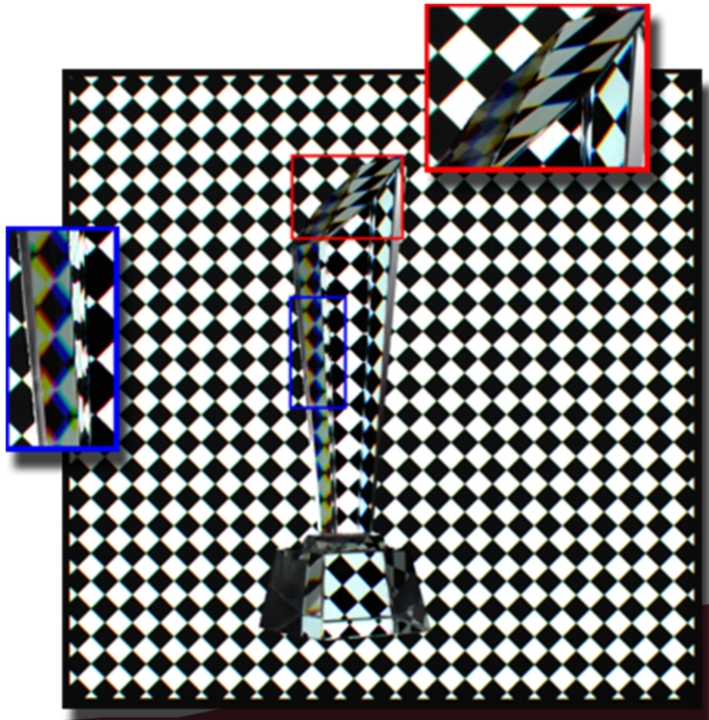
Rendering and Modeling of Transparent Objects

Minglun Gong

Dept. of CS, Memorial Univ.

- **Capture transparent object appearance**
  - Using frequency-based environmental matting
  - Reduce number of input images needed through compressive sensing
- **Reconstruct transparent surface shape**
  - Measure how the light travels from the source to the camera
  - Jointly optimize the 3D positions and normals of the refractive surface using a novel position-normal consistency constraint

Two Recent Projects



Part I

Capture Transparent Object Appearance

- **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 and fuzzy object boundaries**

Extraction of Opaque Objects

- $C = \alpha F + (1 - \alpha)B$
  - 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

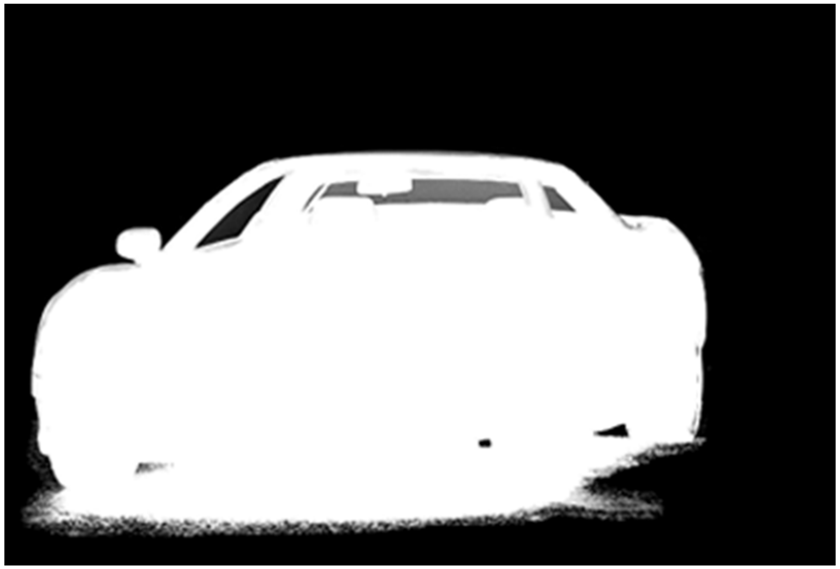


Image Matting Formulation

Transparent Objects

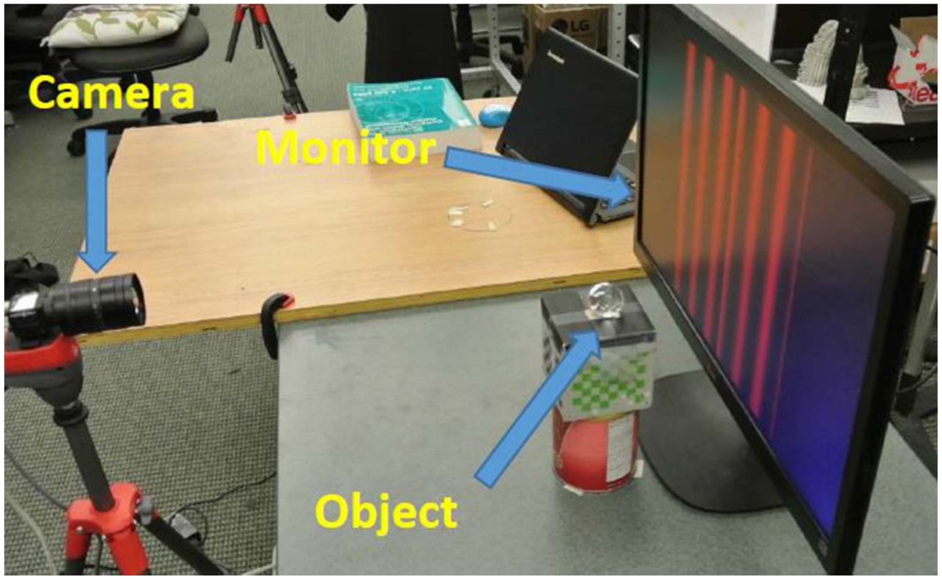
- **Do not have their own colors but acquire their appearances from the environments**
  - **Reflect, refract, and scatter environment light**
- **Require environment matting (EM)**

Environment Matting Formulation

- $C = F + \rho\mathbf{WB}$
  - 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**
    - $\|\mathbf{W}\|_1 = 1, \mathbf{W}_i \geq 0$

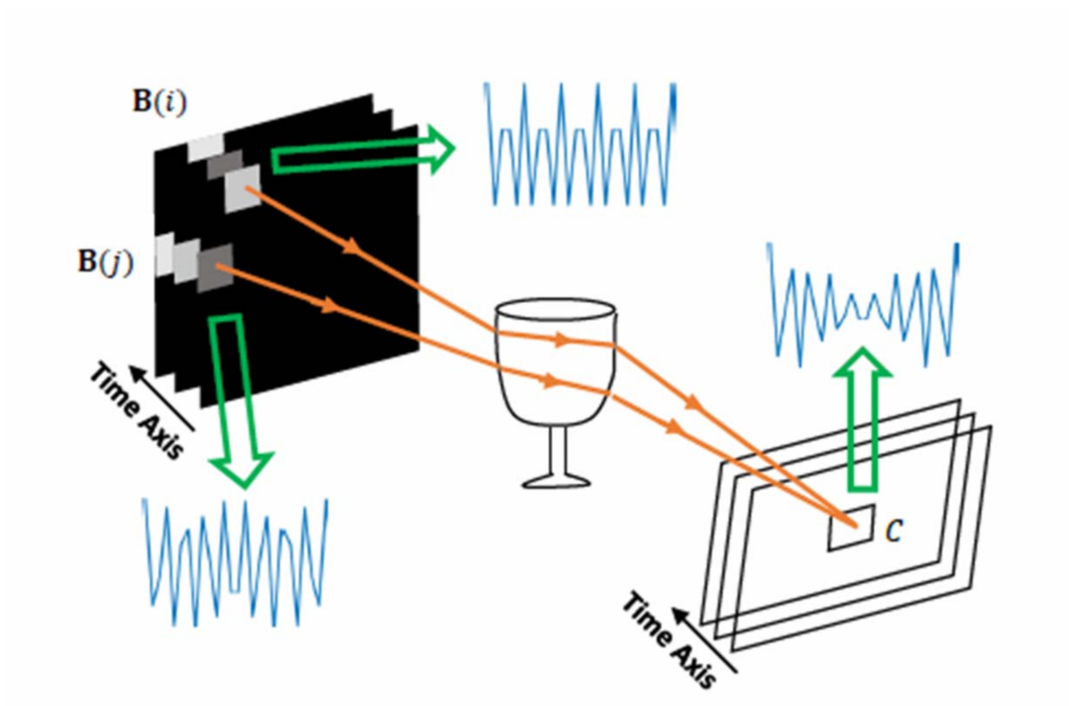
- **Under a solid black backdrop:**
  - $\mathbf{B} = 0$,
  - $C = F$
- **Under a solid white backdrop:**
  - $\mathbf{B} = b$,
  - $\|\mathbf{W}\|_1 = 1$,
  - $C = F + \rho b$

- **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



Environment Matting (Cont'd)

- Proposed by Zhu and 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 W is the element-wise product of a row vector and 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 $M = O\big(s \times log(N/s)\big)$ 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

Reconstruction via CS

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

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

- **Use both frequency and 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** p**th region**
    - $1 \leq p \leq k$
- **Reduce the maximal frequency from** n **to** $\frac{n}{k}$

$1 \leq f \leq \frac{n}{k}, \ \varphi_1$
$1 \leq f \leq \frac{n}{k}, \ \varphi_2$
… …
$1 \leq f \leq \frac{n}{k}, \ \varphi_k$

Augment with Phase Information

- Both frequency search and 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

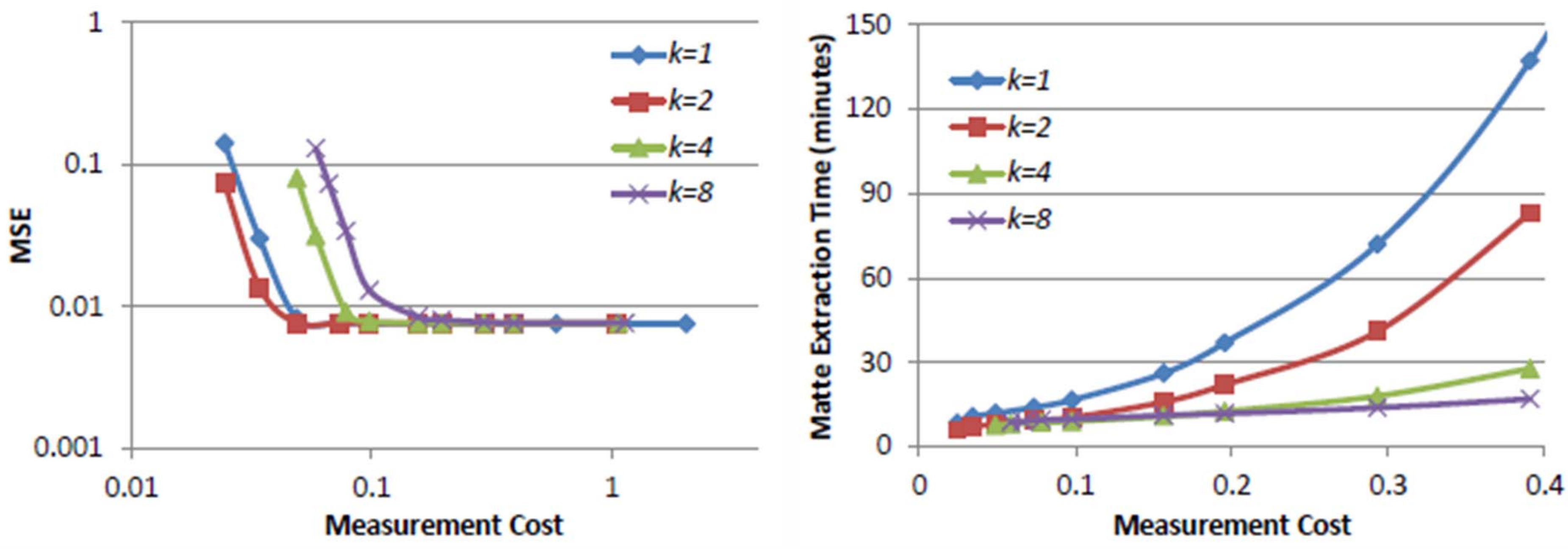
Augment with Phase Information (Cont'd)

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

- **The efficiency of CS is usually quantified using measurement cost:**
  - **The ratio between the number of measurements and the number of unknowns**
    - **Need to compute $\bar{\mathrm{W}}_{row}$ and $\bar{\mathrm{W}}_{col}$, 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**

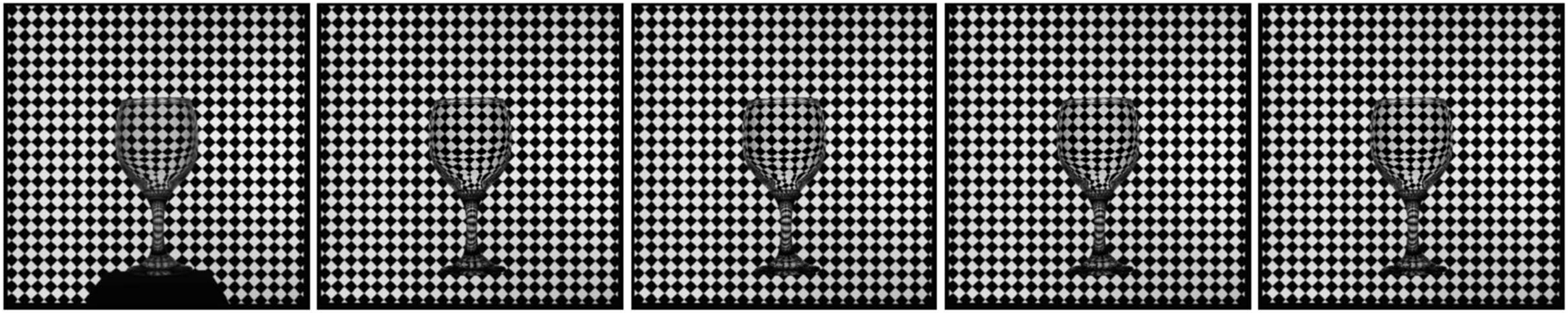
Measurement Cost

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

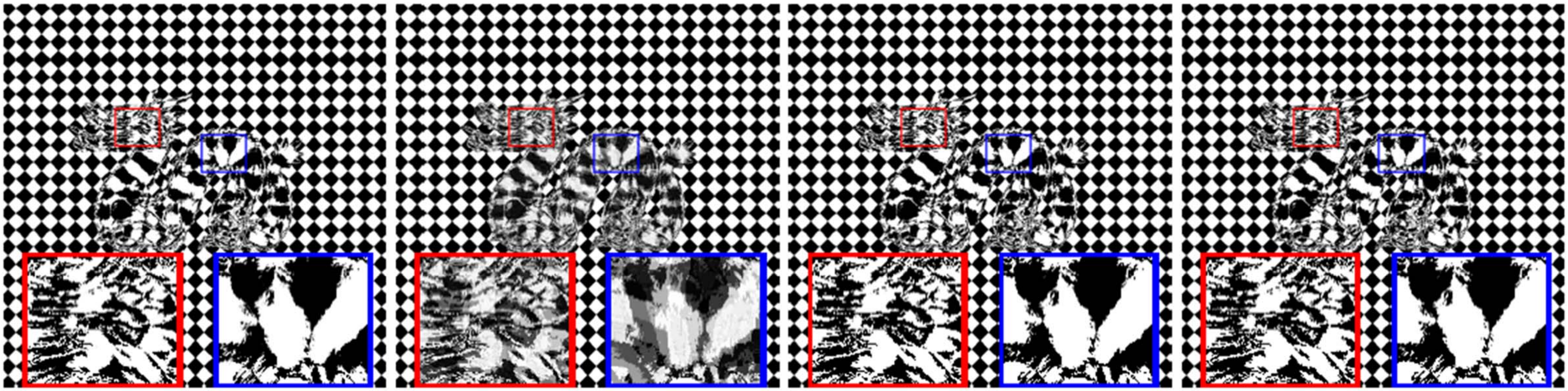


(a) Accuracy (b) Matte Extraction Time

Quantitative Evaluation on Synthetic Object

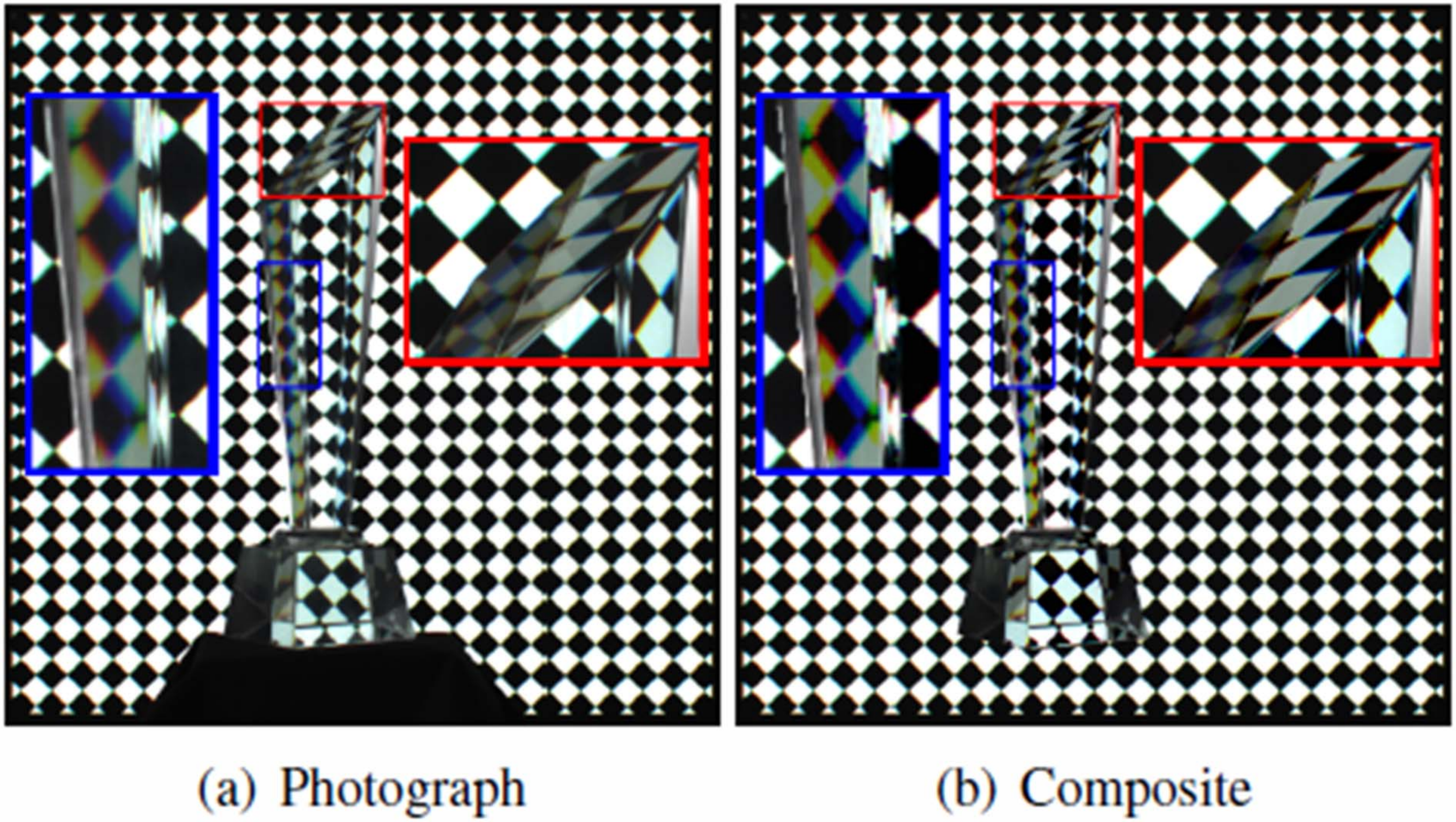


(a) Photograph (b) 200 images by CS (c) 400 images by CS (d) 800 images by CS (e) 2120 images by DFT



(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

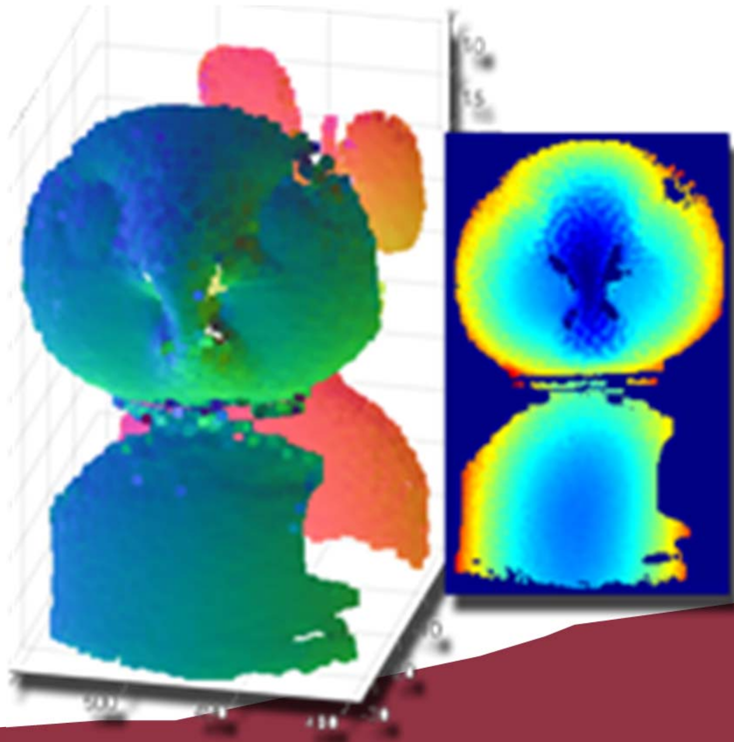


Dispersion Effect

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

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

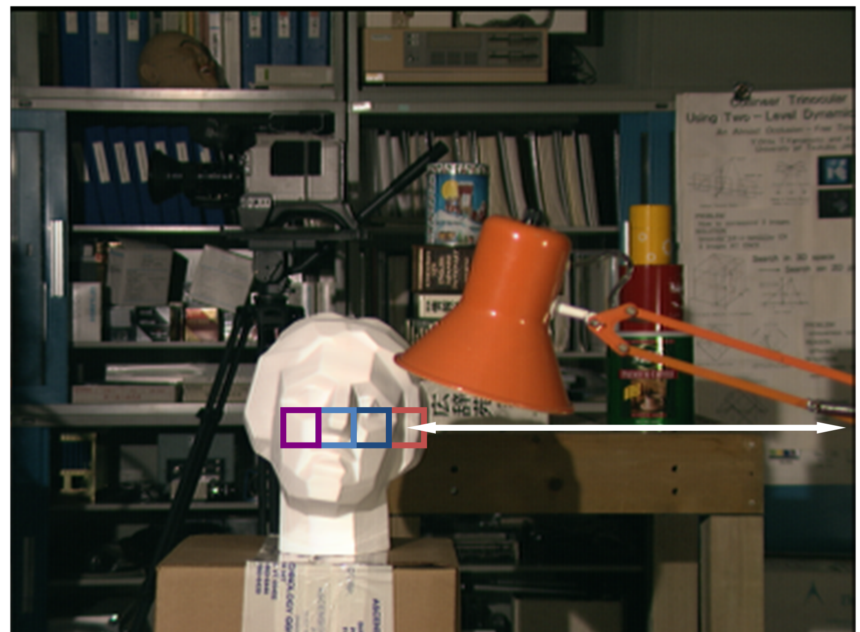
Summary



Part II

Reconstruct Transparent Surface Shape

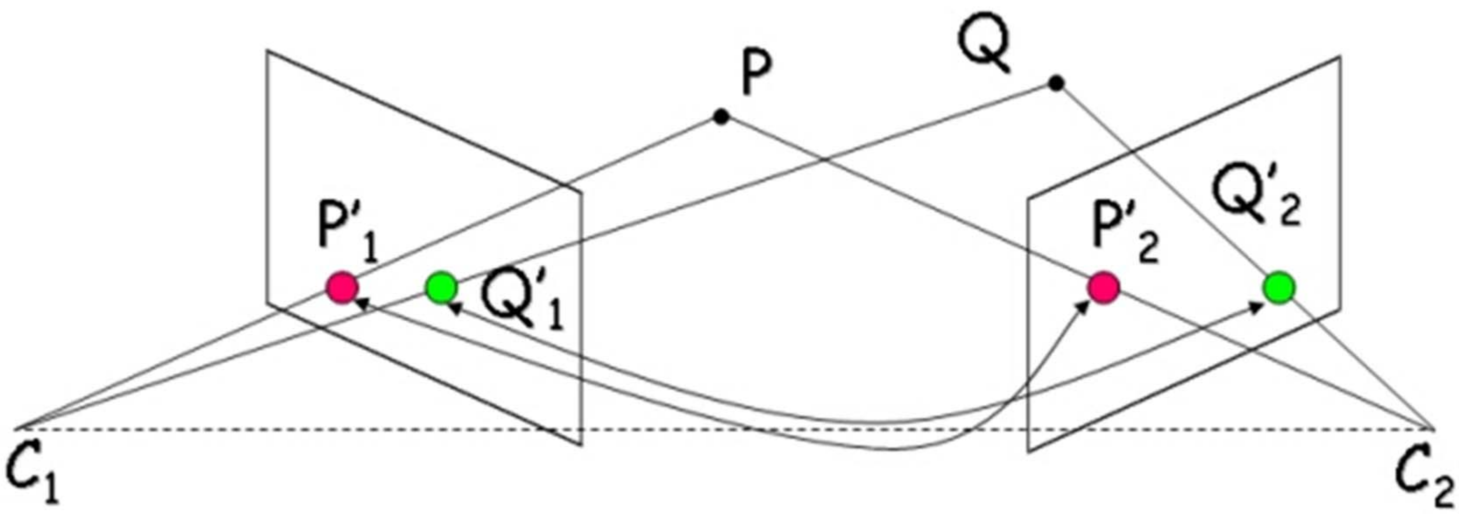
- **Humans rely on stereo matching to extract 3D information**
  - **The disparity between the corresponding pixels indicates the depth of the object**



3D Reconstruction for Opaque Object

Reconstruction for Opaque Object (Cont'd)

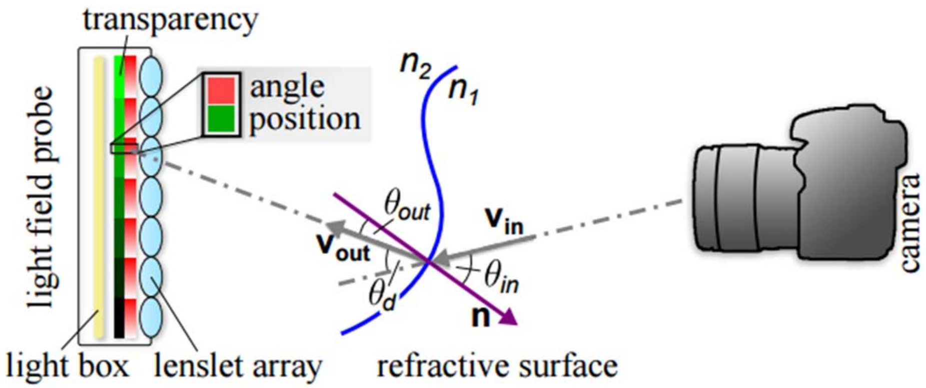
- **Pipeline:**
  - Matching by intensity, followed by triangulation
- **Assumptions:**
  - Each surface point has a unique color/pattern
  - Surface color is the same under different view directions

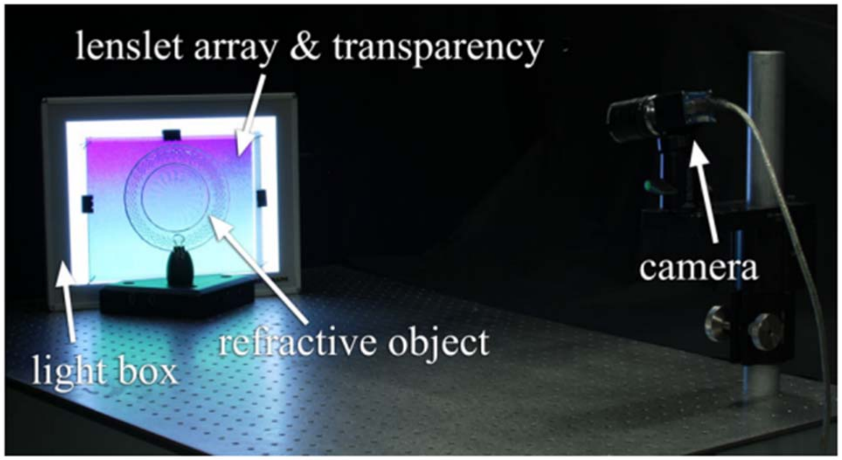


- **The appearance of transparent object is mostly determined by refraction**
- **The intensity is view dependent**

Transparent Objects

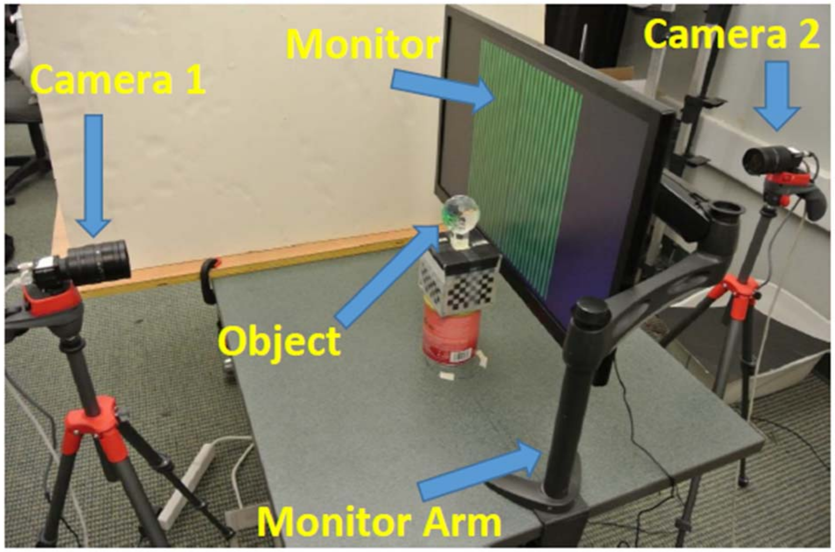
- **Proposed by Wetzstein et al. in ICCV 2011**
  - Use light field probes to acquire the correspondences between the incident and exit rays
  - Assume object is thin and hence light is refracted only once
  - Compute refraction positions through triangulation





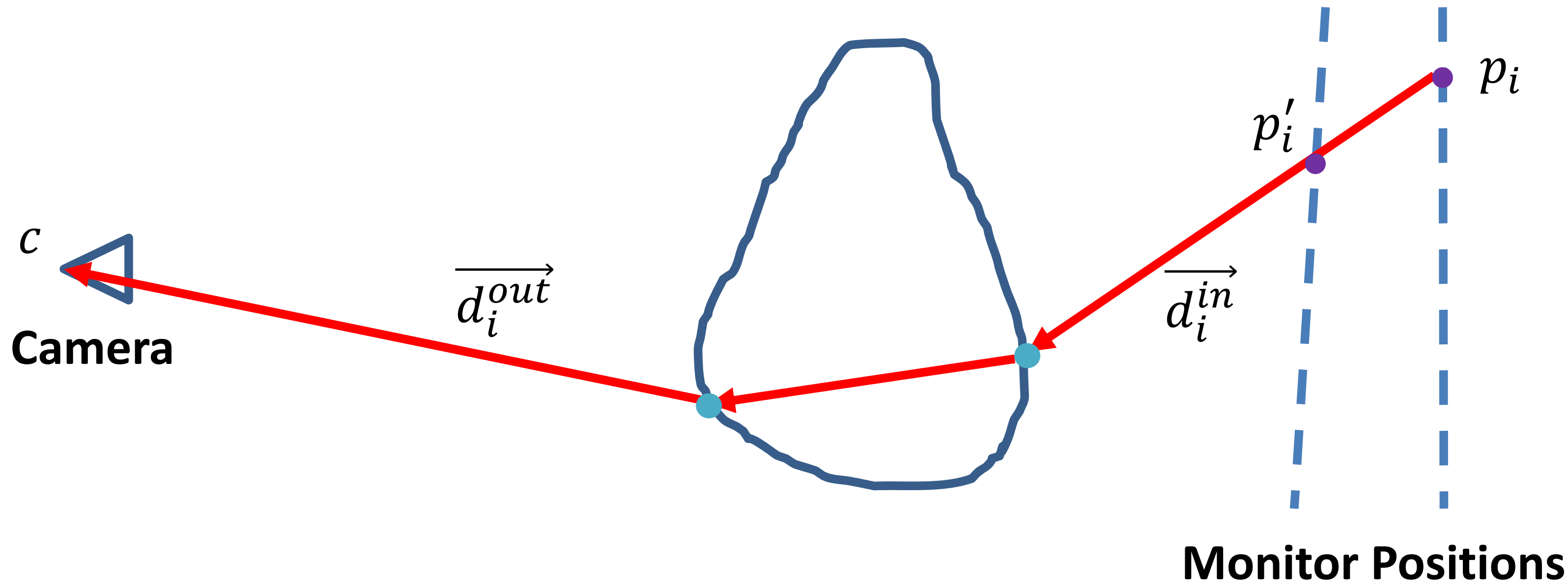
Refraction-based Triangulation

- **Use two cameras and 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



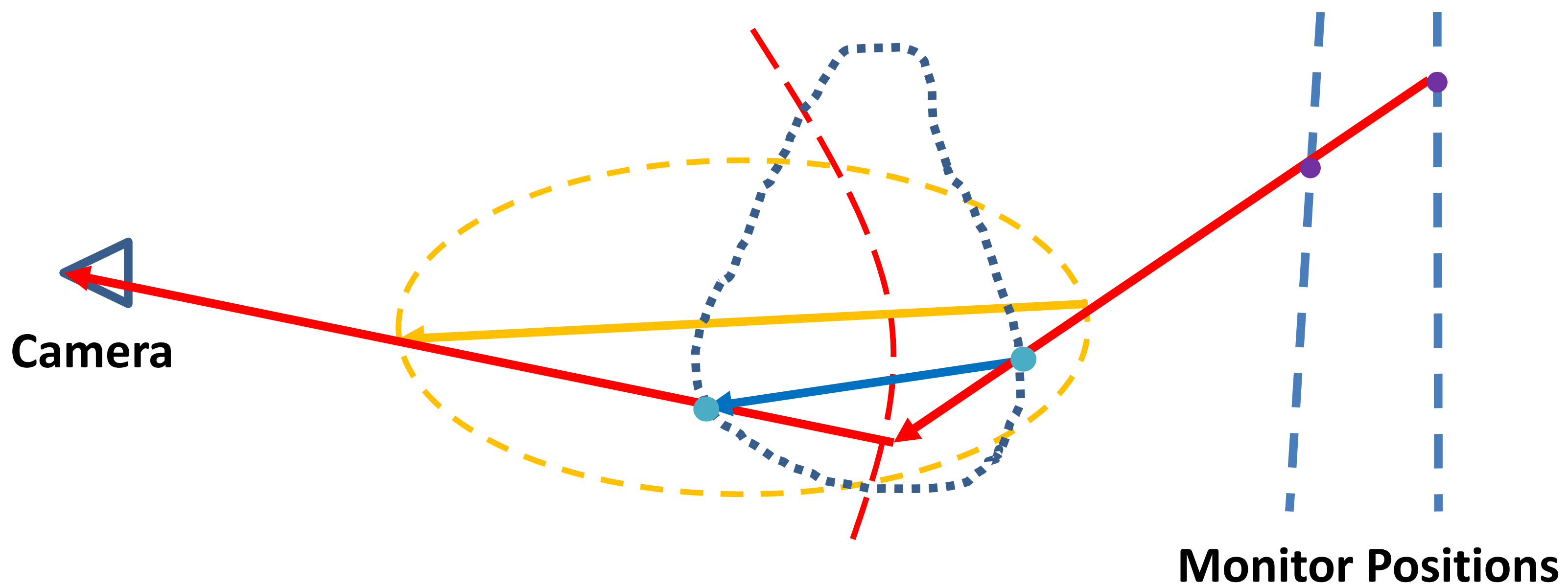
Our Approach

- **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

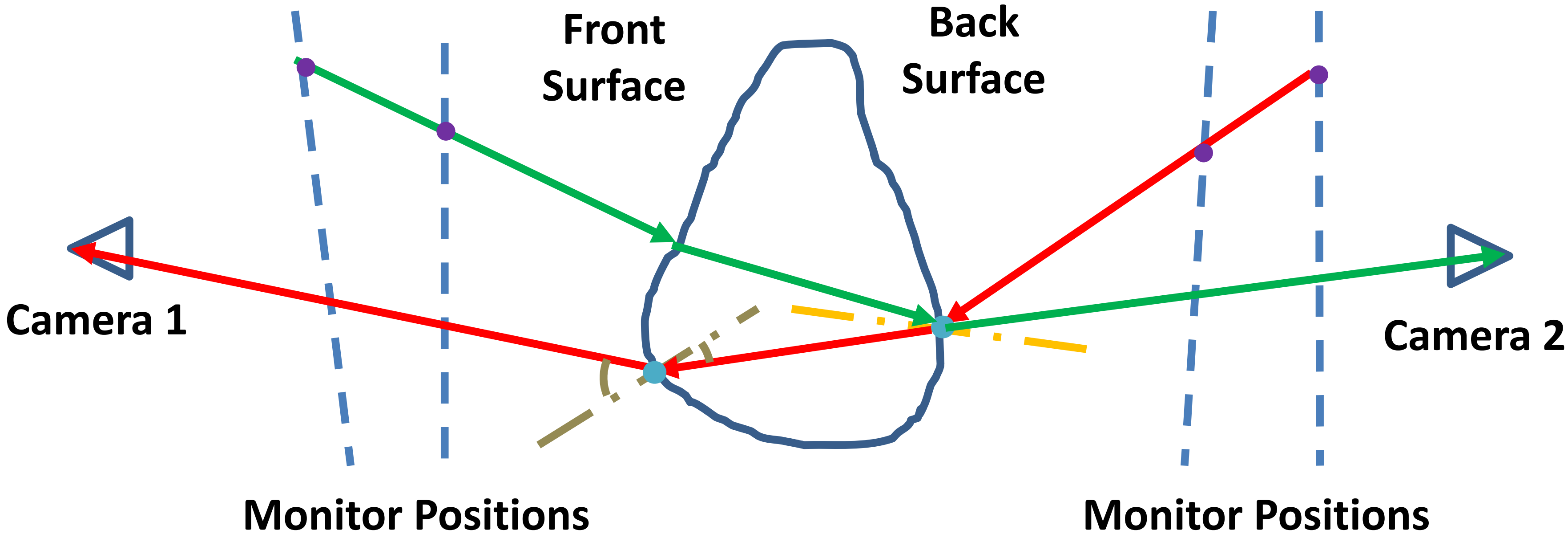


Surface Ambiguities

Position-Normal Consistency (PNC)

- 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 and back surfaces

- **Enforcing PNC at single refraction location does not provide enough constraints**
  - **Capture ray-ray correspondences from both front and 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 and a smoothness term for both front and 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

- 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

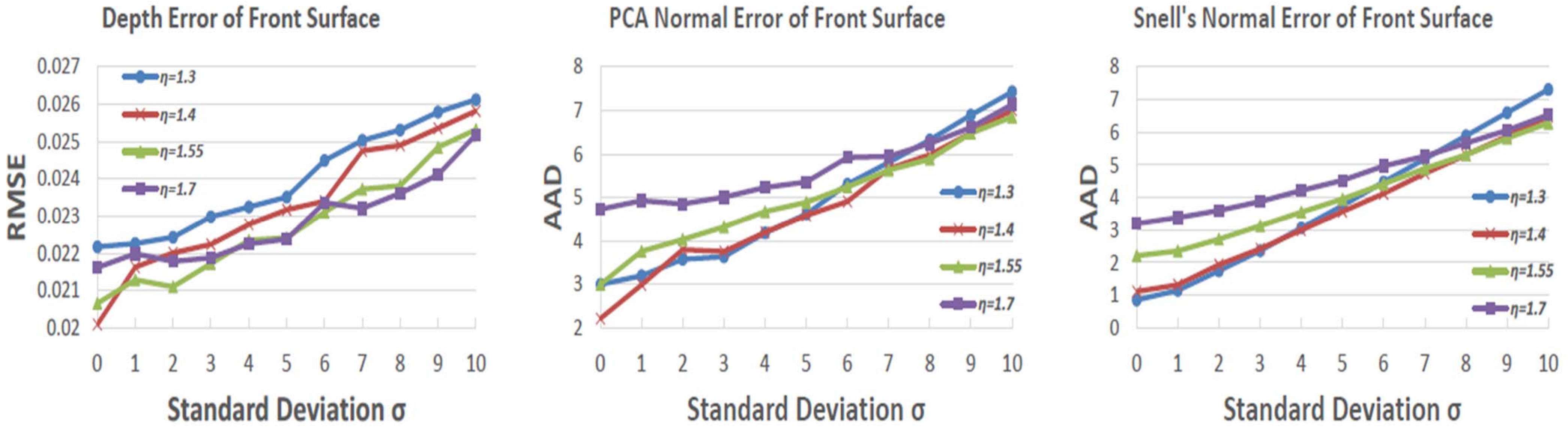
The Two Terms

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

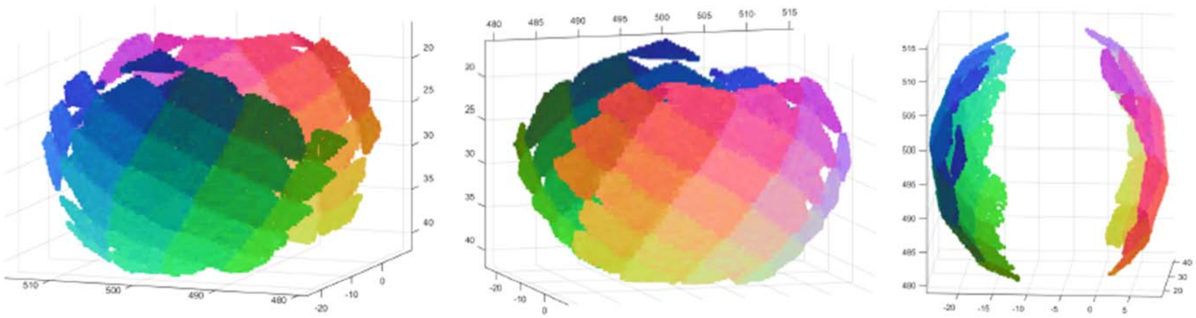
Experiments

- 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 and PCA normal
  - AAD between the true normal and Snell's law normal

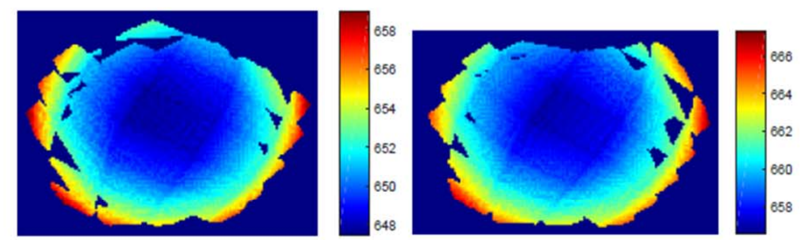
Synthetic Object



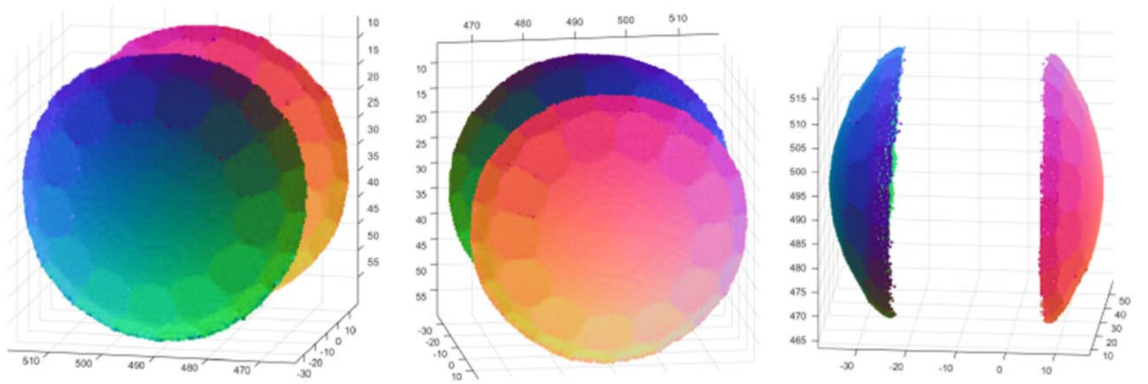
Quantitative Evaluation



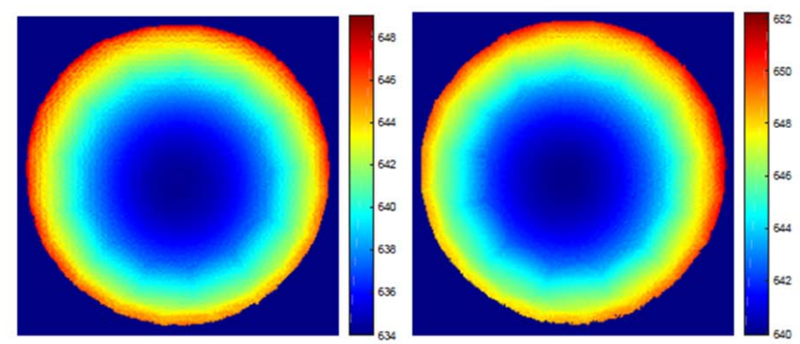
(a) Point cloud of the “ornament” object



(b) Depth maps of the “ornament” object

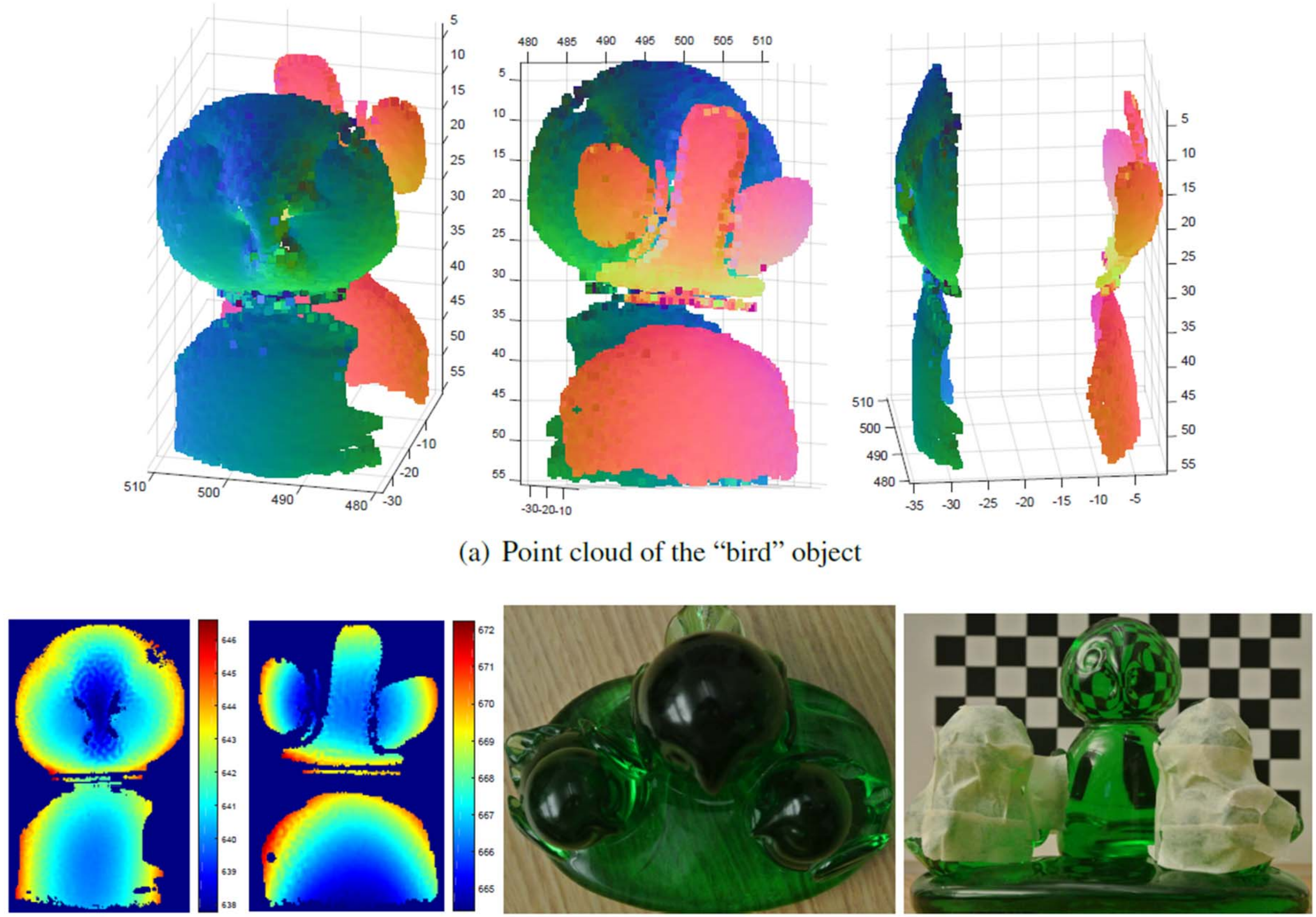


(c) Point cloud of the “ball” object



(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 and normal estimation
  - Refractive index estimation
- **Limitations:**
  - Thousands of images need to be captured
  - Assume homogeneous objects and two refraction events

Summary

Future Work

- **Can we fully capture the appearance of a transparent object and insert it into 3D virtual environment?**
  - **How to capture ray-ray correspondences when there are multiple contribution sources?**
  - **How to capture object from all sides?**
  - **Spatial coherences among contribution sources**
  - **Caustic effect?**

Image-Based Transparent Object

- 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

Questions?

- **Yiming Qian, Minglun Gong, & Yee-Hong Yang: 3D reconstruction of transparent objects with position-normal consistency. IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA, June 27-30, 2016.**

- **Yiming Qian, Minglun Gong, & Yee-Hong Yang: Frequency-based environment matting by compressive sensing. IEEE International Conference on Computer Vision: 3532-3540. Santiago, Chile, December 13-16, 2015.**

Related Publications

- 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 advanced. 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.
- Two recent research projects that advance the state-of-the-art on this front is presented here. The first one investigates how transparent objects interact with the environments using a frequency-based environment matting approach. However, unlike existing approaches 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 approach for reconstructing homogeneous transparent objects. By introducing a novel position-normal consistency constraint, an optimization procedure is designed to jointly reconstruct the 3D positions and normals of transparent surfaces.

Abstract

- Dr. Minglun Gong is a Professor and Head of 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 15 papers 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 and the 2015 Best Paper Award from the Canadian Artificial Intelligence Association.

Biography

Who am I

1994: B.Engr. in Mechanical Engr.

1997: M.Sc. in Computer Sci.

2003: Ph.D. in Computer Sci.

2003~2007: Assist. Prof.

2007~pres: Assist. Prof., Assoc. Prof., Prof., and Dept. Head

2016~pres: Visiting Prof.