Optical Computing for Light Transport Matrix Methods

Anton Schirg

***Abstract*—This paper explains the idea of optical computing on the light transport matrix. It shows how a simple algorithm can be transfered into the optical domain. Besides, it deals with a more complex algorithm for separating direct and indirect light transport, which works by altering the light transport of epipolar and non-epipolar light paths in a camera projector system using high refresh rate masks in front of camera and projector. This allows obtaining images containing only the indirect component of light transport (i.e. light rays that have been reflected multiple times before reaching the camera) by blocking all epipolar light transport. Using a more complex construction of masks, images whose indirect contribution is not dependent on the projected pattern can be captured. This is interesting for 3d reconstruction by structured light images. Finally this paper will discuss a compressed sensing technique for 3d reconstruction of fast moving scenes.**

I. Introduction

For a camera projector system (Figure 1) the light transport matrix contains the contribution of any projector pixel to any camera pixel.

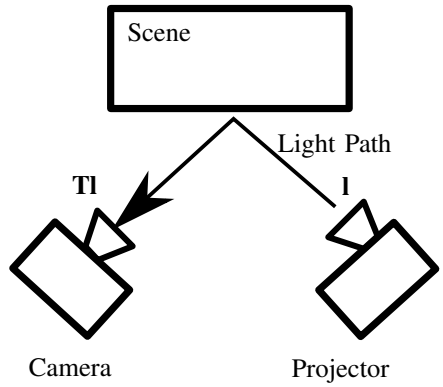

Fig. 1. Camera Projector System

For example, by its definition it can be used to calculate the image $\mathbf{i} = \mathbf{T}\mathbf{l}$ the camera would observe if the scene was illuminated with the projector image $\mathbf{l}$. Of course the light transport matrix has to be known for this computation. As shown further below, by manipulating the light transport matrix direct light transport can be removed and an indirect only image can be calculated.

But in practice most often it is not possible to directly obtain the light transport matrix as it would take a long time and consume a lot of memory. For a camera and projector both having a resolution of 10^6 pixels at 30 frames per second this would take more than 9 hours and consume 1 terabyte of memory assuming 8 bits per pixel.

One can avoid having to obtain the transport matrix by performing calculations in the optical domain. For example, by definition of the light transport matrix, multiplying a vector to the right side of the matrix is equivalent to projecting the corresponding image onto the scene and capturing an image (illuminate and capture operation).

There exist methods using optical calculations to obtain an approximation of the light transport matrix sufficient for many applications [1].

Other algorithms, like the one discussed in this paper, can do all calculations in the optical domain.

This paper is structured as follows: Section 2 demonstrates the conversion of a simple algorithm into the optical domain on the example of power iteration. Section 3.A covers the separation of direct and indirect light transport using high refresh rate masks. In section 3.B this paper will discuss 3D reconstruction as an application of indirect invariant imaging.

II. Optical Power Iteration

Power iteration is a simple algorithm (Alg. 1) to find the largest eigenvector v of a matrix.

Input: Matrix $\mathbf{M}$
Result: largest eigenvector $\mathbf{v}$ of $\mathbf{M}$
$\mathbf{l} =$ random vector;
for *i = 1 to n* **do**
  $\mathbf{l} = \mathbf{M}\mathbf{l}$;
  normalize $\mathbf{l}$;
end
$\mathbf{v} = \mathbf{l}$;
return $\mathbf{v}$;

Algorithm 1: Power iteration

This algorithm can easily be transferred into the optical domain by replacing the matrix multiplication with an illuminate and capture operation (Alg. 2).

Once the largest eigenvector has been obtained the next one can be computed by finding the largest eigenvector of the matrix $\mathbf{M} - \lambda\mathbf{I}$, where $\mathbf{I}$ is the identity matrix and λ the eigenvalue corresponding to the largest eigenvector.

Multiple eigenvectors can be used to retrieve a low rank approximation of the light transport matrix.

But as the convergence properties of power iteration may be arbitrarily bad for ill-formed matrices and power iteration computes only one eigenvector at a time, [1] looks at more efficient algorithms of similar design.

Input: Scene $\mathbf{S}$
Result: largest eigenvector $\mathbf{v}$ of the light transport matrix of $\mathbf{S}$
$\mathbf{l}$ = random vector;
for *i = 1 to n* **do**
  project $\mathbf{l}$ onto $\mathbf{S}$;
  $\mathbf{l}$ = image of $\mathbf{S}$;
  normalize $\mathbf{l}$;
end
$\mathbf{v} = \mathbf{l}$;
return $\mathbf{v}$;

Algorithm 2: Optical power iteration

III. Separation of Direct and Indirect Light

By manipulating the flow of light through a scene one can achieve interesting results. In [2] the authors show how light transport can be manipulated so that only certain light paths contribute to the final image.

A. Operating Principle

For a stereo camera projector system for every point on one image exists an epipolar line on the other image on which the point of direct light transport must lie (Figure 2). Thus light paths can be classified into epipolar and non-epipolar paths. All direct light transport is epipolar, while most of the indirect light transport is non-epipolar.

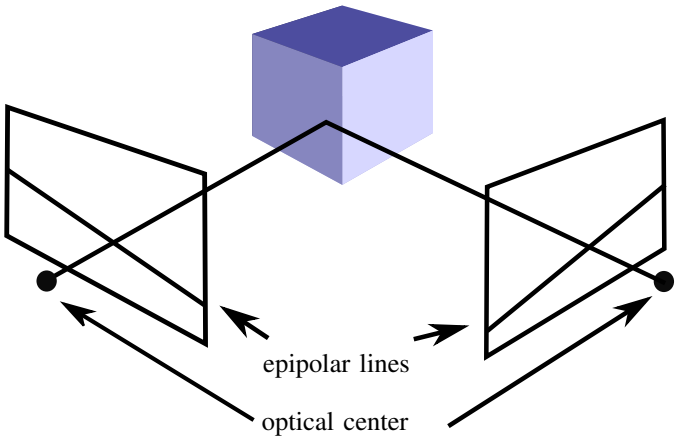

Fig. 2. Epipolar light transport

To manipulate the flow of light two DMD masks (digital micromirror device) operating at a frequency much higher than the camera frame rate are used, mask $\mathbf{q}$ in the projector and mask $\mathbf{m}$ in front of the camera (Figure 3).

DMD masks are used as they can reach much higher refresh rates than LCDs. As a downside they can only represent binary masks which had to be taken into account in the choice of masks.

For the following considerations we imagine the images and masks to be transformed in such a way that epipolar lines are horizontal and as such occupy a continuous section in the vectors (Figure 4).

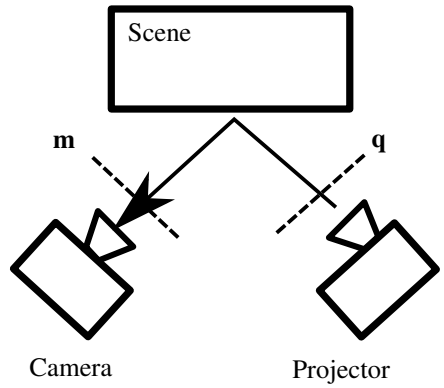

Fig. 3. Camera Projector System with Mask

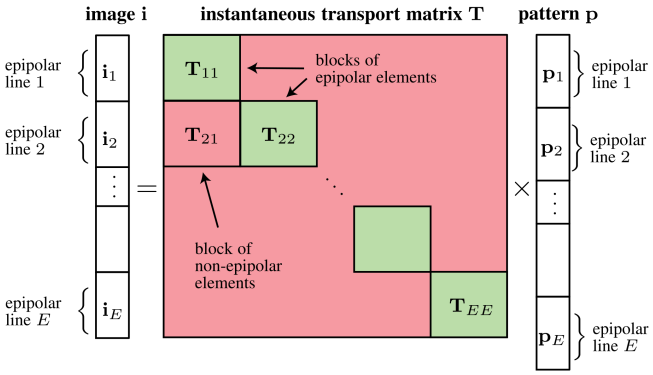

Fig. 4. Transport matrix if epipolar lines are horizontal (From [2])

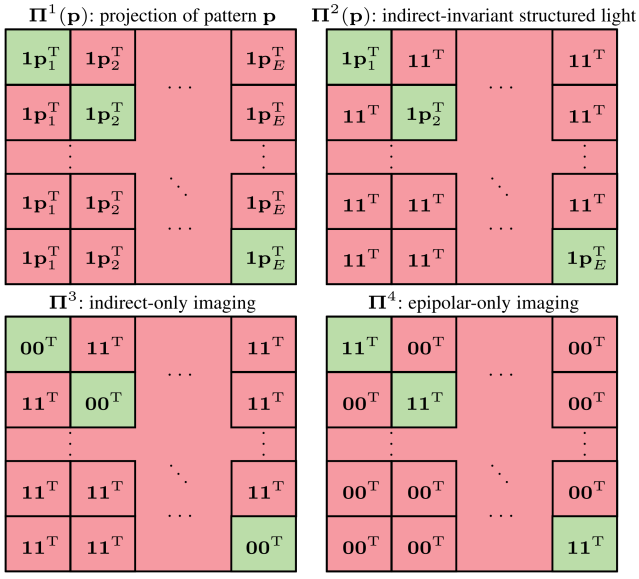

Fig. 5. Probing matrices (From [2])

1) Indirect Only: An indirect only image (Figure 6) can be obtained by removing all epipolar light transport. This will also remove some of the indirect light but the contribution of this light would have been negligible.

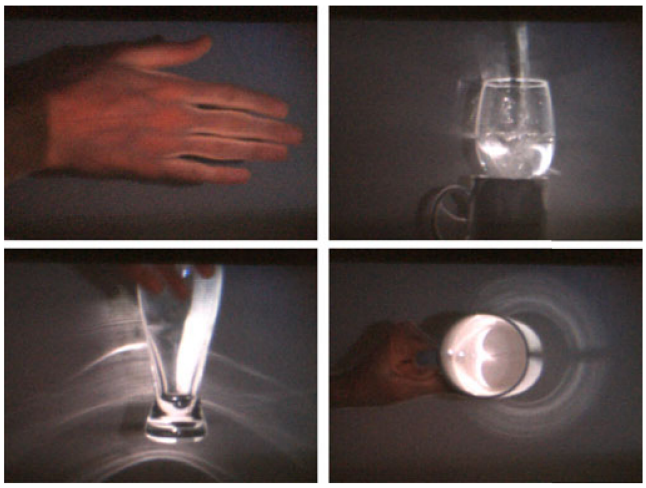

Fig. 6. Indirect only images. Top left: In the image of a hand vein patterns are visible. Other images: Caustics in glasses and a mug are emphasized. (From [2])

Letting through only one epipolar line on the projector and all other lines on the camera will remove all direct light transport.

This can also be thought of as manipulating the light transport matrix by changing the influence of different light paths, i.e. weight single entries in the transport matrix by entrywise product with a so called "probing matrix" Π:

$$\mathbf{i} = (\Pi \circ \mathbf{T})\mathbf{1}\,, \tag{1}$$

where $\mathbf{i}$ is the captured image, $\mathbf{1}$ is a vector of ones and $\circ$ denotes the entrywise product.

The resulting image from the camera projector system is

$$\mathbf{i} = \mathbf{m} \circ (\mathbf{T}\mathbf{q})\,, \tag{2}$$

where $\mathbf{m}$ and $\mathbf{q}$ are the masks in front of camera and projector. This is equal to equation (1) for $\Pi = \mathbf{m}\mathbf{q}^\intercal$.

Such a decomposition of Π is possible only for very specific matrices. But as the masks operate at a higher frequency than the camera we can switch between T masks per frame. The resulting image will be

$$\mathbf{i} = \sum_{t=1}^{T} \mathbf{m}_t \circ (\mathbf{T}\mathbf{q}_t)\,. \tag{3}$$

This is equal to equation (1) for $\Pi = \sum_{t=1}^{T} \mathbf{m}_t\mathbf{q}_t^\intercal$.

For $\Pi = \Pi^3$ (Figure 5 indirect-only) this is satisfied for the case mentioned above where one epipolar line is activated at a time.

However this method has poor light efficiency as only $1/T$ of the projected light is let through.

To improve this a randomized approach can be used: half of all lines are activated in the projector and the rows corresponding to the other half in the camera, i.e. $\mathbf{q}_t$ is a sample of

$$\mathfrak{q} = \{\text{each epipolar line is turned on with probability } 0.5\} \tag{4}$$

and $\mathbf{m}_t = \overline{\mathbf{q}_t}$, where $\overline{\mathbf{q}}$ denotes the entrywise not operation on the binary vector $\mathbf{q}$.

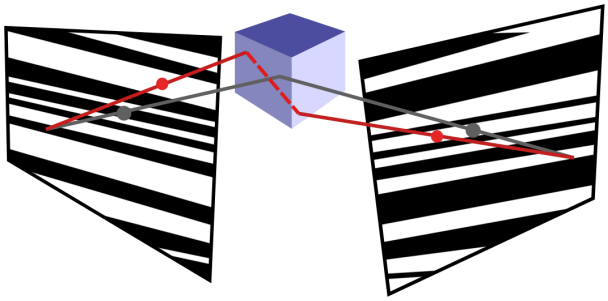

Fig. 7. Randomized indirect-only mask. Direct light transport (gray) is always blocked, while nearly $\frac{1}{4}$ of indirect light (red) is let through.

To further support this approach the authors of [3] use the vectorization scheme of T shown above (Figure 4):

Following (3) a single row of the image is

$$\mathbf{i}_e = \frac{1}{T}\sum_{t=1}^{T}\sum_{f=1}^{E} \overline{\mathbf{q}_{t,e}} \circ (\mathbf{T}_{ef}\mathbf{q}_{t,f})\,. \tag{5}$$

Thanks to the vectorization scheme this can be split into epipolar and non-epipolar terms.

$$\mathbf{i}_e = \frac{1}{T}\sum_{t=1}^{T}\left(\overline{\mathbf{q}_{t,e}} \circ (\mathbf{T}_{ee}\mathbf{q}_{t,e}) + \sum_{f=1,f\neq e}^{E} \overline{\mathbf{q}_{t,e}} \circ (\mathbf{T}_{ef}\mathbf{q}_{t,f})\right) \tag{6}$$

$\mathbf{q}_{t,e}$ represents a single epipolar line so by the definition of $\mathbf{q}_t$ in (4) either $\mathbf{q}_{t,e}$ or $\overline{\mathbf{q}_{t,e}}$ is a zero vector and thus the first term is zero.

$$\mathbf{i}_e = \frac{1}{T}\sum_{t=1}^{T}\sum_{f=1,f\neq e}^{E} \overline{\mathbf{q}_{t,e}} \circ (\mathbf{T}_{ef}\mathbf{q}_{t,f}) \tag{7}$$

For $T \to \infty$ this corresponds to the expected value

$$\mathcal{E}[\mathbf{i}_e] = \mathcal{E}[\sum_{f=1,f\neq e}^{E} \overline{\mathfrak{q}_e} \circ (\mathbf{T}_{ef}\mathfrak{q}_f)]\,. \tag{8}$$

And this by linearity of the expected value and by the fact that $\mathfrak{q}_e$ and $\mathfrak{q}_f$ are independent results in

$$\mathcal{E}[\mathbf{i}_e] = 0.25\sum_{f=1,f\neq e}^{E} \mathbf{T}_{ef}\mathbf{1} \tag{9}$$

$$\mathcal{E}[\mathbf{i}] = 0.25\left(\Pi^3 \circ \mathbf{T}\right)\mathbf{1}\,. \tag{10}$$

So we see that the expected value of the image is equal to the exact method but with a light efficiency of $1/4$.

2) Epipolar Only: For the epipolar only probing matrix no short decomposition exists [2]. Instead one can take a conventional and a non-epipolar image in a row and subtract the two to obtain an epipolar only image.

3) Indirect-Invariant: Instead of an epipolar only image for some applications — like the one explained below — an indirect-invariant image (i.e. where the indirect contribution is not dependent on the projected pattern) is sufficient. This can be obtained from only one frame. Mask $\mathbf{m}$ is again chosen as a sample of

$$\mathfrak{m} = \{\text{each epipolar line is on with probability } 0.5\}\,. \quad (11)$$

Mask $\mathbf{q}$ is chosen as follows:

$$\mathbf{q}_t = \mathbf{m}_t \circ \mathbf{r}_t + \overline{\mathbf{m}_t} \circ \overline{\mathbf{r}_t} \quad (12)$$

where $\mathbf{r}_t$ is a sample of

$$\mathfrak{r} = \{\text{pixel } p \text{ on epipolar line } e \text{ is } 1 \text{ with probability } p_e[p]\} \quad (13)$$

where p_e is the e-th epipolar line in the projected image.

In [3] the authors show that this is sensible similar to indirect only:

Following (3) a single row of the image is given by

$$\mathbf{i}_e = \frac{1}{T}\sum_{t=1}^{T}\sum_{f=1}^{E} \mathbf{m}_{t,e} \circ \left(\mathbf{T}_{ef}\left(\mathbf{m}_{t,f} \circ \mathbf{r}_{t,f} + \overline{\mathbf{m}_{t,f}} \circ \overline{\mathbf{r}_{t,f}}\right)\right)\,. \quad (14)$$

Thanks to the vectorization scheme this can be split into epipolar and non-epipolar terms.

$$\begin{aligned}\mathbf{i}_e = \frac{1}{T}\sum_{t=1}^{T}&(\mathbf{m}_{t,e} \circ (\mathbf{T}_{ee}(\mathbf{m}_{t,f} \circ \mathbf{r}_{t,f})) + \\ &\mathbf{m}_{t,e} \circ (\mathbf{T}_{ee}(\overline{\mathbf{m}_{t,f}} \circ \overline{\mathbf{r}_{t,f}})) + \\ &\sum_{f=1,f\neq e}^{E} \mathbf{m}_{t,e} \circ (\mathbf{T}_{ef}(\mathbf{m}_{t,f} \circ \mathbf{r}_{t,f} + \overline{\mathbf{m}_{t,f}} \circ \overline{\mathbf{r}_{t,f}})))\end{aligned} \quad (15)$$

$\mathbf{m}_{t,e}$ represents a single epipolar line so by the definition of $\mathbf{m}_t$ in (11) either $\mathbf{m}_{t,e}$ or $\overline{\mathbf{m}_{t,e}}$ is a zero vector and thus the second term is zero.

$$\begin{aligned}\mathbf{i}_e = \frac{1}{T}\sum_{t=1}^{T}&(\mathbf{m}_{t,e} \circ (\mathbf{T}_{ee}(\mathbf{m}_{t,f} \circ \mathbf{r}_{t,f})) + \\ &\sum_{f=1,f\neq e}^{E} \mathbf{m}_{t,e} \circ (\mathbf{T}_{ef}(\mathbf{m}_{t,f} \circ \mathbf{r}_{t,f} + \overline{\mathbf{m}_{t,f}} \circ \overline{\mathbf{r}_{t,f}})))\end{aligned} \quad (16)$$

For $T \to \infty$ this corresponds to the expected value

$$\begin{aligned}\mathcal{E}[\mathbf{i}_e] = \mathcal{E}[&\mathfrak{m}_e \circ (\mathbf{T}_{ee}(\mathfrak{m}_e \circ \mathfrak{r}_e)) + \\ &\sum_{f=1,f\neq e}^{E} \mathfrak{m}_e \circ (\mathbf{T}_{ef}(\mathfrak{m}_e \circ \mathfrak{r}_e + \overline{\mathfrak{m}_e} \circ \overline{\mathfrak{r}_e}))]\,.\end{aligned} \quad (17)$$

And this since $\mathfrak{m}_e$ is either $\mathbf{1}$ or $\mathbf{0}$ and by the fact that $\mathfrak{m}_e$ is independent from $\mathfrak{m}_f$, $\mathfrak{r}_e$ and $\mathfrak{r}_f$ and since by its definition (13) $\mathcal{E}[\mathfrak{r}_e] = \mathbf{p}_e$ results in

$$\begin{aligned}\mathcal{E}[\mathbf{i}_e] =&0.5\mathbf{T}_{ee}\mathcal{E}[\mathfrak{r}_e] + \\ &0.5\sum_{f=1,f\neq e}^{E} \mathbf{T}_{ef}\mathcal{E}[\mathfrak{m}_e \circ \mathfrak{r}_e + \overline{\mathfrak{m}_e} \circ \overline{\mathfrak{r}_e}]\,.\end{aligned} \quad (18)$$

$$\begin{aligned}=&0.5\mathbf{T}_{ee}\mathbf{p}_e + \\ &0.5\sum_{f=1,f\neq e}^{E} (\frac{1}{2}\mathbf{T}_{ef}\mathbf{p}_e + \frac{1}{2}\mathbf{T}_{ef}(1-\mathbf{p}_e))\,.\end{aligned} \quad (19)$$

$$\begin{aligned}=&0.5\mathbf{T}_{ee}\mathbf{p}_e + \\ &0.25\sum_{f=1,f\neq e}^{E} \mathbf{T}_{ef}\mathbf{1}\,.\end{aligned} \quad (20)$$

$$\mathcal{E}[\mathbf{i}] = 0.25(\Pi^2(2p) \circ \mathbf{T})\mathbf{1}\,. \quad (21)$$

B. Application - Depth and Albedo Reconstruction

An application of indirect-invariant imaging shown in [2] is depth and albedo reconstruction in scenes with specular reflections as these pose a problem when using conventional imaging (Figure 8). One can see that the reconstruction is greatly improved when using indirect-invariant imaging (Figure 9).

conventional imaging (1 of 9) indirect-invariant imaging (1 of 9)

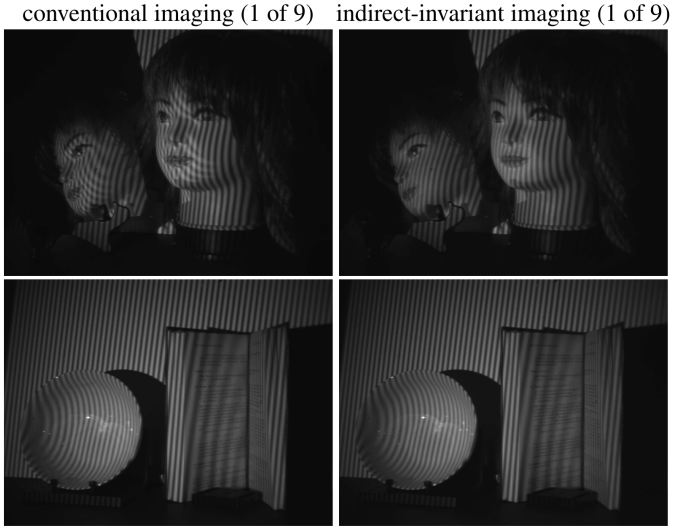

Fig. 8. Comparison of conventional and indirect-invariant imaging. Note that in the indirect-invariant image the projected pattern is not reflected from the mirror. (From [2])

The working principle of the reconstruction is described in [5, Section 4.1]. Multiple sine waves oriented along the epipolar lines are projected onto the scene. Using the arctangent of the ratio of these waves at a specific location the phase can be reconstructed and from it the depth at this location. As the phase wraps at $-\pi/\pi$ multiple frequencies are projected to obtain absolute phase. In this case 3 frequencies and 3 phase offsets were used. This results in 9 images total.

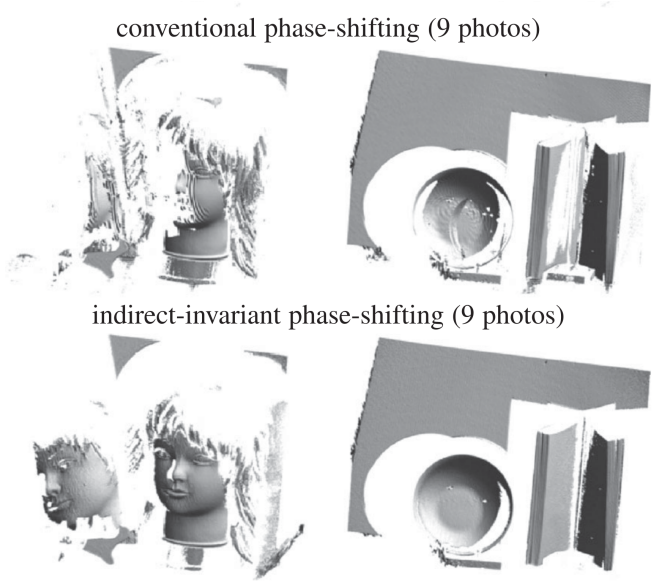

Fig. 9. Comparison of the 3D reconstruction result using conventional and indirect-invariant imaging. (From [2])

For fast moving scenes taking multiple images per reconstruction is not possible. Instead compressed sensing techniques can be used to interleave S structured light patterns into a single image [4]. In this case $S = 6$ (2 frequencies, 3 phase offsets).

The camera pixels are partitioned into S sets, $\mathbf{b}_s$ are vectors indicating the set membership of each pixel. These membership vectors are used as masks cycled though in front of the camera synchronously to switching the projected pattern.

This results in each pixel containing only contributions from one of the projected patterns. The missing pixels are restored by solving the linear system

$$\|\sum_{s=1}^{S} \mathbf{b}_s \circ \mathbf{i}_s - \mathbf{i}\| \leq \epsilon \tag{22}$$

for $\mathbf{i}_s$ for $s = 1 \dots S$. This system is naturally underdefined as one image is used to obtain S images.

To be able to solve the system anyway, spacial coherence is exploited by using JPEG2000 wavelets as basis $\mathbf{W}$ and finding a sparse solution by minimizing

$$\|\mathbf{W}^\intercal[\mathbf{i}_1 \dots \mathbf{i}_S]\| \,. \tag{23}$$

using a non-linear optimizer [6].

IV. Summary and Outlook

We have seen the idea of optical computing on the light transport matrix and how a simple algorithm can be transfered into the optical domain. As a more complex algorithm we looked at separation of direct and indirect light. A key observation is that indirect light transport is dominated by non-epipolar paths and direct light transport contains only epipolar paths. This allows obtaining indirect-only images by blocking all epipolar light transport. Interestingly direct-only imaging can not be achieved in a similar fashion. But using a more complex construction of masks one can produce images whose indirect contribution is not dependent on the projected pattern, which is interesting for 3d reconstruction by structured light images. Finally we saw how compressed sensing techniques can be used to take multiple of these structured light images in a single shot which is useful for 3d reconstruction of fast moving scenes.

Another interesting topic that follows on from the above might be how the compressed sensing for one shot 3d reconstruction can be optimized as it requires heavy post-processing. Maybe new sensor technologies like multi-bucket pixels [7] can remove the need for post processing.

References

[1] Matthew O'Toole and Kiriakos N Kutulakos. Optical computing for fast light transport analysis. *ACM Trans. Graph.*, 29(6):164–1, 2010.

[2] Matthew O'Toole, John Mather, and Kiriakos N Kutulakos. 3d shape and indirect appearance by structured light transport. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3246–3253, 2014.

[3] Matthew O'Toole, John Mather, and Kiriakos N Kutulakos. Supplemental material to 3d shape and indirect appearance by structured light transport, 2014. https://web.stanford.edu/%7Emotoole/slt.

[4] Dikpal Reddy, Ashok Veeraraghavan, and Rama Chellappa. P2c2: Programmable pixel compressive camera for high speed imaging. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 329–336. IEEE, 2011.

[5] Joaquim Salvi, Sergio Fernandez, Tomislav Pribanic, and Xavier Llado. A state of the art in structured light patterns for surface profilometry. *Pattern recognition*, 43(8):2666–2680, 2010.

[6] E. van den Berg and M. P. Friedlander. SPGL1: A solver for large-scale sparse reconstruction, June 2007. http://www.cs.ubc.ca/labs/scl/spgl1.

[7] Gordon Wan, Xiangli Li, Gennadiy Agranov, Marc Levoy, and Mark Horowitz. Cmos image sensors with multi-bucket pixels for computational photography. *IEEE Journal of Solid-State Circuits*, 47(4):1031–1042, 2012.