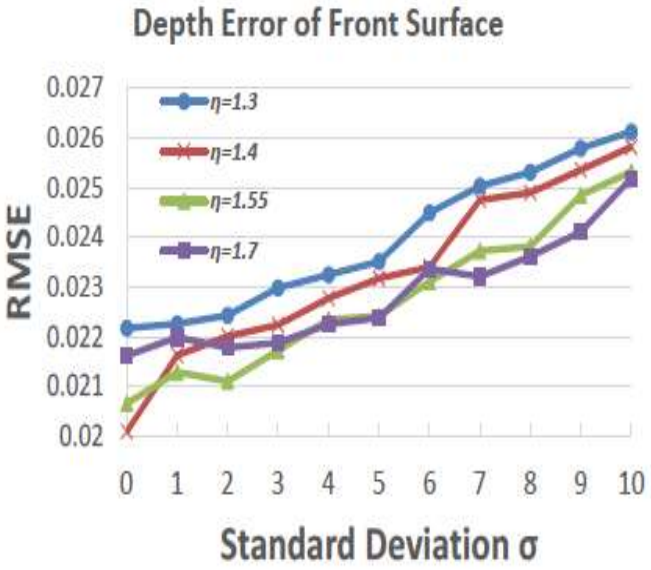
- **For the i^{th} ray-ray correspondence, the normal consistency term is measured as:**
  - $E_{pnc}(i) = 1 - |P(i) \cdot S(i)|$
    - $P(i)$**: the PCA normal**
    - $S(i)$**: the Snell's law normal**
- **The smoothness term is defined as:**
  - $E_{so}(D) = \sum_{s \in D} \sum_{t \in \mathcal{N}(s)} \big(D(s) - D(t)\big)^2$
    - D **: the depth map of refraction surface**
    - $\mathcal{N}(s)$**: the local neighborhood of pixel** s

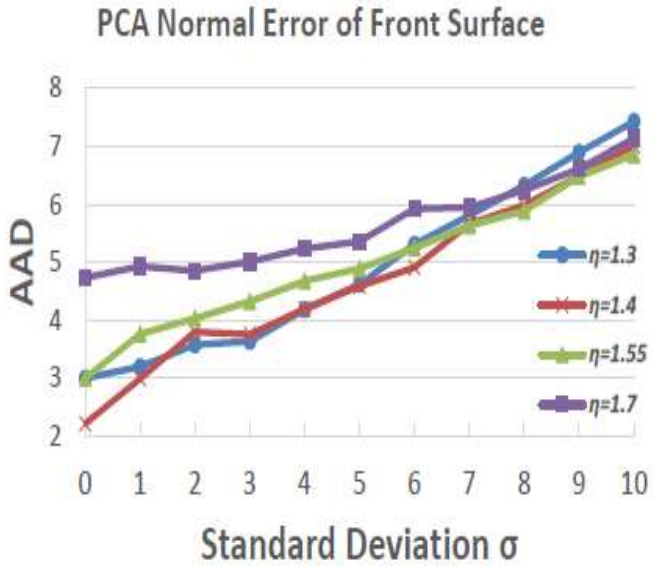
- **Implemented in MATLAB R2014b**
  - **The PCA & Snell normal calculations for different pixels are independent & are computed in parallel**
- **Run on an 8-core PC with 3.4GHz Intel Core i7 CPU & 24GB RAM**
  - **Processing time varies between 1-2 hours**

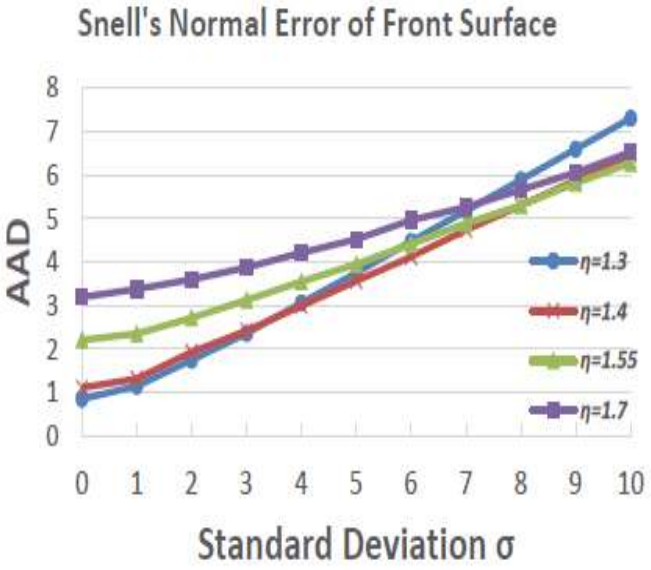
Experiments

- **Use a ray-tracer to simulate the refraction effect of a sphere**
- **Three metrics for evaluation:**
  - **Root mean square error (RMSE) of depths**
  - **Average angular difference (AAD) between the true normal & PCA normal**
  - **AAD between the true normal & Snell's law normal**

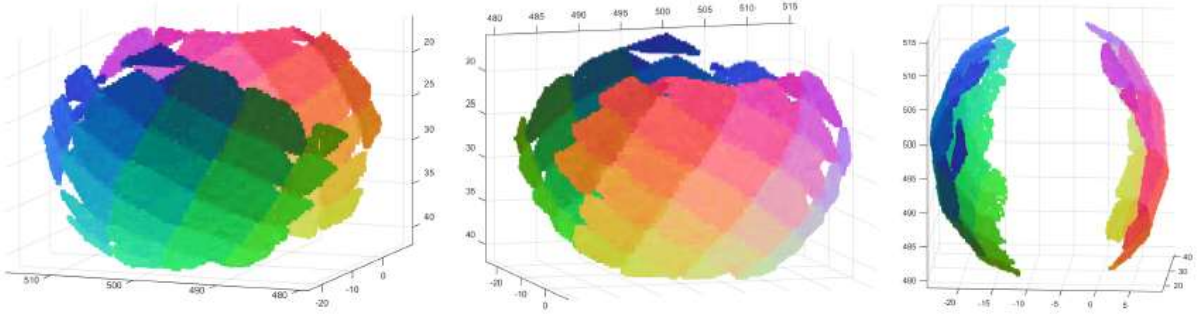
Synthetic Object



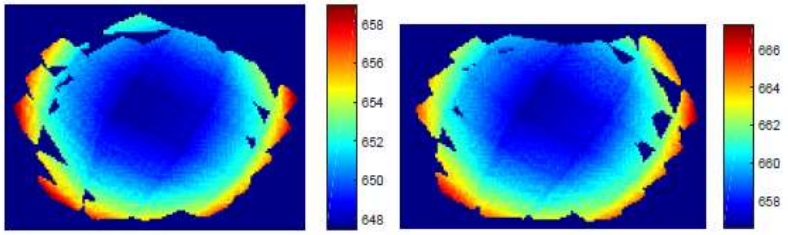




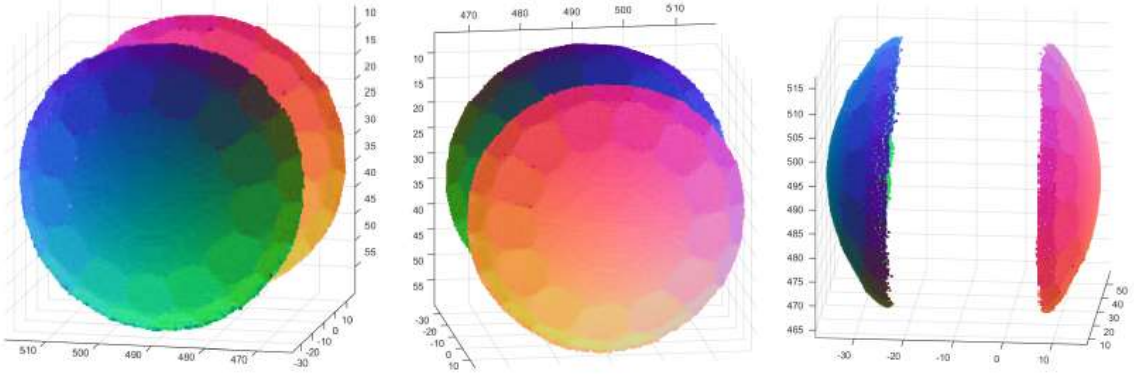
Quantitative Evaluation



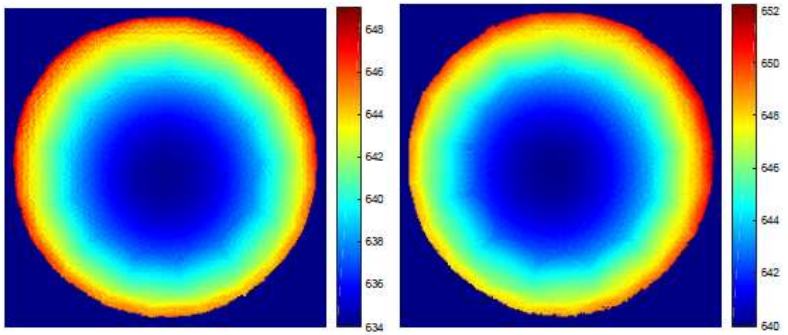
(a) Point cloud of the "ornament" object



(b) Depth maps of the "ornament" object

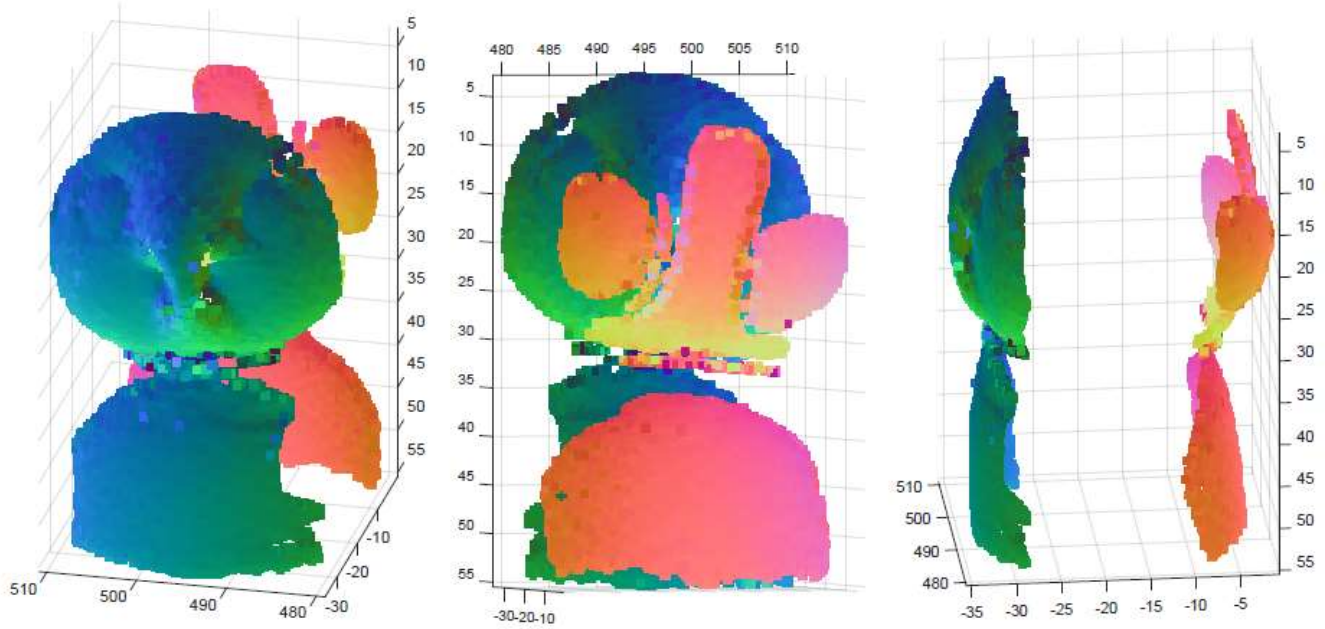


(c) Point cloud of the "ball" object



(d) Depth maps of the "ball" object

Real Objects



(a) Point cloud of the "bird" object

(b) (c) (d) (e)

Real Object (Cont'd)

- **Contributions:**
  - Simultaneous 3D position & normal estimation
  - Refractive index estimation

- **Limitations:**
  - The estimated Point cloud is incomplete
  - Thousands of images need to be captured
  - Assume homogeneous objects & two refraction events

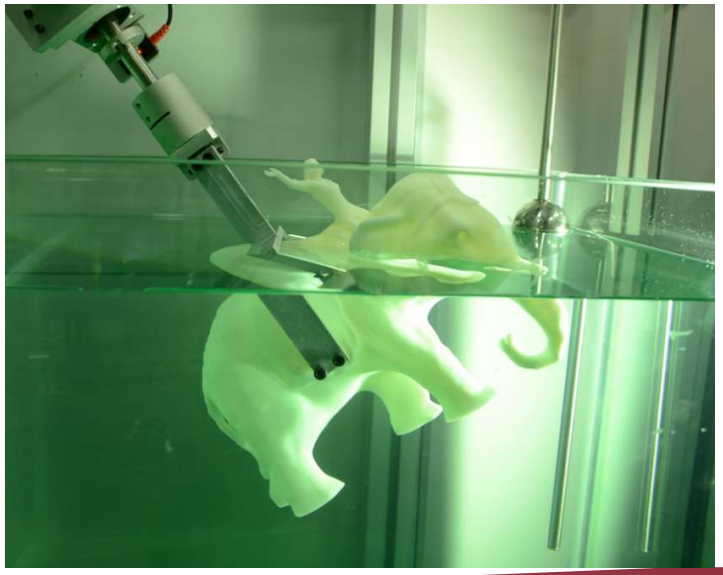
Summary

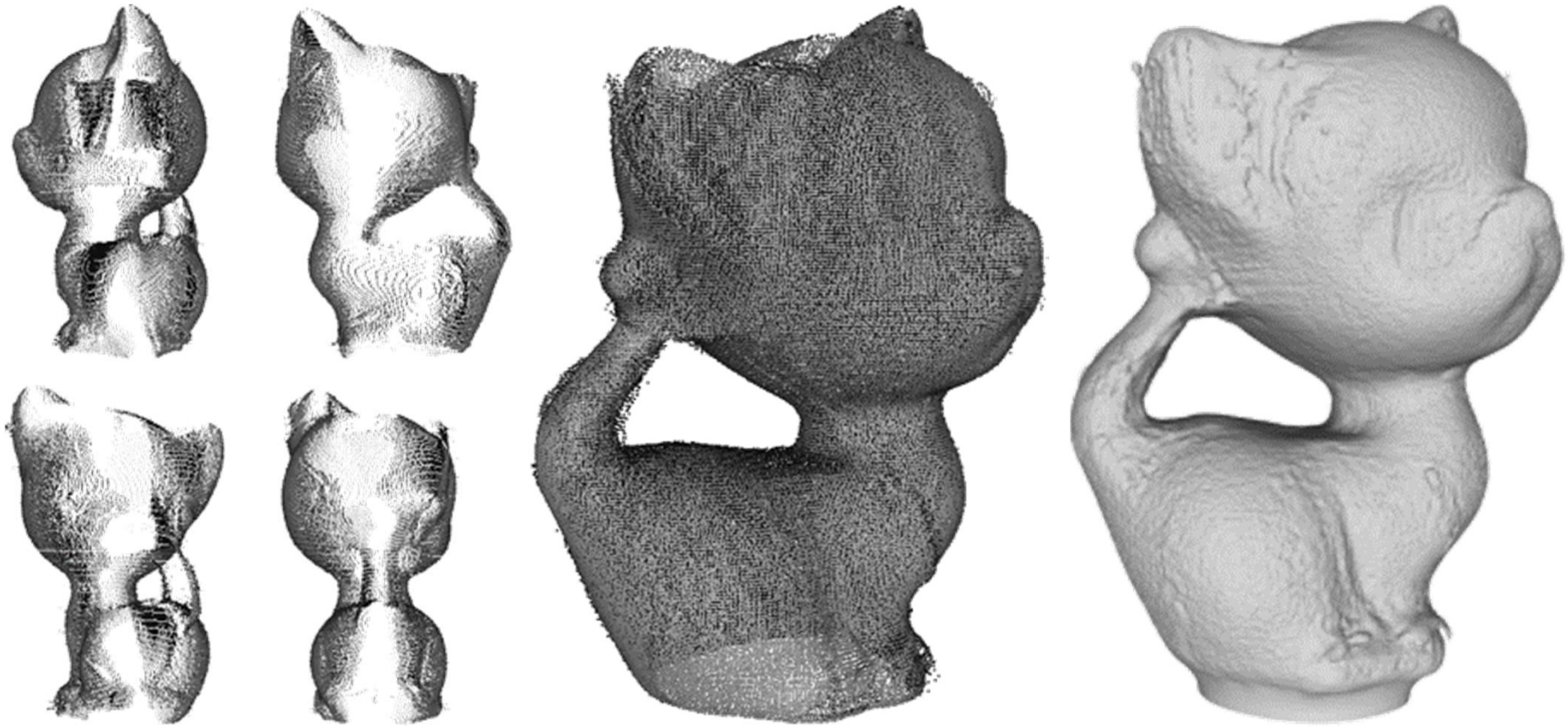
Part III

Reconstruct Full Transparent Object Model

Intrusive Acquisition

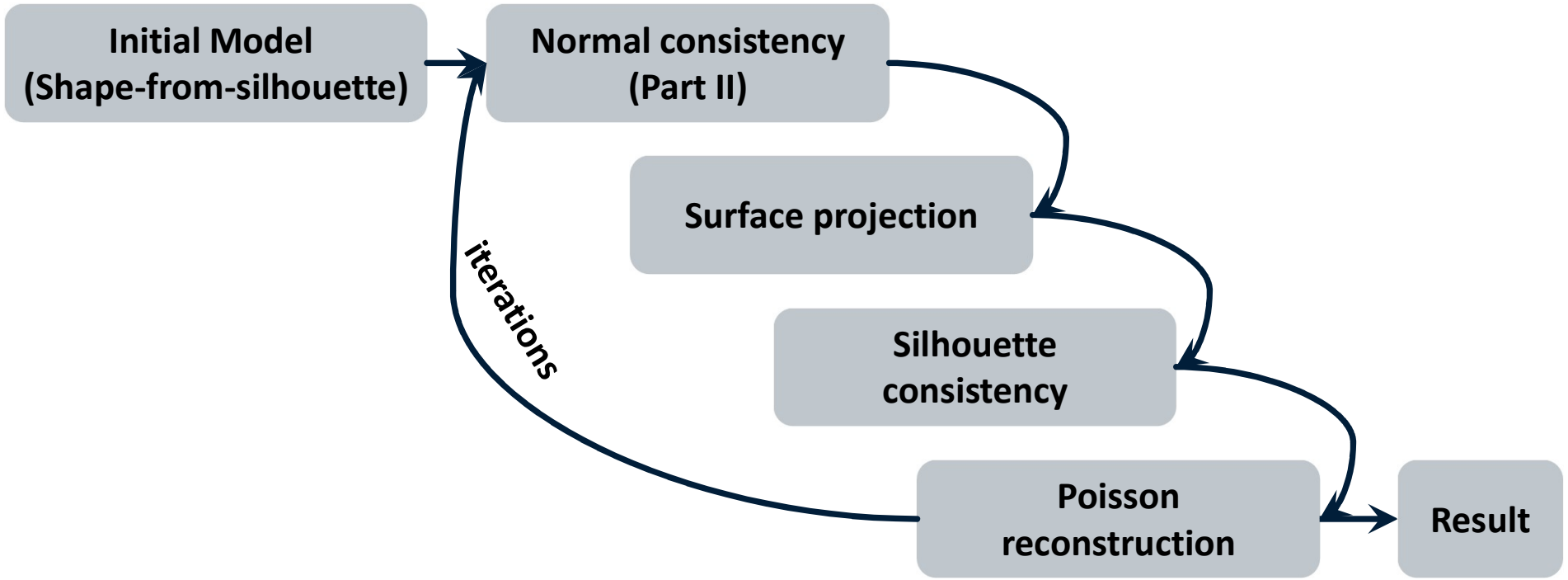
- Paint then scan
- Fluorescent immersion
  - Put object in fluorescent liquid & analyze the light rays that are visible due to fluorescence
- Dip transform
  - Dip object in liquid in different orientations & measure volume displacement





Directly Merge Point Clouds from Part II

- **Start from initial model obtained by space carving**
- **Progressively recovers geometric details through optimizing light refraction, silhouette, & smoothness constraints**

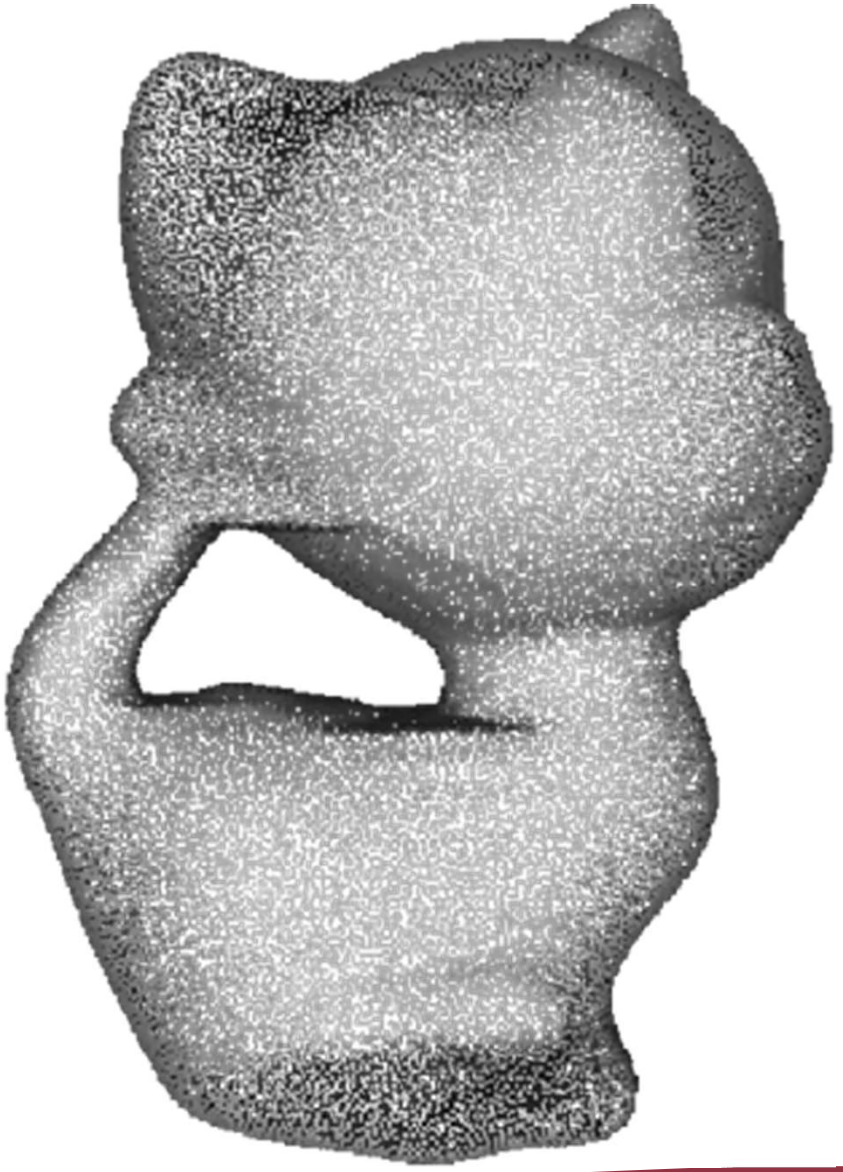


Our Approach

- **Object is placed on Turntable #1**
  - **LCD monitor on Turntable #2 serves as light source**
  - **Camera #1 captures silhouettes & ray-pixel correspondences**
  - **Camera #2 looks at Turntable #1 for its rotation axis calibration**

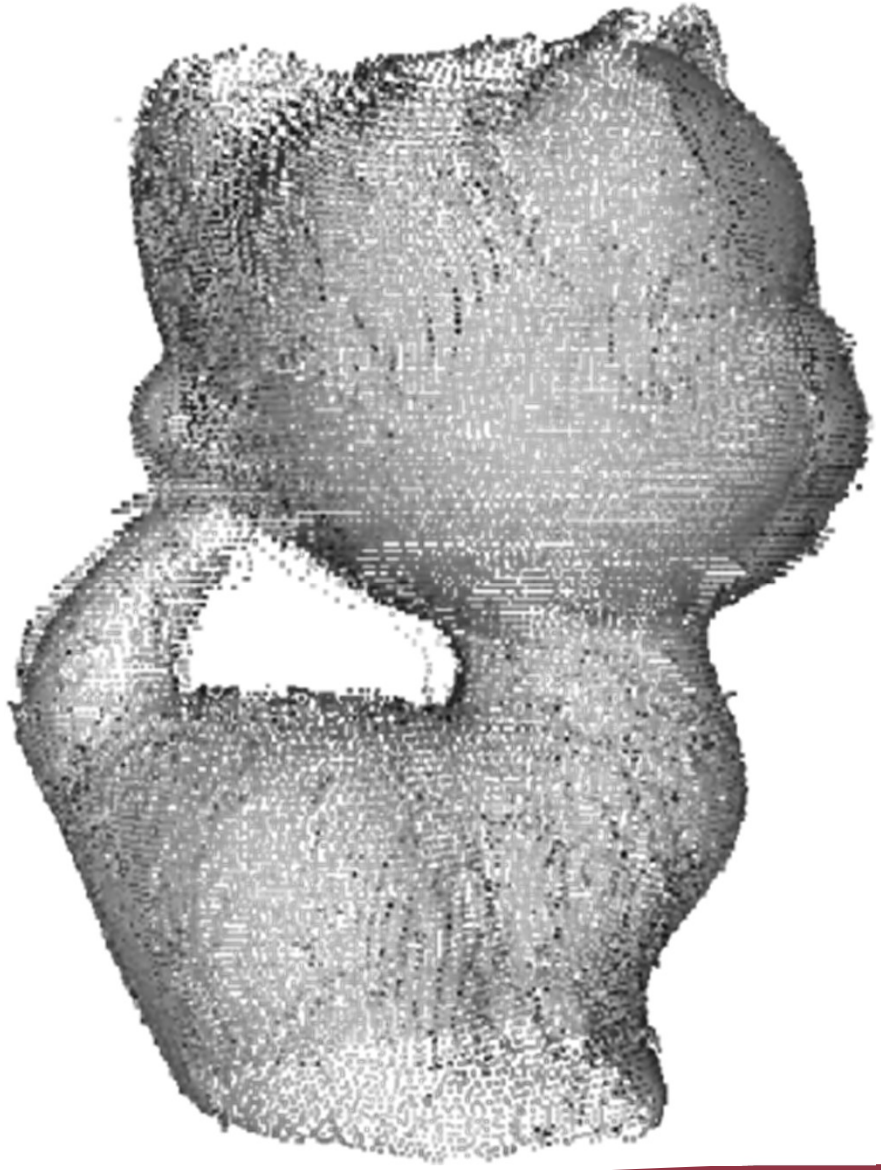
Automatic Capturing Setup

- **Generate initial model:**
  - **Gather the silhouettes of the object under different viewpoints**
  - **Compute the space within the silhouettes by space carving**
    - **[Kutulakos & Seitz 2000]**
- **The model is complete but inaccurate**
  - **Does not capture concave areas well**



Initial Rough Model

- Generate point clouds
  - Each set of front & back viewpoints is used to compute a point cloud
  - Different sets merge into a point set surface
- The point set surface is incomplete & noisy
  - No data on top/bottom
  - Not well-aligned
  - Better capture concave areas



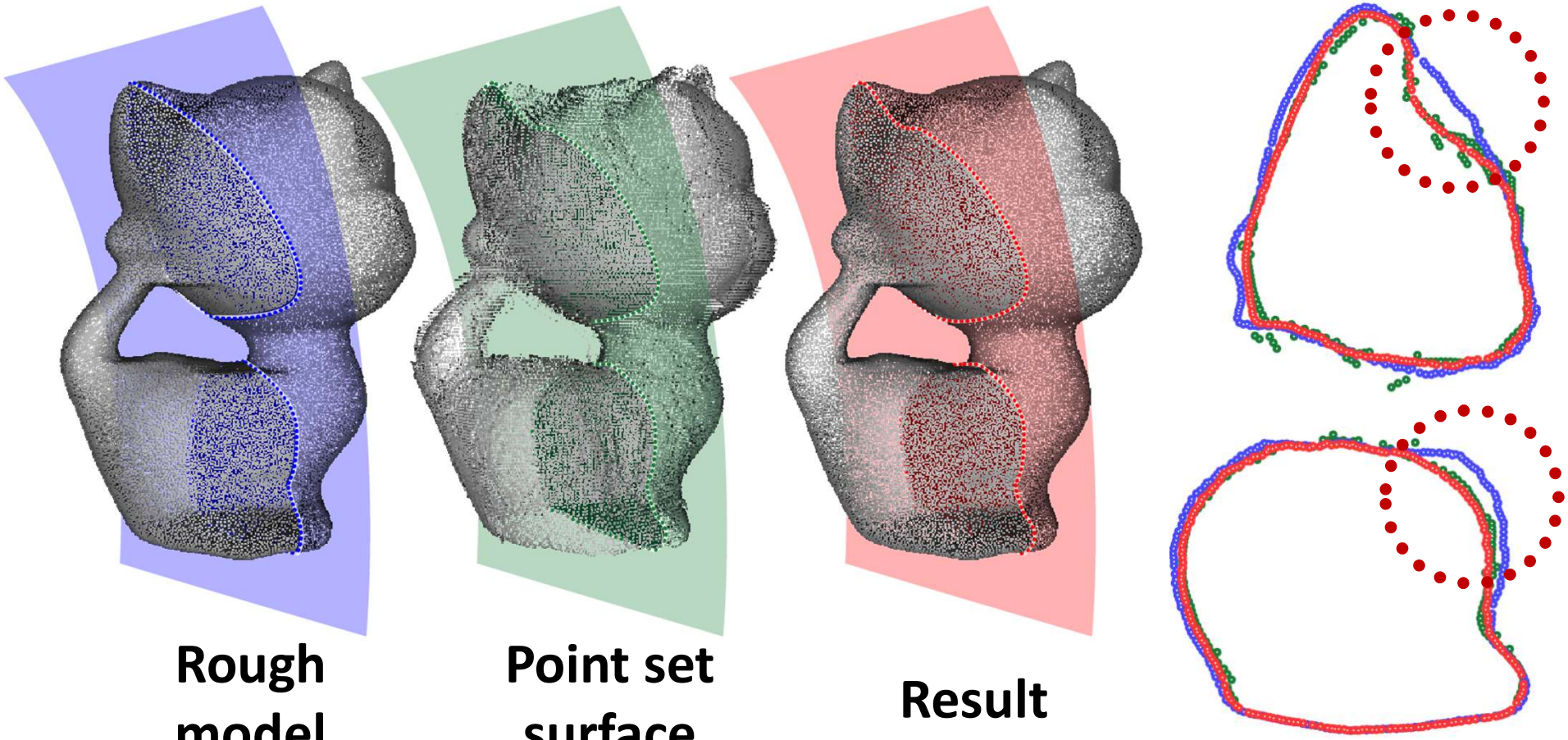
Point Cloud from Part II

- **Sample points on the complete model**
- **Project them toward the point set surface**

$$-\sum_{j\in J}\left(\begin{array}{c}\sum_{i\in I}\left\|x_j^{k+1}-p_i\right\|e^{\frac{-\left\|x_j^k-p_i\right\|^2}{\left(h_j/4\right)^2}}\\ +\frac{\alpha}{|\mathcal{N}_j|}\sum_{j'\in\mathcal{N}_j}\left\|\Delta_j-\Delta_{j'}\right\|^2\end{array}\right)$$

  - **1st term moves toward the point set surface**
  - **2nd term maintains smoothness & completeness**

Point Consolidation



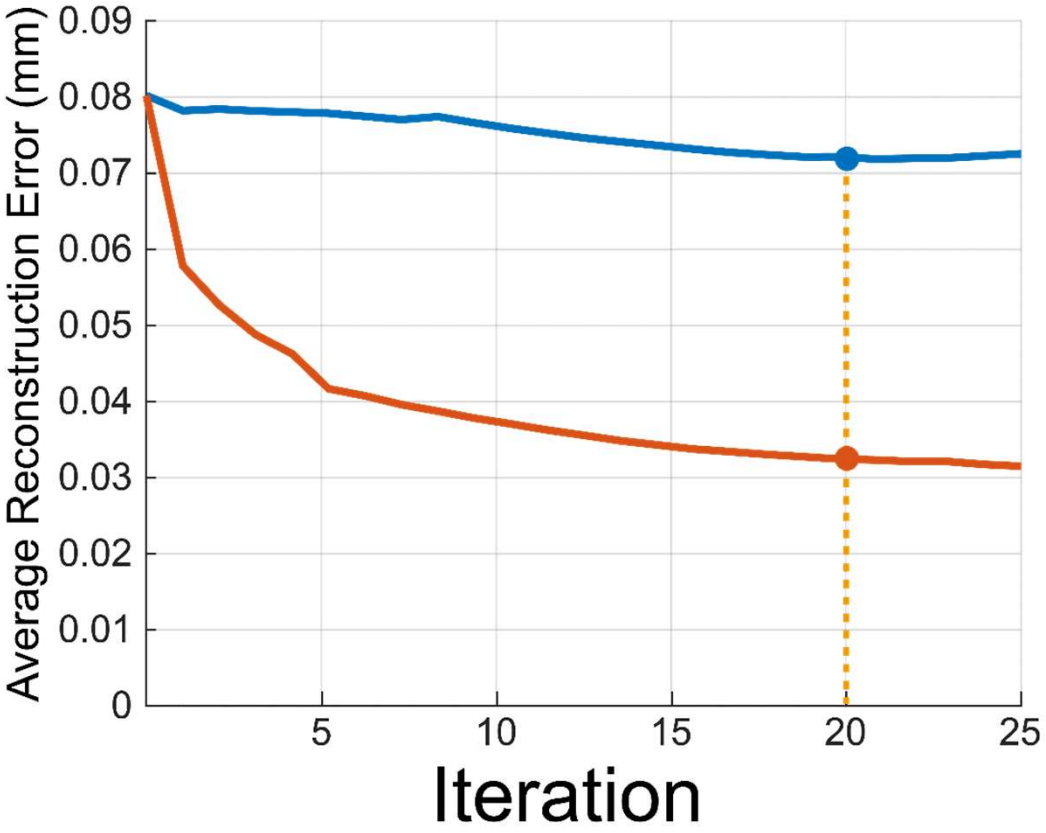
Point Consolidation (Cont'd)

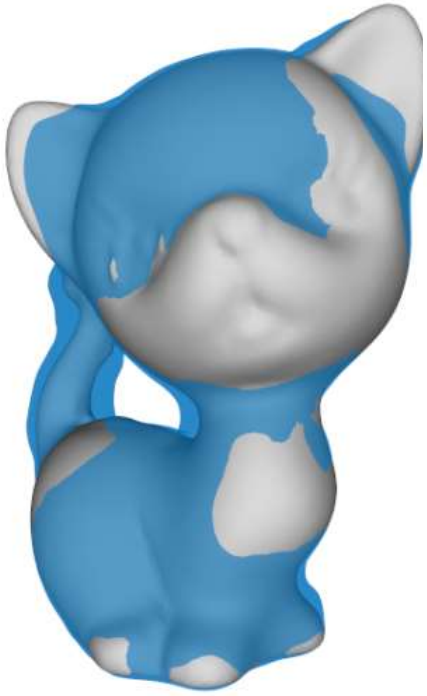
- **Point estimated from Part II are not accurate**
  - **Can lead the projection away from silhouettes**
  - **Enforcing consistency with silhouettes helps**

$$- \min_{X} \sum_{j \in J} \left(\begin{array}{c} \sum_{v=1}^{V} \delta_j^v D\left(q_j^v, \partial\Omega_v\right) \\ + \frac{\beta}{|\mathcal{N}_j|} \sum_{j' \in \mathcal{N}_j} \left\| \Delta_j - \Delta_{j'} \right\|^2 \end{array} \right)$$

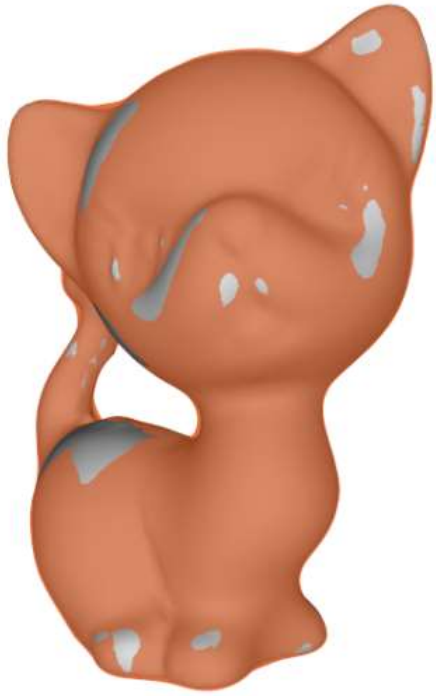
  - **1st term minimizes distance between surface projection & silhouettes**
  - **2nd term maintains smoothness**

Silhouette Consistency



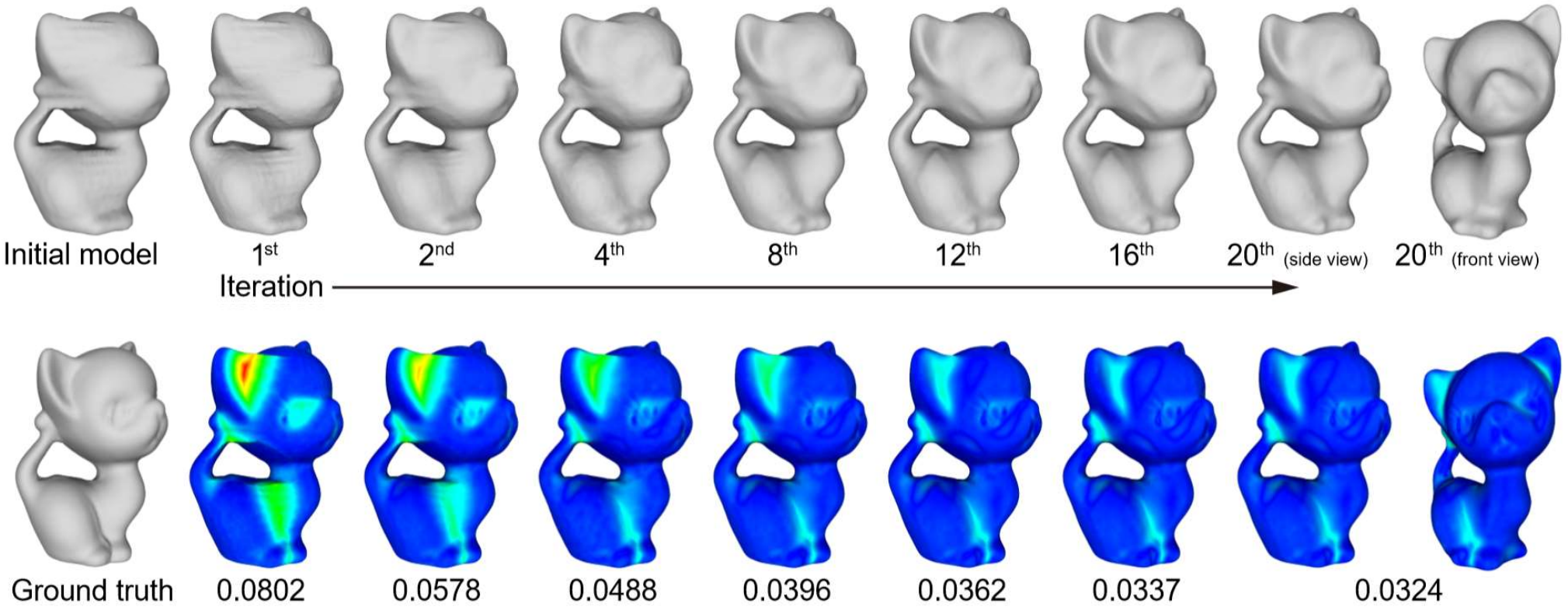


Without silhouette constraints



With silhouette constraints

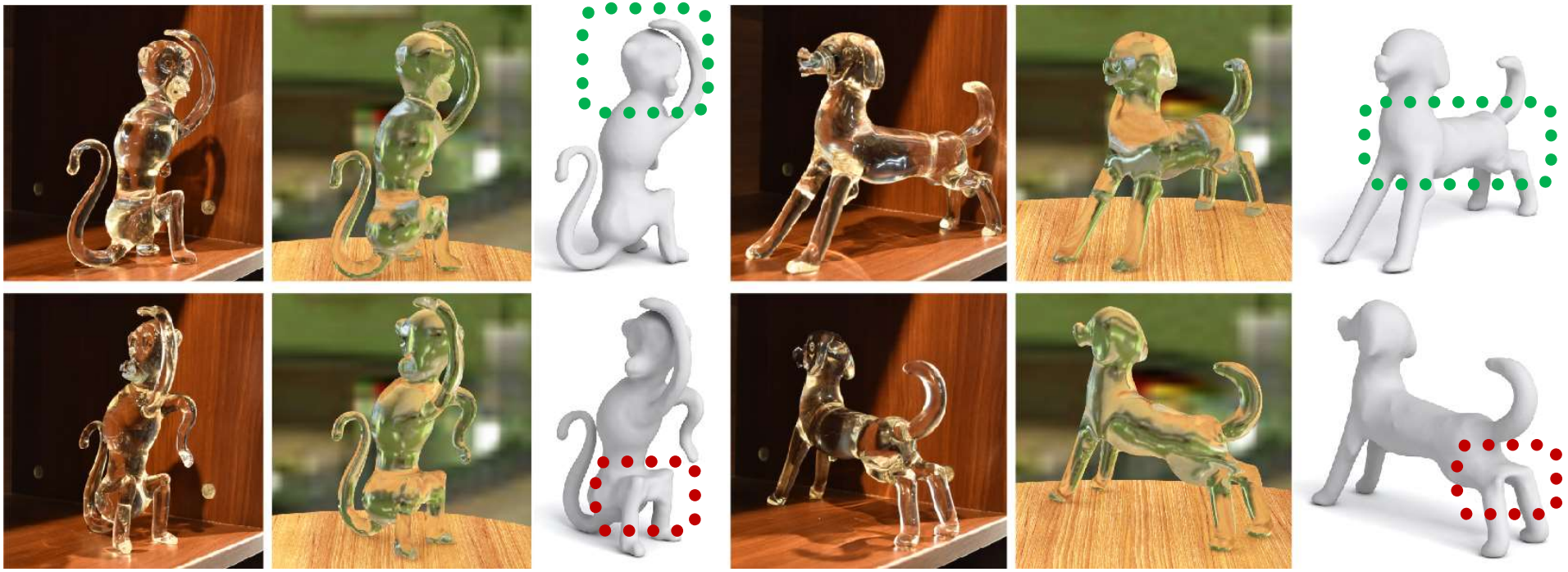
Silhouette Consistency (Cont'd)



Repeat until Converge

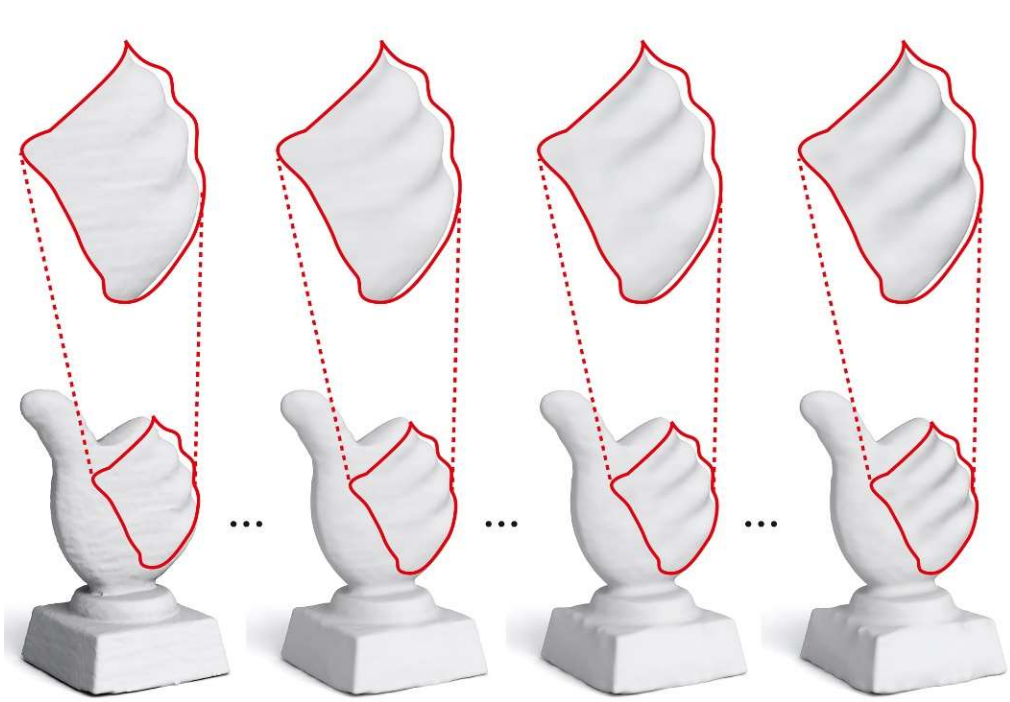
Experiment on Synthetic Object

Experiment on Real Object



Real image | Rendered image | Reconstruction | Real image | Rendered image | Reconstruction

Experiment on More Real Objects

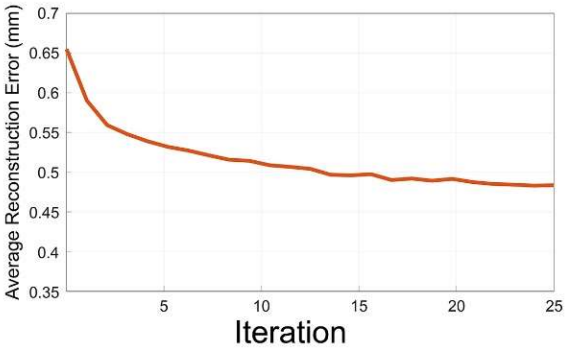


(a)

(b)

(c)

(d)



(e)

Quantitative Evaluation

- **Contributions:**
  - An automatic setup for capturing ray-ray correspondences
  - Adaptive surface projection for point consolidation
  - Silhouette consistency
- **Limitations:**
  - Thousands of images need to be captured
  - Assume homogeneous objects & two refraction events

Summary

Future Work

- **How to reconstruct the 3D shape of time-varying surfaces, such as water?**
  - **Cannot capture multiple images with different backdrops at the same time**
  - **Have to make estimation based on a single image**

Dynamic Transparent Surface

- **Yiming Qian, Minglun Gong, & Yee-Hong Yang: Frequency-based environment matting by compressive sensing. *ICCV, December 2015*.**
- **Yiming Qian, Minglun Gong, & Yee-Hong Yang: 3D reconstruction of transparent objects with position-normal consistency. *CVPR, June 2016*.**
- **Bojian Wu, Yang Zhou, Yiming Qian, Minglun Gong, & Hui Huang: Full 3D reconstruction of transparent objects. *Siggraph, August 2018.***

Related Publications

Questions?

- Modeling and rendering real objects are active topics in both computer vision and graphics. Many powerful techniques are available for capturing the 3D shapes and photorealistic appearances of opaque objects, but the ones for handling transparent objects are not as capable. The challenges are due to the facts that transparent objects do not have their own colors but acquire their appearances from the environments and that these objects interact with light in complex manners including reflection, refraction, and scattering.
- Three research projects that advance the state-of-the-art on this front are presented here. The first one investigates how transparent objects interact with the environments using a frequency-based environment matting approach. Unlike existing methods that require thousands of captured images and/or long processing time, our approach exploits compressive sensing theory to extract the matte effectively and efficiently. The second project develops a new refraction-based algorithm for estimating 3D point positions on transparent object surfaces. By introducing a novel surface and refraction normal consistency constraint, an optimization procedure is designed to jointly reconstruct the 3D positions and normals of these points. Finally, the third project aims at reconstructing full 3D models for transparent objects. Starts from a rough but complete 3D model generated from space carving, our algorithm progressively optimizes the model under three constraints: surface and refraction normal consistency, surface projection and silhouette consistency, and surface smoothness.

Abstract

- **Dr. Minglun Gong is a Professor and Head at the Department of Computer Science, Memorial University of Newfoundland. He obtained his Ph.D. from the University of Alberta in 2003, his M.Sc. from the Tsinghua University in 1997, and his B.Engr. from the Harbin Engineering University in 1994. After graduation, he was a faculty member at the Laurentian University for four years before joined the Memorial University in 2007.**
- **Minglun's research interests cover various topics in the broad area of visual computing (including computer graphics, computer vision, visualization, image processing, and pattern recognition). So far, he has published over 100 referred technical papers in journals and conference proceedings, including 19 articles in ACM/IEEE transactions. He is the inventor of an awarded patent and 6 pending patents. Currently an associate editor for Pattern Recognition, he has also served as program committee member for top-tier conferences (e.g. ICCV and CVPR) and reviewer for prestigious journals (e.g. IEEE TPAMI and ACM TOG). He was the recipient of the Izaak Walton Killam Memorial Award, the 2015 Best Paper Award from the Canadian Artificial Intelligence Association, and the 2016 Best Paper Award from the International Symposium on Visual Computing.**

Biography