High Resolution Diffuse Optical Tomography using Short Range Indirect Subsurface Imaging

Chao Liu[1] Akash K. Maity[2] Artur W. Dubrawski[1] Ashutosh Sabharwal[2] Srinivasa G. Narasimhan[1]
[1] Carnegie Mellon University [2] Rice University

Abstract—Diffuse optical tomography (DOT) is an approach to recover subsurface structures beneath the skin by measuring light propagation beneath the surface. The method is based on optimizing the difference between the images collected and a forward model that accurately represents diffuse photon propagation within a heterogeneous scattering medium. However, to date, most works have used a few source-detector pairs and recover the medium at only a very low resolution. And increasing the resolution requires prohibitive computations/storage. In this work, we present a fast imaging and algorithm for high resolution diffuse optical tomography with a line imaging and illumination system. Key to our approach is a convolution approximation of the forward heterogeneous scattering model that can be inverted to produce deeper than ever before structured beneath the surface.We show that our proposed method can detect reasonably accurate boundaries and relative depth of heterogeneous structures up to a depth of $8\ mm$ below highly scattering medium such as milk. This work can extend the potential of DOT to recover more intricate structures (vessels,tissue, tumors, etc.) beneath the skin for diagnosing many dermatological and cardio-vascular conditions.

1 Introduction

Imaging below the skin and through tissue is important for diagnosis of several dermatological and cardiovascular conditions. MRI remains the best current approach to obtain a 3D dimensional visualization below the skin. But MRI is expensive, requires visits to a hospital or imaging center, and the patients are highly inconvenienced. Non-invasive imaging using visible or near-infrared light has the potential to make devices portable, safe, and convenient to use at home or at point-of-care centers.

While light penetrates deep through tissue, it scatters continuously resulting in poor image contrast. This makes it challenging to recover useful properties about the anatomical structures below the skin. Further, the anatomical structures include a complex heterogeneous distribution of tissue, vasculature, tumors (benign or malignant) that vary in optical properties (density, scattering, absorption) and depths below the skin. This makes the modeling of light propagation below skin challenging.

Fortunately, under the highly scattering regime, the photons can be assumed to be traveling diffusely in the medium and can be described as a random walk. This has enabled accurate forward models under diffuse photon propagation. In order to improve contrast, imaging detectors and sources are placed at different locations on the skin. This arrangement captures only indirectly scattered light while eliminating the dominant direct reflection and backscatter [1]. The approaches that attempt to invert the diffusion model with such indirect light imaging systems are commonly classified as "Diffuse Optical Tomography" (DOT). Due to their portability and ease of use, DOT is becoming an attractive choice over traditional modalities like MRI for cerebral as well as hemodynamic imaging [26], [35]. More recently, DOT has been shown to be a promising tool in detecting strokes [9], breast cancer [19], and thyroid imaging [17].

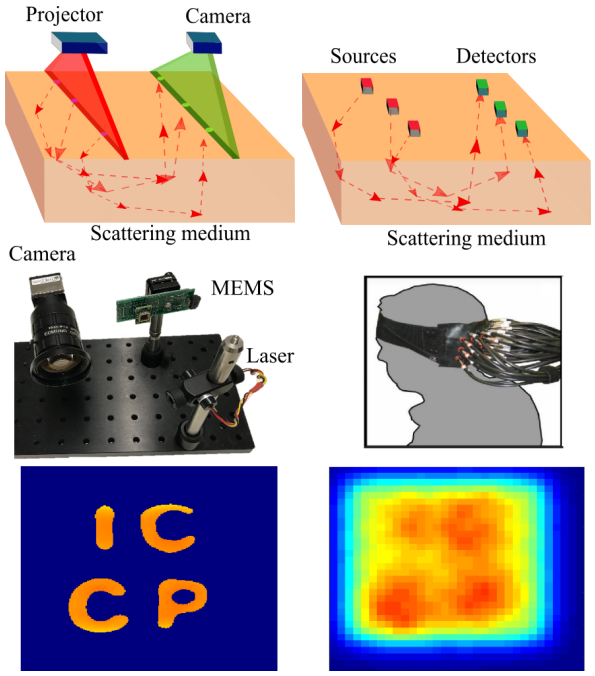

Fig. 1. Diffuse optical tomography (DOT) with line-scanned camera and line scanning MEMS projector (left) compared with traditional DOT [1] with 25 source-detector pairs (right). Both arrangements capture short range indirect subsurface scattered light but our approach is more efficient and recovers the medium (bottom row) at much higher resolution. The results shown here are from simulated data. Please refer to Section 6 for more details and real results.

But there are two important drawbacks to existing DOT approaches. First, the number of source-detector pairs limits the form-factor of devices built so far. Even with multiple source-detector pairs, applying traditional inverse methods for DOT

1. Analogously, in vision and graphics, works measure the Bi-directional Sub-surface Scattering Reflectance Distribution Function (BSSRDF) [12], [13], [16], [22], [34]

results in poor resolution, as shown in the second column of Figure 1. Second, as the number of source-detector pairs increases, the computational complexity of the algorithms that recover the volumetric medium increases prohibitively [7]. In this work, we present an imaging and algorithmic approach to resolve these fundamental drawbacks in DOT systems. Instead of separate detector-source pairs, we use a high resolution 2D camera and a MEMS projector to obtain a large number of effective source-detector pairs, as is commonly done in vision and graphics. This makes the DOT system much more compact and programmable. Second, to increase the speed of acquisition, we illuminate a line on the skin and capture a different line in the sensor, as described in [24]. This arrangement captures short-range indirect light transport much faster than point-to-point illumination and sensing. But [24] uses a rectified configuration where the projector and camera planes are parallel [28], leading to a low spatial resolution over a large overlapping stereo volume. We develop a new design with a verged configuration that enables high spatial resolution imaging within a small region on the skin (approximately 8 cm x 8 cm).

Using this verged design, we develop an efficient algorithm that is based on the convolution approximation to the forward model for light propagation in a heterogeneous subsurface medium. We show that the convolution kernel is independent of the heterogeneous structure and only depends on the imaging configuration and the scattering properties of the homogeneous scattering medium. This allows us to recover the heterogeneous structures at much higher spatial resolution compared with the traditional DOT, as shown in the last row of Figure 1.

We evaluate our imaging and algorithmic approaches on simulated data by borrowing optical parameters of skin, tissue, blood and vasculature from bio-optical literature [25]. We then demonstrate our approach with an imaging setup that consists of a high resolution 2D camera and a custom-built MEMS based laser projector that are verged to capture near-microscopic spatial resolution beneath a small area of the surface. This imaging system is used to recover heterogeneous structures immersed within a highly scattering medium, such as milk. We show that with the proposed hardware and algorithm, we can detect reasonably accurate boundaries of structures up to a depth of $8mm$ below the surface of milk. We believe this work represents strong progress in achieving high-spatial resolution diffuse optical tomography for the first time at subsurface depths of several millimeters in highly scattering media.

2 Related Works

Over the past few years, there have been developments mainly on two aspects of DOT - improving the instrument system design [32], [33], [38], and secondly, on theoretical aspects that involves accurate forward modeling and rendering [2], [4]. The traditional DOT setup consists of illumination sources and detectors which are placed on the tissue or skin surface. To improve the reconstruction of optical parameters of the tissue volume, multiple configurations of DOT have been explored. In general, it is not possible to obtain depth information from a single source-detector system and therefore multiple source-detector configurations are needed [31].

More recent systems utilize the time-domain (TD) information of photon propagation. TD-DOT systems consist of a source emitting a pulse of light and fast-gated detectors that capture the time-profile of photon arrival. The detectors and the sources are located on a probe with fixed strategic distances between them, so that the photons have traversed a certain depth and rejecting early arriving photons [29]. The most important drawback in these systems is that the limited number of detectors constrains the spatial resolution of the reconstructed optical parameters. More recent DOT systems use structured illumination, which involves projection of patterns instead of discrete sources [3], [18], [23], [30]. The light after interaction with the tissue is captured by either a single-pixel detector or a 2D CCD camera. The use of structured illumination addresses the issues of low speed and sparse spatial sampling, which are associated with traditional DOT systems [30].

The reconstruction of optical parameters involves fitting a forward model to the acquired measurements. The forward model can be obtained either from mesh-based Monte-Carlo simulations [15], [37] or from a diffusion approximation [5], [6], [14] derived from radiative-transfer equations (RTE). While the Monte-Carlo based forward model is more accurate, it requires hours of computing time. Under Born or Rytov approximations, the forward model relates the optical parameters and the measurements by a linear set of equations [27]. Generally, the number of optical parameters to be reconstructed per voxel is very large compared to the number of measurements, and therefore the inverse problem is severely ill-posed. Tikhonov regularizer or sparsity-inducing regularizers are commonly applied for solving the inverse problem [10], [20]. However, with dense sampling the computational load increases as the reconstruction process involves inversion of a large-scale Jacobian matrix [36].

3 Forward Model

The radiative transfer equation (RTE) [8], [21] describes the light radiance, which models light propagation, at a particular position in the scattering medium at a specific time instant. It is generally difficult to solve the RTE in closed form. When the medium is highly scattering, as in the case of biological tissue, the diffusion approximation is commonly applied to obtain the diffusion equation [14], [22]. The photon diffusion equation models the fluence rate Φ, that is defined as the total light radiance integrated over all directions, at a position $\mathbf{r}$ in the scattering medium for a continuous intensity light source S, given as,

$$\left(D(\mathbf{r})\nabla^2 + \mu_a(\mathbf{r}) \right) \Phi(\mathbf{r}) = S(\mathbf{r}); \tag{1}$$

where $\mu_a(\mathbf{r})$; $\mu_s'(\mathbf{r})$ are the absorption coefficient and the reduced scattering coefficient of the medium respectively, and $D(\mathbf{r}) = 1/(3(\mu_s'(\mathbf{r}) + \mu_a(\mathbf{r})))$ is known as the diffusion coefficient of the scattering medium. The tissue is commonly modeled as a semi-infinite scattering homogeneous medium, with the source and the detector positioned at the air-medium interface. When the light source is treated as a constant pencil beam source, i.e. $S(\mathbf{r}) = S(\mathbf{r}_s)$, the solution for fluence rate in (1) for the configuration in Figure 2 can be written in a analytical form using extrapolated zero boundary conditions (EBC):

$$\Phi_0(\mathbf{r}_d;\mathbf{r}_s) = \frac{S}{4\pi D_0}\left[\frac{e^{-\mu_{tr} r_1}}{r_1} - \frac{e^{-\mu_{tr} r_2}}{r_2}\right]; \tag{2}$$

where, $\Phi_0(\mathbf{r}_d;\mathbf{r}_s)$ is the fluence rate at detector kept at a position $\mathbf{r}_d$ with a constant source at $\mathbf{r}_s$ [5]. The diffusion coefficient of the homogeneous medium is denoted by D_0 and the term $\mu_{tr} = \sqrt{3\mu_s'\mu_a}$ depend on the optical properties of the homogeneous scattering medium. The extrapolated boundary condition (EBC) accounts for the refractive index mismatch of air and the medium.

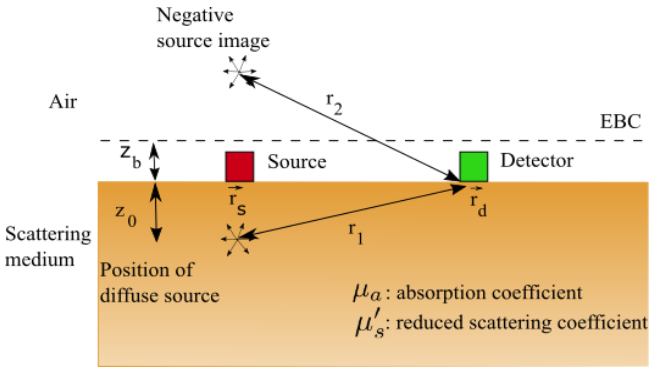

Fig. 2. Source-detector configuration in typical DOT system. The fluence rate at the detector is given by superposition of the real diffuse source located z_0 below the surface, and a negative image source around the zero flux line denoted by EBC.

Solving for the boundary condition defines a zero fluence rate line located z_b above the air-medium interface. This boundary line is imposed by placing a negative image of the source around the zero-crossing line [5]. The terms r_1 and r_2 are the distances from the detector to the real and the negative image source respectively, and they are defined as:

$$\begin{aligned} r_1 &= j\mathbf{r}_s + z_0 \quad \mathbf{r}_d j; \\ r_2 &= j\mathbf{r}_s \quad z_0 \quad 2z_b \quad \mathbf{r}_d j; \end{aligned} \tag{3}$$

where, $z_0 = 3D$ is the location of diffused source in the medium. The term z_b is the distance of the zero fluence rate boundary from the air-medium interface.

Often, we are interested in reconstructing objects like veins or tumors embedded within human tissue. Typically these objects have different optical parameters than the background medium. In the presence of heterogeneity, the change in absorption coefficient of the medium can be defined as,

$$_a(\mathbf{r}) = \quad _{a_0} + \quad _a(\mathbf{r}) \tag{4}$$

where $_a(\mathbf{r})$ is the difference in absorption coefficient of the heterogeneous object over the background medium. We assume that the change in the reduced scattering coefficient $^{0}_{s}(\mathbf{r})$ is negligible and can be ignored. The resultant fluence rate at the detector position $\mathbf{r}_d$ for a source at $\mathbf{r}_s$ is written as a linear addition of fluence rate from homogeneous component $_0(\mathbf{r}_d;\mathbf{r}_s)$ and the change in fluence rate $(\mathbf{r}_d;\mathbf{r}_s)$ due to the heterogeneous object,

$$(\mathbf{r}_d;\mathbf{r}_s) = \quad _0(\mathbf{r}_d;\mathbf{r}_s) + \quad (\mathbf{r}_d;\mathbf{r}_s): \tag{5}$$

When the change in the absorption coefficient due to the heterogeneity $_a$ is small, the change in fluence rate due to the heterogeneous object, also known as the perturbation component, is minimal compared to the homogeneous component and can be ignored using the first Born approximation. The perturbation term is then given by [5],

$$(\mathbf{r}_d;\mathbf{r}_s) = \int \quad _0(\mathbf{r}_s;\mathbf{r}_j)\frac{_a(\mathbf{r}_j)}{D_0}G_0(\mathbf{r}_j;\mathbf{r}_d)d\mathbf{r}_j; \tag{6}$$

where G_0 represents the Green's function for a homogeneous slab and is related to $_0$ in (2) as $G_0 = D_0 \quad _0$=S.

We acquire images using a CCD camera, which records the radiant exitance at the surface. The radiant exitance is proportional to the diffuse reflectance R, which is the projection of current density along the surface normal of the detector for a unit-power source,

$$R(\mathbf{r}_d;\mathbf{r}_s) = D_0 \quad \overline{z_d} \quad _{z_d=0}; \tag{7}$$

where z_d is the z component of the detector location $\mathbf{r}_d$.

Applying a derivative to (5) with respect to z_d and multiplying by D_0, we obtain,

$$R(\mathbf{r}_d;\mathbf{r}_s) = R_0(\mathbf{r}_d;\mathbf{r}_s) + \quad R(\mathbf{r}_d;\mathbf{r}_s); \tag{8}$$

where $R_0 = D_0[\quad _0 = z_d]_{z_d=0}$ is the diffuse reflectance due to the homogeneous background medium and is obtained by taking a derivative of $_0$ in (2) with respect to z_d, given by [22],

$$R_0 = \frac{1}{4}\left[\frac{z_0(1+\quad r_1)e^{\quad r_1}}{r_1^3} + \frac{(z_0+2z_b)(1+\quad r_2)e^{\quad r_2}}{r_2^3}\right] \tag{9}$$

Similarly, R represents the change in diffuse reflectance for the heterogeneous object. The expression for R is obtained by taking a derivative of (6) with respect to z_d and multiplying by D_0, resulting in the following expression,

$$\begin{aligned} R(\mathbf{r}_d;\mathbf{r}_s) &= \int \quad _0(\mathbf{r}_s;\mathbf{r}_j) \quad _a(\mathbf{r}_j) \quad \frac{G_0(\mathbf{r}_j;\mathbf{r}_d)}{z_d} \quad _{z_d=0} d\mathbf{r}_j; \\ &= \int \quad _0(\mathbf{r}_s;\mathbf{r}_j) \quad _a(\mathbf{r}_j)R_0(\mathbf{r}_j;\mathbf{r}_d)d\mathbf{r}_j: \end{aligned} \tag{10}$$

We discretize the integral above by dividing the medium into N voxels, and the optical properties are defined for each voxel. If the medium is discretized into N voxels with volume of each voxel as h^3, then (10) can be written in the discrete summation form given by

$$R(\mathbf{r}_d;\mathbf{r}_s) = \quad \sum_{j=1}^{N} P(\mathbf{r}_s;\mathbf{r}_j;\mathbf{r}_d) \quad _a(\mathbf{r}_j); \tag{11}$$

with

$$P(\mathbf{r}_d;\mathbf{r}_j;\mathbf{r}_s) = \quad _0(\mathbf{r}_s;\mathbf{r}_j)R_0(\mathbf{r}_j;\mathbf{r}_d)h^3: \tag{12}$$

The term $P(\mathbf{r}_s;\mathbf{r}_j;\mathbf{r}_d)$ is commonly known as the phase function defined at each voxel position $\mathbf{r}_j$ in the medium. The values of the phase function depend on the optical properties of the background homogeneous medium as well as the location of the source $\mathbf{r}_s$ and the detector $\mathbf{r}_d$. Note that the phase function is independent from the structure of the heterogeneous object.

4 Convolution Approximation of Heterogeneous Model

In this section, we describe how the diffuse forward model can be adapted to our experimental setup. We project a line illumination on the scene using a laser projector as in [24]. So the light source is now considered as a slit source instead of a point source. By using a slit source we reduce the acquisition time since line scanning is significantly faster than point scanning. We incorporate a slit source in the forward model by using the linear superposition principle. The quantities described in the previous section which are functions of the source location $\mathbf{r}_s$ are now obtained by adding up the contributions corresponding to all the point sources on the illumination line.

On the detector side, we use a rolling shutter camera synchronised with the laser projector. The advantage of using a camera is that each pixel in the camera sensor can now be considered

Fig. 4. Visualization of phase function for different pixel to illumination line distance in y-z plane (top row), and x-y plane (bottom row). S and D represents the illumination line and pixel location respectively. As the pixel to illumination line distance increases, the photons tend to travel deeper into the scattering medium but leads to reduced number of photons reaching the pixel, thereby reducing the signal-to-noise ratio.

Fig. 3. Generation of short range indirect images for a small (a) and a large (b) pixel to illumination distance Δy. The simulated scene consists of three cylinders embedded in a scattering medium. The offset Δy is kept constant while scanning the entire scene to obtain an image. For a shorter Δy as in (a), the rods are distinctly visible, whereas for longer Δy, the blurring increases with reduction of signal-to-noise ratio.

as an individual detector, and hence we have a detector array with millions of detectors. Secondly, since the camera sensor can be considered as a grid array of detectors, we can derive a convolution form of the forward model, significantly speeding up the computation time. We acquire several images by varying the pixel to illumination line distance shown in Figure 3. These images are referred to as short-range indirect images. The boundaries of the heterogeneous structures become more blurry in the short range indirect image as the pixel to illumination line distance Δy increases. The blurring effect is related to Δy and the depth of the structures. This is similar to the depth from (de)focus methods, where the blurring effect is related to the focal setting of the camera and the scene depth.

The values of phase function at each voxel for the short-range indirect images can be interpreted as the number of photons that have traversed through the corresponding voxel for a given pixel to illumination line distance. Typically, the most probable path between a pixel and the source illumination line follows a well-known "banana shape" [11] and is shown for different pixel to illumination line distances in the Figure 4.

We note two important properties of the phase function P. Firstly, in case of simple geometry like the semi-infinite homogeneous background medium under consideration, the expression for the Green's function G_0 as well as σ_0 can be written in terms of relative voxel location rather than the absolute location, i.e,

$$\begin{aligned} P(\vec{r}_d; \vec{r}_j; \vec{r}_s) &= \sigma_0(\vec{r}_s - \vec{r}_j) R_0(\vec{r}_j - \vec{r}_d); \\ &= P(\vec{r}_j - \vec{r}_d; \vec{r}_s - \vec{r}_d). \end{aligned} \tag{13}$$

Secondly, we note that the values of the phase function decays fast spatially as the distance between a voxel and source or detector position increases. Hence, we can neglect the contribution of voxels that are far away from both the illumination line and the pixel. Since we are using a projected line illumination as our light source, we approximate the phase function in (13) by the summation of truncated phase function for each source point along the illumination line. Additionally, as evident from the figure, the contribution of light from the illumination line to a center pixel is dominant only near the center of the illumination line, and hence we can use a spatially-invariant phase kernel. We define the pixel to illumination line distance $\Delta y = y_s - y_d$, where y_s and y_d are the y component of illumination row $\vec{r}_s$ and the pixel location $\vec{r}_d$ respectively. The phase kernel for a line illumination can then be written as,

$$\Phi(\vec{r}_j - \vec{r}_d; \Delta y) = \sum_{\vec{r}_s} P(\vec{r}_j - \vec{r}_d; \vec{r}_s - \vec{r}_d); \tag{14}$$

where the summation over $\vec{r}_s$ is for all the point sources lying on the illumination line. In the following, we will denote the phase kernel as $\Phi(\Delta y)$ for denotation simplicity unless the spatial dependency needs to be emphasized.

Similarly, the diffuse reflectance $R(\vec{r}_d; \vec{r}_s)$, the change in diffuse reflectance $\Delta R(\vec{r}_d; \vec{r}_s)$ and the homogeneous diffuse reflectance $R_0(\vec{r}_d; \vec{r}_s)$ in (8) are modified for line illumination as the sum of contribution from all point sources lying on the illumination line, and are defined as $R(\vec{r}_d; \Delta y)$, $\Delta R(\vec{r}_d; \Delta y)$ and $R_0(\vec{r}_d; \Delta y)$ respectively. We denote $(x_d; y_d)$ as the surface coordinates of the pixel location $\vec{r}_d$ as shown in Figure 3. If the change in absorption coefficient $\Delta\sigma_a(\vec{r}_j)$ in (11) is represented by a 3D volume Q, then the change in diffuse reflectance ΔR in (11) can now be expressed in a convolution notation as

$$\begin{aligned} \Delta R(x_d; y_d; \Delta y) &= \sum_{\vec{r}_s} \sum_{j=1}^{N} P(\vec{r}_j - \vec{r}_d; \vec{r}_s - \vec{r}_d) \, \Delta\sigma_a(\vec{r}_j) \\ &= \sum_{j=1}^{N} \Phi(\vec{r}_j - \vec{r}_d; \Delta y) \, \Delta\sigma_a(\vec{r}_j) \end{aligned} \tag{15}$$

where $\Delta R \in R^{M \times N}$ is defined over a sensor array of dimension $M \times N$ and corresponds to each pixel to illumination line distance Δy as shown in Figure 3. By representing the change of absorption coefficient $\Delta\sigma_a$ by a 3D volume Q, we can rewrite (15) as the sum of a 3D convolution results:

$$\Delta R(x_d; y_d; \Delta y) = \sum_{z} \Phi(\Delta y) \, Q(x_d; y_d; z) \tag{16}$$

The change in absorption coefficient in the 3D volume is denoted by $Q \in R^{M \times N \times D}$, where D is the depth resolution. The 3D truncated kernel $\in R^{m \times n \times D}$ is the defined for each y, and has the same depth resolution as that of the 3D volume Q. Using (8), the resultant diffuse reflectance acquired at each pixel to illumination line distance y can be written as a linear summation of the contribution from homogeneous background medium R_0 and the perturbation term due to presence of heterogeneity R_y

$$R(x_d; y_d; \quad y) = R_0(x_d; y_d; \quad y) \quad \sum_z \quad (\quad y) \quad Q(x_d; y_d; z)$$

where $R \in R^{M \times N}$ is the diffuse reflectance on a $M \times N$ grid.

4.1 Reconstruction of heterogeneous structure

For the set of captured images which correspond to different pixel to illumination line y, we capture a set of short range indirect images $I(\quad y)$. For the given set of images, we reconstruct the volume Q of unknown optical parameters by solving the following optimization problem,

$$\min_Q : \sum_{y=T_{d\,min}}^{T_{max}} \|I(\quad y) \quad I(R_0(\quad y) \quad (\quad y) \quad Q)\|_F^2 + \|Q\|_1; \tag{17}$$

where $\|:\|_F$ denotes the Frobenius norm, and $\quad$ is an unknown scaling factor which depends on the intensity and width of the laser profile and the sensitivity of the camera. The procedure for determining this factor $\quad$ is highlighted in more detail in Section 6:2. We also assume the reconstructed volume to be sparse, which essentially implies that the heterogeneous object only occupies a fraction of the total reconstructed volume.

The optimization is done over a range of y values. For smaller y values, the diffusion approximation breaks down, as the photon propagation is largely governed by single or very few scattering events. For very large y, not enough photons reach the camera pixels, and therefore the measurement images have a poor signal-to-noise ratio. Therefore, the range of y values needs to be chosen appropriately.

If we know the exact optical parameters $_s^0$ and $_a$ of the homogeneous background medium, then we can construct the kernel $(\quad y)$ as in (14). However in some cases, the background optical parameters of the material are not known. In those cases, we select a homogeneous patch inside the field of view, and the pixel intensity measurements with R_0 with respect to the unknown optical coefficients as in (9). We then use the estimated values of the coefficients to construct the phase kernel $(\quad y)$ for solving the optimization in (17).

We use PyTorch for implementation given it is highly optimized for convolution operations. The running time on a workstation with TianX GPU is around 5 minutes for 300 iterations for Q with a depth resolution of 64. The value in (17) is heuristically chosen to be 0:0001. We start the optimization with an initial value of all zeros for Q, and the reconstruction accuracy can be further improved if a better initialization is provided based on prior knowledge of the scene.

5 Hardware

In this section, we describe our imaging setup for capturing short range indirect images. In [24], a rectified configuration where the projector and camera are parallel is used for capturing the short-range images. That setup leads to a low spatial resolution over a large overlapping stereo volume. To capture high resolution images for small area of interest, we need a high spatial resolution over a smaller overlapping stereo volume. One way to achieve smaller overlapping stereo volume is to verge the projector and camera. This motivates us to design a verged setup for capturing high resolution short-range indirect images.

Our setup consists of a pair of synchronized rolling shutter camera and a laser projector implemented with Micro-Electro-Mechanical-Systems (MEMS). Our imaging setup is shown in Figuer 1 and Figure 6 (a). We use IDS-3950CPv2 industrial camera and Mirrorcle MEMS development kit. The central wavelength for the laser light is 680 nm. The MEMS mirror reflects the laser beam from the laser diode and the orientation of the MEMS mirror can be controlled in terms of two rotation axes (vertical and horizontal). The size of the imaged area on the sample is 8 cm by 8 cm. We model the laser diode and MEMS mirror pair as a pinhole projector whose center of projection is the center of rotation of the MEMS.

During the imaging process, the projector is scanned through the epipolar planes of the projector-camera pair. The camera is synchronized such that the pixels having a pre-defined offset from the corresponding epipolar line on the camera image plane are exposed. Each offset corresponds to one pixel to illumination line distance y as discussed in Section 4. For the physically rectified projector-camera pair as in [24], the epipolar lines on the projector and camera image are horizontal. This simply corresponds to illuminating and exposing the corresponding rows of projector and camera. In contrast, in our setup, the epipolar lines in the projector and camera are not horizontal due to the verged setup. So we cannot capture the short range indirect images by illuminating and exposing corresponding rows. Instead, on the projector side, we control the angle of the MEMS mirror to scan the light laser beam across a pencil of epipolar lines with different 2D slopes in the projector image plane. On the camera side, we interpolate over offset epipolar lines to get the short range indirect images. As a special case, for $y = 0$, the interpolated line overlaps with the epipolar. The resultant image is the direct light image.

Our image setup has smaller FOV than the rectified system in [24] due to the non-zero vergence angle between the project and camera. As a result, we can place the sample closer to the camera while the sample can still be illuminated by the projector. This enables higher image resolution for smaller area of interest so that more fine-grained (sub)surface details can be captured. In Figure 5, we show the images of a paper sticker captured with different devices. The sticker page with letters is occluded by several other pages so no letters can be seen under regular lighting. The occluded letters are visible in the short range indirect images from both [24] and our setup. Our device has smaller FOV and higher spatial resolution over the region of interest due to the verged configuration. In addition, we have better contrast and higher SNR because the laser light source used in our setup is of higher intensity compared to the pico-projector in [24]. The bright spot in the center and the lines in(d) is due to the reflection and inter-reflections from the protective glass in front of the MEMS mirror.

5.1 Calibration

The device is mounted vertically above the liquid container as shown in Figure 6 (a), with no cover above the scattering medium.

Fig. 5. Images of a paper sticker captured using different devices. The sticker page with letters is occluded by several other pages so no letters can be seen under regular lighting. (a) Image captured with cellphone camera under regular lighting. ; (b) Short-range indirect image captured with the device in [24]; (c) Enlarged image for the sticker region in (b); (d) Short-range indirect image captured with our device. Our device has smaller FOV due to non-zero vergence angle. The images captured with our device has higher resolution, SNR and contrast as shown in the insects in (c) and (d). The bright spot in the center and the lines in (d) is due to the re ection and inter-re ections from the protective glass in front of the MEMS mirror.

Fig. 6. Experiment setup and calibration to compensate the laser-mirror misalignment and non-linearity of MEMS. (a) The device is mounted vertically above the sample container, with no cover above the scattering medium. (b) During MEMS calibration, we consider the misalignment between the laser and mirror (above) and non-linearity of MEMS mechanics. Due to misalignment, the incident laser beam onto the MEMS mirror will not be perpendicular to the mirror surface and align with the MEMS rotation center; Due to non-linearity of MEMS mechanics, the input control signal and degrees of rotation are not linearly related.

Fig. 7. The MEMS-camera pose calibration. We illuminate a plane with known pose with 2D array of beams as shown in (a) and (b). Given the plane orientations, we can get the 3D parameters for the rays from multiple such images. Then we triangulate all the tted rays to determine the center of projection for the projector, in our case, the rotation center for the MEMS mirror.

We model the laser-MEMS as a pinhole projector whose center of projection is the rotation center of the MEMS mirror. During the calibration process, we estimate the relative pose between the MEMS mirror and the camera. For MEMS, we compensate for the non-linear mapping between the input voltage for the MEMS and the mirror tilt angle, and account for the mis-alignment of the MEMS mirror and the laser, as shown in Figure 6 (b).

More speci cally, we illuminate planes with given poses relative to the camera with a set of dot patterns. As shown in Figure 7, given the laser dot images for different plane positions, we can t the laser rays in 3D and triangulate the rays to get the origin of the rays,i.e. the rotation center of the MEMS mirror. Due to the laser-MEMS misalignment and tting error for the rays, the rays will not intersect at one 3D point. We solve a least square problem for the intersection point where the point to ray distances are minimized. The tted rays are also used to account for the non-linear relation between the MEMS input voltage and the rotation angle. In calibration, we build a lookup table relating the input voltage for the MEMS and the rotation angle for the mirror to account for their non-linear relation. During imaging, given the target laser ray direction, we can estimate the required input voltage by interpolating over lookup table.

6 Experiment Results

6.1 Simulation

We test the proposed algorithm using Monte Carlo rendered images. For the homogeneous medium, we use the scattering coef cients of human skin measured in [22]. The heterogeneous inclusions are located up to 4 mm below the surface. For the imaging setup, the width of the laser illumination line is 1 mm. The distance between the illumination line and the camera pixel ranges from 0 to 15 mm. To make the diffusion approximation valid for the algorithm, we only use the images with the illumination to pixel distance y larger than 2 mm.

The simulated direct and global light images are shown in the rst two rows in Figure 8. The global light image is the sum of the images captured with different y's except for $y = 0$. The inclusions can not be seen in the direct only image. For the global light image, because highly scattering property of skin, the contrast is low for some of the deeper inclusions, such as the solid circle in the right column. This makes the detection and localization for such objects (e.g. tumor beneath the skin surface) dif cult. For each short-range indirect image, the image intensity is contributed in part by the indirect light that travels from the illumination line with a preset distance to the imaged scene point. As a result, compared with the global light image, the contrast of