

Optical Engineering

SPIDigitalLibrary.org/oe

Surface reconstruction from multiview projection data employing a microlens array-based optical detector: a simulation study

Xiaoming Jiang
Liji Cao
Jörg Peter

Surface reconstruction from multiview projection data employing a microlens array-based optical detector: a simulation study

Xiaoming Jiang,* Liji Cao, and Jörg Peter

German Cancer Research Center, Division of Medical Physics in Radiology, Im Neuenheimer Feld 280, Heidelberg 69120, Germany

Abstract. This article proposes a surface reconstruction method from multiview projectional data acquired by means of a rotationally mounted microlens array-based light detector (MLA-D). The technique is adapted for *in vivo* small animal imaging, specifically for imaging of nude mice, and does not require an additional imaging step (e.g., by means of a secondary structural modality) or additional hardware (e.g., laser-scanning approaches). Any potential point within the field of view (FOV) is evaluated by a proposed photo-consistency measure, utilizing sensor image light information as provided by elemental images (EIs). As the superposition of adjacent EIs yields depth information for any point within the FOV, the three-dimensional surface of the imaged object is estimated by a graph cuts-based method through global energy minimization. The proposed surface reconstruction is evaluated on simulated MLA-D data, incorporating a reconstructed mouse data volume as acquired by x-ray computed tomography. Compared with a previously presented back projection-based surface reconstruction method, the proposed technique yields a significantly lower error rate. Moreover, while the back projection-based method may not be able to resolve concave surfaces, this approach does. Our results further indicate that the proposed method achieves high accuracy at a low number of projections. © The Authors. Published by SPIE under a Creative Commons Attribution 3.0 Unported License. Distribution or reproduction of this work in whole or in part requires full attribution of the original publication, including its DOI. [DOI: [10.1117/1.OE.53.2.023104](https://doi.org/10.1117/1.OE.53.2.023104)]

Keywords: surfaces; optical design; simulations.

Paper 131467 received Sep. 24, 2013; revised manuscript received Jan. 22, 2014; accepted for publication Jan. 27, 2014; published online Feb. 27, 2014.

1 Introduction

Recently, a noncontact light detector utilizing a purposely designed microlens array (MLA-D) has been developed.¹ The detector consisted of a MLA, a septum mask, and a photon sensor [cf. Fig. 1(a)]. In contrast to conventional lens-based imaging systems, an MLA-D possesses a thin construction size (<10 mm including cooling). Since its imaging performance is optimal for objects positioned close to its optics, a single or a multitude of MLA-Ds can be arranged circumferentially in close proximity to the imaged object, hence allowing for confined system assembly. Furthermore, the whole optical imaging system can potentially be encircled by another imaging system such as positron emission tomography or magnetic resonance imaging (MRI), in which MLA-Ds are compatible or nearly “invisible.”²

An MLA-D does not form an immediate observer image. Corresponding to each microlens, a local lattice of sensor pixels forms a so-called elemental image (EI), as shown in Fig. 1(b). Therefore, an image needs to be calculated by inverse mapping³ or an iterative algorithm, as depicted in Ref. 4. For noncontact optical *in vivo* imaging, the procedure for localizing emitted photons within the imaged object—with the purpose to further solve the inverse problem involved in image reconstruction^{5,6}—is crucially affected by the exact knowledge of the three-dimensional (3-D) surface of the imaged object. As of today, surface information is mainly derived from secondary data as provided by

(1) imaging the object with another modality such as x-ray computed tomography (CT)⁵ or MRI⁷ or (2) conducting a structured light 3-D scanning procedure.⁸ Alternatively, direct optical methods such as back-projection-based (silhouette or visual hull) based approaches have also been proposed.^{9–11} While the first class of methods requires additional hardware and scanning steps, the second class is generally considered less accurate, as concave areas cannot always be resolved.

As both classes of approaches operate independently from the light camera being used, they have been proposed in the literature (a survey is given in Ref. 12) as either conventional or multilens-based camera approaches such as the MLA-D employed here.^{13,14} For the latter systems though, it has been shown recently that 3-D imaging, including surface reconstruction, can be achieved by so-called integral imaging techniques.^{15,16}

Using MLAs to record and recover light fields has become a recent research highlight in the fields of optics and computer vision.^{17–19} Shape reconstruction or depth retrieval using MLAs can be performed from a single projection by finding corresponding points among neighboring EIs or adjacent focal point images.^{15,20} Due to the limited number of microlenses and/or light trajectories, the reconstructed surface possesses a rather low-depth resolution and is insufficient for the downstream task of image reconstruction from optical data (tomography).

Considering that the MLA-D is either part of a multidetector imaging setup or is positioned or rotated around the imaged object, the aforementioned problem could be improved upon by combining multiple projections from

*Address all correspondence to: Xiaoming Jiang, E-mail: xiaoming.jiang@dkfz.de

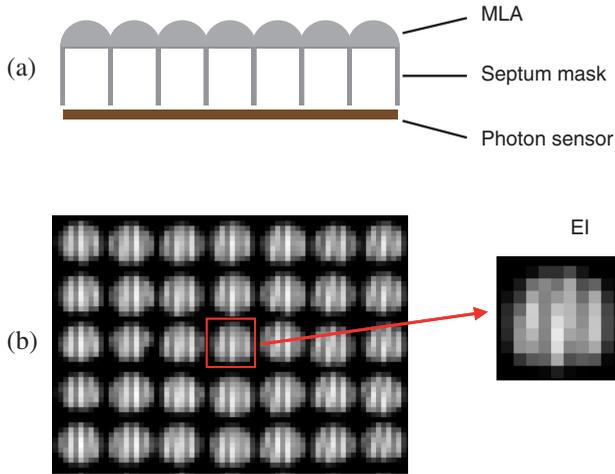


Fig. 1 Cross-sectional view of the microlens-based optical detector (MLA-D) containing a microlens array (MLA), a septum mask, and a photon sensor (a). Experimental data of adjacent elemental images (EIs) using an assembled MLA-D, as listed in setup 1 (Table 1). Each EI contains 10×10 pixels (b).

multiview data (at unchanged illumination condition), which has been extensively studied in computer vision.^{21–23} However, the direct application of these algorithms in the MLA-D system is difficult for two main reasons. First, global illumination either by environmental light or by fixed distributed light sources is difficult to achieve due to the enclosed design.¹ Second, the EIs of the MLA-Ds possess a comparatively low resolution so the visibility computation is difficult to perform.

Therefore, an integrated instrumentation setup for *in vivo* small animal (mice) imaging is presented by means of a simulation study employing MLA-Ds of various sizes as well as by a new light source setup for object illumination. Furthermore, a dedicated surface reconstruction algorithm is presented for this setup, which is based on a newly defined photo-consistency measure (the concept of photo-consistency measure can be referred in Ref. 22) and volumetric graph cuts. It is particularly tailored to the comparatively large number of EIs in our design, albeit each EI possesses a rather low-spatial resolution.

2 Methods

2.1 MLA-D

As depicted in Fig. 2, the binary sensor data of an MLA-D represents a two-dimensional (2-D) matrix $\mathbf{B} = [b_{m,n}] (m = 1, \dots, M; n = 1, \dots, N)$ with square pixel size p , where M and N are the number of pixels in x and y directions on the photosensor plane, respectively. The arrangement of sensor-aligned $G \times H$ microlenses with lens pitch d_l is represented

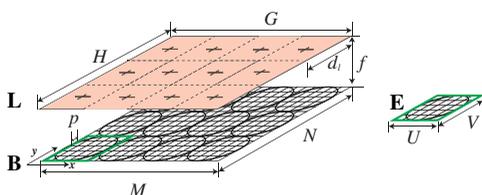


Fig. 2 Illustration of MLA-D notation.

Table 1 Microlens array-based light detector (MLA-D) specification for the simulation setups.

Parameter	Setup 1	Setup 2
Focal length, f (mm)	2.2	2.4
Lens pitch, d_l (μm)	480	520
Half-cone angle, θ (deg)	6.2	6.2
Pixel size, p (μm)	48	6.5
Lens grid, $G \times H$	70×60	64×54
Sensor resolution, $M \times N$	700×600	5120×4320
Elemental image (EI) resolution, $U \times V$	10×10	80×80
Detector size (mm^2)	33×28	33×28

by $\mathbf{L} = [l_{g,h}] (g = 1, \dots, G; h = 1, \dots, H)$. An elemental image \mathbf{E} is formed with local pixels corresponding to microlens $l_{g,h}$, represented as $\mathbf{E} = [e_{u,v}] (u = 1, \dots, U; v = 1, \dots, V)$, where U and V are the numbers of pixels in x and y directions, respectively.

Currently, two types of MLA-D setups have been proposed,^{1,14} for which the parameters are summarized in Table 1, respectively. The first design has been assembled and is being used in practice, whereas the second is still under construction.

2.2 System Simulation

The simulation is performed within the physically based rendering (PBRT) framework, which includes interfaces for detector, lighting, and scenario (imaged object) descriptions.²⁴ The configuration of the MLA-Ds is defined by the parameters in Table 1. As a result of the previously mentioned illumination source confinement, a local illumination setup as illustrated in Fig. 3 is selected. As can be seen, adjacent MLA-Ds share the same illumination. The whole assembly is also rotatable around the long axis of the imaged object to acquire the data at A angular views. At each projectional view, two sensor data, \mathbf{B}_a^1 and \mathbf{B}_a^2 ($a = 1, \dots, A$), are obtained from the detectors.

In order to define a realistic phantom representing an imaged object, a triangulated mesh extracted from a nude mouse CT dataset²⁵ is adopted as input for the simulation,

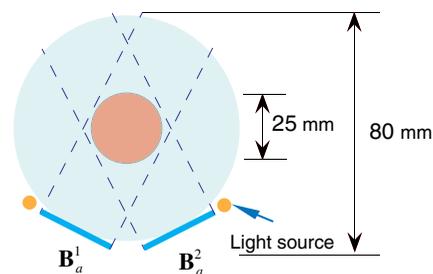


Fig. 3 Geometry of the proposed system design and notations. This design ensures a local illumination condition, because two MLA-Ds share the illumination by a pair of fixed diffuse line sources. When the system rotates, both MLA-Ds and light sources rotate at the same offset.



Fig. 4 Exemplary projection image of a real mouse CT data acquisition. A triangulated mesh is derived from the CT volume and acts as the phantom for the simulation. The box marked in red refers to the desired region of interest, as imaged by the MLA-Ds.

as shown in Fig. 4. The actual (identical) fields of view (FOVs) of the simulated MLA-Ds are marked by the red box. Once the surface data have been derived, its reflectance is set to be ideally diffusive (Lambertian surface).^{26,27} Considering the size of normal mice (approximately 25 mm in diameter) as well as the depth of field for the individual microlenses, the radius of rotation is set to be 40 mm. The oblique angle between the two detectors is set to be 40 deg in order to maintain the condition that any point on the surface of the imaged object is being seen on either detector (photo-consistency). Further details about the simulation can be found in Ref. 14. The simulation is carried out by employing a high-performance computer cluster with 128 CPU cores (Intel E5450, 3.0 GHz, 16 GB memory for each core) for parallel implementation of different microlens units.

2.3 Surface Reconstruction

Centrally important to the surface reconstruction procedure, a photo-consistency measure is proposed by comparing the formed EIs following the system shown in Fig. 3. By calculating and super-positioning all photo-consistency measures, a photo-consistency volume of the imaged object is obtained. The problem of surface reconstruction is applied onto the formed volume, concretely to decide whether any voxel belongs to the imaged object, also known as labeling problem.²⁸ The labeling problem can be expressed as an energy minimization form and is further solved by the graph cuts-based method.²⁹

2.3.1 Photo-consistency measure

To solve the problem of surface reconstruction from multi-view stereo in computer vision, a number of approaches have been proposed for evaluating the visual compatibility of a reconstruction with a collection of input images.³⁰ Most of these measures (often entitled photo-consistency measures) operate by comparing pixels in one image to pixels in other images to see how well they correlate.²¹ To define a suitable measure for the surface reconstruction employing the MLA-D configuration, notations are first introduced.

In the process of image formation, any point $\mathbf{P}(x, y, z)$ on the surface of the imaged object, which is within the FOV of either MLA-D, can potentially be detected by a number of microlenses. A ray originating from $\mathbf{P}(x, y, z)$ that forms a trajectory forward an EI is called a valid ray. A set of EIs containing all valid rays $\hat{\mathbf{T}}$ of a given $\mathbf{P}(x, y, z)$ is referred to as $\hat{\mathbf{E}}$, as shown in Fig. 5. The intensities as collected by the sensor of the formed trajectories $\hat{\mathbf{T}}$ tend to be similar since the corresponding sensor pixels in $\hat{\mathbf{E}}$ measure the same point \mathbf{P} in space, only from slightly different angles. Instead of comparing the absolute differences among the collected

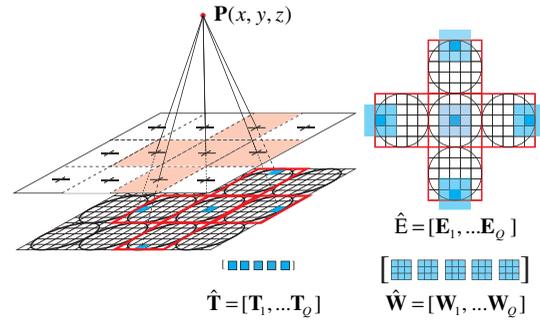


Fig. 5 Illustration of image formation by a single point. Any point \mathbf{P} would form a set of trajectories ($\hat{\mathbf{T}}$) corresponding to multiple EIs ($\hat{\mathbf{E}}$). In order to estimate similarities among trajectories, the correlation among pixel windows ($\hat{\mathbf{W}}$) is derived.

trajectories directly, a more robust method is to compare the similarities between two pixel windows since area matching (pixel window) compares regularly sized regions of pixel values in two images and, hence, is well suited for a texture scene.³¹ For example, two arbitrary trajectories \mathbf{T}_r and \mathbf{T}_p from $\hat{\mathbf{T}}$ can be extended to form corresponding pixel windows \mathbf{W}_p and \mathbf{W}_r , respectively (for setup 1, a 3×3 pixel window is used, while for setup 2, a 7×7 pixel window is used). The collection of pixel windows is represented as $\hat{\mathbf{W}}$. Following common strategies assessing similarities between data formed by two pixel windows, the normalized cross-correlation (NCC) is employed in this article as a similarity measure. It is defined as³¹

$$\text{NCC}(\mathbf{W}_p, \mathbf{W}_r) = \frac{\sum_{i,j} (\mathbf{W}_p^{i,j} - \bar{\mathbf{W}}_p)(\mathbf{W}_r^{i,j} - \bar{\mathbf{W}}_r)}{\sqrt{\sum_{i,j} (\mathbf{W}_p^{i,j} - \bar{\mathbf{W}}_p)^2} \sqrt{\sum_{i,j} (\mathbf{W}_r^{i,j} - \bar{\mathbf{W}}_r)^2}}, \quad (1)$$

where $\bar{\mathbf{W}}_p$ and $\bar{\mathbf{W}}_r$ represent the average value of two corresponding pixel windows, respectively. This function yields return values between -1 and 1 . A perfect match would reach the maximum of 1 . In our context, a high correlation is defined ($\text{NCC} > 0.7$), which is used in the following part. Employing the NCC function as described decreases the effect of noise in the data due to, among others, the inhomogeneous responses of sensor pixels.

In surface reconstruction procedure, a volume model is adopted to represent the surface for its simplicity and uniformity.²¹ A reasonable simplification can be made, though. Any voxel $\mathbf{V}_{i,j,k}$ (the concept ‘‘voxel’’ here refers to a point in a spatial grid) can form trajectories onto the photon sensor plane assuming that there would be no self-occlusion caused by other voxels.²² Using the proposed design as illustrated in Fig. 3, a voxel $\mathbf{V}_{i,j,k}$ can be seen through Q^1 and Q^2 microlenses in the two MLA-Ds simultaneously. Hence, $\hat{\mathbf{E}}$, $\hat{\mathbf{T}}$, as well as $\hat{\mathbf{W}}$ are formed by two subsets: $\hat{\mathbf{E}}^1$ and $\hat{\mathbf{E}}^2$; $\hat{\mathbf{T}}^1$ and $\hat{\mathbf{T}}^2$; and $\hat{\mathbf{W}}^1$ and $\hat{\mathbf{W}}^2$, as shown in Fig. 6.

Because of the specific design and configuration of MLA-Ds, the problem of defining a photo-consistency measure is extended from pair comparison between pixel windows, as seen in Eq. (1), to formalize a new measure over two subsets of EIs ($\hat{\mathbf{E}}^1$ and $\hat{\mathbf{E}}^2$) and the corresponding pixel windows ($\hat{\mathbf{W}}^1$ and $\hat{\mathbf{W}}^2$). A simple method is to choose one reference pixel window \mathbf{W}_r within the elemental image \mathbf{E}_r from $\hat{\mathbf{W}}$ and to

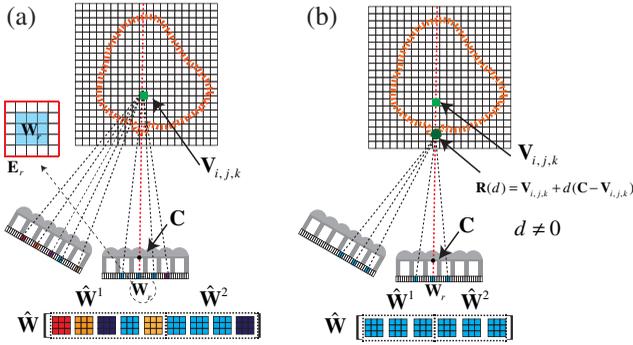


Fig. 6 An arbitrary point $V_{i,j,k}$ forms pixel windows on MLA-Ds represented as \hat{W} (\hat{W}_1 and \hat{W}_2 on two detectors, respectively). One pixel window W_r is chosen as reference window (correspondingly, the EI formed is represented as E_r). Some pixel windows in \hat{W} possess high correlations with W_r , while others do not (a). Supposing that there exists an optic ray (dashed line marked in red) between $V_{i,j,k}$ and the optical center C of the corresponding microlens forming E_r , any point $R(d)$ on the optic ray would form the same pixel window W_r within E_r . When $R(d)$ is on the surface, the formed pixel windows in \hat{W} have high correlations with W_r , as seen in (b).

compare it with the other formed pixel windows from the same subset \hat{W} , one by one. For an arbitrary $V_{i,j,k}$, it might occur that some of the pixel windows possess a high degree of correlation with W_r , while some other pixel windows possess a low degree of correlation, as seen in Fig. 6(a). Although the direct average of all NCC values could be applied, the resolvability among adjacent voxels might be hampered.²²

Therefore, an alternative method is used. The assumption is made that there exists an optic ray between $V_{i,j,k}$ and the optical center C of a chosen microlens, as depicted in Fig. 6(b). Correspondingly, the formed pixel window of this microlens is chosen as reference pixel window W_r . As depicted in Fig. 6(b), an arbitrary point $R(d)$ on the supposed optic ray defined by $V_{i,j,k}$ and C could be expressed according to

$$\mathbf{R}(d) = \mathbf{V}_{i,j,k} + d(\mathbf{C} - \mathbf{V}_{i,j,k}), \quad (2)$$

where d is a variable to control the position of $\mathbf{R}(d)$. $\mathbf{R}(d)$ would form the same pixel window W_r within the corresponding E_r as d changes, as shown in Fig. 6(b).

However, in any other EI, e.g., a fixed E_f from \hat{E} that can detect $V_{i,j,k}$, $\mathbf{R}(d)$ would form different pixel windows as compared with the pixel window formed by $V_{i,j,k}$ within E_f . The newly formed pixel window series by $\mathbf{R}(d)$ is represented as \hat{W}_f as d changes. When comparing W_r with the formed set \hat{W}_f , a maximum NCC value can be found with respect to d . If $V_{i,j,k}$ is on the surface (or very close to the surface of the object), the degree of correlation becomes a maximum at $d = 0$, as seen in Fig. 7. Contrarily, if $V_{i,j,k}$ is not on the surface, the degree of correlation is lower at $d = 0$ or becomes a maximum at $d \neq 0$, as shown in Fig. 6(b). In other words, the maximum value of NCC at $d = 0$ can be used as a strong indicator for determining whether $V_{i,j,k}$ is a point on the surface or not.

Having discussed the case of mapping $V_{i,j,k}$ with one microlens, we extend this principle to combine $V_{i,j,k}$ with multiple microlenses. Supposing N_a microlenses are chosen in the a -th projection, the number of occurrences, M_a , in

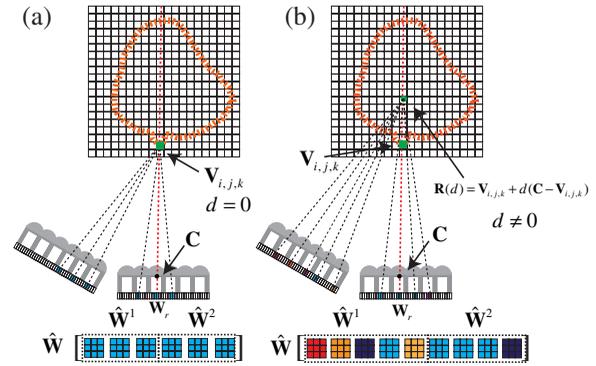


Fig. 7 When voxel $V_{i,j,k}$ is on the surface of the imaged object, the chosen reference window W_r possesses a high degree of correlation with other pixel windows at $d = 0$ (a). When d changes, some pixel windows in \hat{W} possess high correlations with W_r , while others do not (b).

which a maximum NCC value is derived at $d = 0$, would be counted. The ratio M_a/N_a can be used to describe the degree whether $V_{i,j,k}$ is on the surface concerning the normalization effect.

In order to reduce the computational complexity, the following strategies are employed. First, W_r and W_f are chosen from two different detectors. Reasoning is due to the fact that the MLA-Ds configuration is symmetric, as seen in Fig. 3. If there is a case that N_a microlenses are chosen from the first MLA-D detector while M_a occurrences with maximum NCC are obtained on the second MLA-D, equivalently, then it could occur that N'_a microlenses are chosen from the second MLA-D detector while M'_a occurrences are obtained on the first MLA-D. Second, only EIs within limited s rows (\hat{E}_1) are chosen for W_r , and t rows of EIs (\hat{E}_2) are chosen for W_f (s and t are set as 3, in this article). The photo-consistency measure $\varphi(V_{i,j,k})$ is defined as

$$\varphi(V_{i,j,k}) = \exp \left[-\mu \frac{\sum \frac{M_a(V_{i,j,k})}{N_a(V_{i,j,k})} + \frac{M'_a(V_{i,j,k})}{N'_a(V_{i,j,k})}}{2V_A} \right], \quad (3)$$

in which μ is a rate-of-decay parameter (set to be 0.5, in this article), and V_A is the number of effective projectional views ($M_a/N_a > 0.001$ and $M'_a/N'_a > 0.001$). The details about the implementation could be found in Algorithm 1.

2.3.2 Volumetric surface reconstruction based on graph cuts

By calculating the proposed photo-consistency for each voxel employing Eq. (3), a volume of the imaged object is derived. The surface reconstruction task in this context is to assign a label to each voxel of the image object, i.e., a voxel either belongs to the background or to the imaged object. In accordance with labeling a voxel, an energy cost is paid. The objective is to find a discrete labeling of voxels with minimum systematic total energy, known as labeling problem.²⁸ In our context of a binary labeling problem (only two labels for the background and the imaged object), two energy parts are included. One part of cost concerns the label itself (data cost), i.e., the volume of imaged object in this context. This part of total cost can be seen as a volume integration of the imaged object. The other part of cost

Algorithm 1 Photo-consistency measure calculation.

Input:

$$\mathbf{V}_{i,j,k}, \hat{\mathbf{B}}_{fov} = [\mathbf{B}_a^1, \mathbf{B}_a^2] (a = 1 \dots A)$$

Begin:

 Initialize $V_A = 0$
for each $\mathbf{B}_a^1, \mathbf{B}_a^2$ in $\hat{\mathbf{B}}_{fov}$

 Choose $\hat{\mathbf{E}}_1$ within s rows of El's from first detector and corresponding lens center information

 Choose $\hat{\mathbf{E}}_2$ within t rows of El's from second detector

Begin Pixel windows-comparison:
 $\mathbf{V}_{i,j,k}$ is visible through $N_a(\mathbf{V}_{i,j,k})$ microlenses in $\hat{\mathbf{E}}_1$

 Reference pixel windows set: $\hat{\mathbf{W}}_r = [\mathbf{W}_r^l] (l = 1 \dots N_a)$

 Optical center of N_a microlenses: $\hat{\mathbf{C}} = [\mathbf{C}_l] (l = 1 \dots N_a)$
 $\mathbf{V}_{i,j,k}$ is visible through $F(\mathbf{V}_{i,j,k})$ microlenses in $\hat{\mathbf{E}}_2$

$$\hat{\mathbf{E}}_2^f = [\mathbf{E}_f] (f = 1 \dots F)$$

 Initialize $M_a(\mathbf{V}_{i,j,k}) = 0$
for each \mathbf{W}_r^l in $\hat{\mathbf{W}}_r$
for each \mathbf{E}_f in $\hat{\mathbf{E}}_2^f$
for each d

$$\mathbf{R}(d) = \mathbf{V}_{i,j,k} + d(\mathbf{C}_l - \mathbf{V}_{i,j,k})$$

 Calculate formed pixel window \mathbf{W}_f within \mathbf{E}_f by $\mathbf{R}(d)$

 Calculate NCC value between \mathbf{W}_r^l and \mathbf{W}_f

 end loop for d
if NCC between \mathbf{W}_r^l and \mathbf{W}_f reaches maximum at $d = 0$

$$M_a(\mathbf{V}_{i,j,k}) ++$$

 end loop for \mathbf{E}_f

 end loop for \mathbf{W}_r^l

 end *Pixel window-comparison*:

 Choose $\hat{\mathbf{E}}_1$ within s rows of El's from second detector and corresponding lens center information

 Choose $\hat{\mathbf{E}}_2$ within t rows of El's from first detector

 Do *Pixel window-comparison* again to obtain N'_a and M'_a
if $M_a/N_a > 0.001$ and $M'_a/N'_a > 0.001$

$$V_A ++$$

 end loop for $\mathbf{B}_a^1, \mathbf{B}_a^2$

$$\text{Photo-consistency: } \varphi(\mathbf{V}_{i,j,k}) = \exp\left(-\mu \frac{\sum M_a(\mathbf{V}_{i,j,k}) + M'_a(\mathbf{V}_{i,j,k})}{2V_A + N_a(\mathbf{V}_{i,j,k}) + N'_a(\mathbf{V}_{i,j,k})}\right)$$

concerns the regularity of labels among neighboring voxels (regularity cost). When two neighboring voxels have the same label, the regularity cost is zero, whereas two neighboring voxels with the different labels yield nonzero regularity cost values. The regularity cost in this case is set to be the value of the photo-consistency measure calculated by Eq. (3). Because regularity cost is valid across the boundary between the imaged object and its background, the total cost of this part can be seen as a surface integration of the photo-consistency measure. The surface reconstruction problem is transformed to extract an optimal surface S_{opt} , over which the surface integral of the photo-consistency measure $\varphi(\mathbf{V})$ is minimized as well as the volume $V(S)$ enclosed by S_{opt} is maximized. This can be expressed as minimization of the following equation:²²

$$E(S) = \iint_S \varphi(\mathbf{V}) dA - \lambda \int \int \int_{V(S)} dV, \quad (4)$$

in which λ is a parameter to control the weight of volume integration (λ is set to be within 0.05 and 0.25, in this article). In this context, the first term in Eq. (4) is comprehended to be a collapsing force, whereas the second term of volume integration is comprehended to be an expansion force.

The optimization problem in Eq. (4) is solved by the graph-cuts-based method using the implementation by Boykov et al. with expansion moves and swap moves.²⁹ Each voxel is treated as a node in a graph. Given the initial estimation of the imaged object as shown in Fig. 8, i.e., voxels inside of surface S_{in} , the total cost would decrease as the nodes expand. This expansion would stop at the surface of the imaged object, i.e., the balance between the collapsing and ballooning forces (provided that any voxel on the surface possesses a much lower photo-consistency measure, and a good choice of λ as indicated above). The convergence and efficiency of this approach has been validated in previous studies.³²

2.4 Comparison with Back-Projection Method

For comparison, we also applied a previously presented back-projection method¹⁴ to reconstruct the object's surface using data in one detector from multiview data. The reconstructed surfaces from 6, 12, 18, 24, 30, and 36 views, respectively, are evaluated with respect to the two

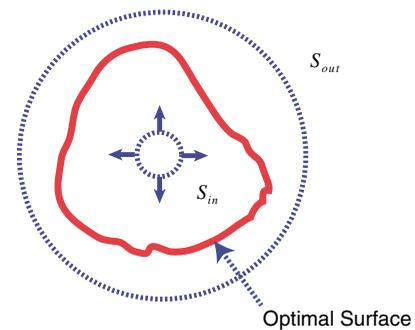


Fig. 8 Illustration of how the optimal surface is obtained between the estimated initial outer boundary S_{out} and the inner boundary S_{in} . Given the initial estimation of the imaged object, the nodes expand until they reach the surface. A minimal systematic cost is achieved when voxels on the surface possess a lower photo-consistency measure value.

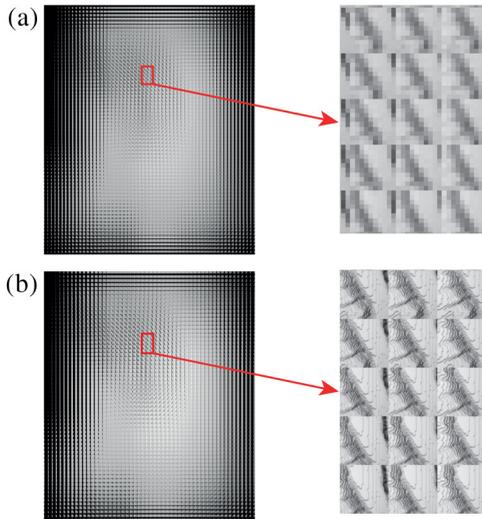


Fig. 9 Exemplary MLA-D raw images resulting simulations from a single view alongside magnified views within the red box. (a and b) The results according to the configuration of setup 1 and setup 2, respectively.

methods, and an error rate $\eta = |\mathbf{V}_{\text{phantom}} \oplus \mathbf{V}_{\text{recon}}| / |\mathbf{V}_{\text{phantom}}|$ is derived, where $\mathbf{V}_{\text{phantom}}$ represents the binary volume of the phantom, and $\mathbf{V}_{\text{recon}}$ is the binary volume after surface reconstruction. \oplus means exclusive or operation between two binary volumes, and $|\cdot|$ is a l_1 -norm operation of binary data.

3 Results

PBRT simulations have been carried out for both setups as listed in Table 1 to generate multiview projection data. Two exemplary MLA-D simulations showing a single view according to the configurations are depicted in Fig. 9. The total computational burden, for example to generate 36 views, was about 4 h for setup 1 and 16 h for setup 2.

Volumetric surface reconstruction is performed on a $256 \times 256 \times 249$ voxel grid with 0.13 mm between two neighboring voxels. Calculated slices of photo-consistency measures according to the 6 and 36 views are compared, as shown in Fig. 10.

For the same slice, the results from back-projection and the proposed method are compared in Fig. 11. The first row

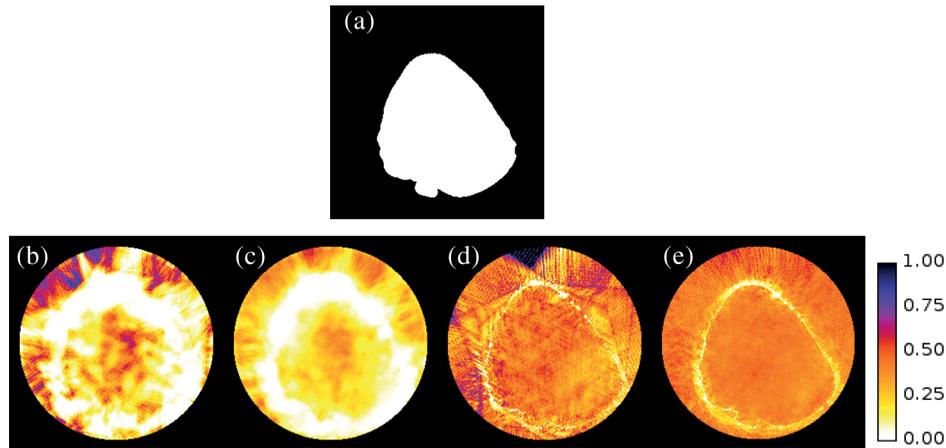


Fig. 10 Photo-consistency measure comparison calculated from 6 and 36 views of data, respectively. (a) The ground truth (from the CT volume); (b and c) results for setup 1; and (d and e) results for setup 2. Note that (b and d) are using 6 projections, while (c and e) are using 36 projections.

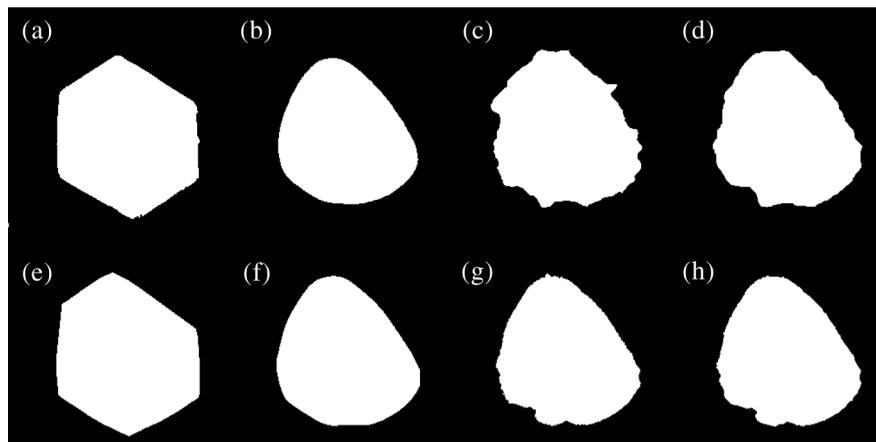


Fig. 11 Comparison of reconstructed slices by the back-projection and the proposed methods using the two setups for the same slice, as in Fig. 10. (a and b) The results by back-projection method using 6 and 36 projections, respectively; (c and d) Results by the proposed method employing the same number of views as for setup 1. (e-h) Results according to setup 2 following the same order as (a-d).

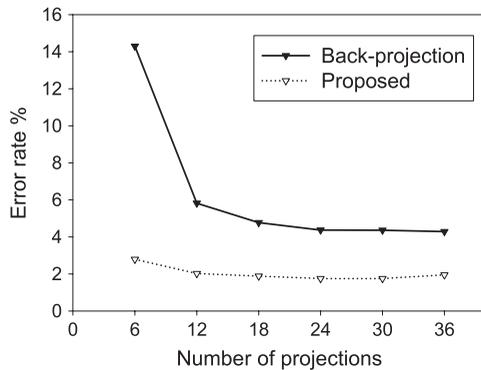


Fig. 12 Measured error rates for the proposed and the back-projection methods using different numbers of views for setup 2.

shows reconstructed slices for setup 1 employing 6 and 36 views of data by the back-projection method [Figs. 11(a) and 11(b), respectively] and by the proposed method [Figs. 11(c) and 11(d)]. The second row of Fig. 11 shows the results using setup 2 following the same order as the first row of figures.

When 36 projectional views are used, setup 1 yields an error rate of 14.18%, while setup 2 results in an error rate of 1.95%. A quantitative evaluation of error rate comparing the back-projection method and the proposed method is performed for setup 2, as shown in Fig. 12. The curves indicate that the proposed method shows a significant improvement of accuracy as compared with the back-projection method.

Results of rendered surfaces for setup 2 are shown in Fig. 13. Considerable errors in shape restoration for results of the back-projection method, such as particularly in the leg areas of the mouse, are clearly evident [cf. Fig. 13(c)]. In contrast, these unresolved structures are being better preserved by using the proposed method, as shown in Fig. 13(b).

4 Discussion and Conclusion

In the field of *in vivo* optical imaging, the back-projection method is a frequently used approach for surface reconstruction. This method not only requires a high number of projections (100 to 300 in general) to produce acceptable results,^{9,10} but it also possesses inherent shortcomings, mainly regarding its inability to resolve concave areas. Surface reconstruction from multiview projection data using the proposed method has been demonstrated to resolve

concave surface areas better. Furthermore, the required number of projections can be significantly reduced when employing MLA-Ds, since MLA-Ds provide additional angular information per detector position. The simulation results indicate that satisfying results can be achieved by 6 to 36 views of projection using the proposed new setup. In contrast to other surface detection approaches used within the research field of *in vivo* optical imaging, such as using structured light or employing a secondary structural imaging procedure, the approach as proposed in this article does not require any other additional hardware.

Because experimentally acquired data were not available as the final MLA-D construction is still to be completed, results are obtained using the PBRT simulation approach. However, as we run the simulation on highly realistic object data, experimentally acquired by x-ray CT of a nude mouse scan, relying on a simulation study does not degrade its value. By incorporating ray-tracing techniques tailored specifically to the optical layout of the MLA-Ds investigated, the PBRT framework has once more proven to be a state-of-the-art tool for PBRT in computer graphics and, in our context, 3-D *in vivo* imaging. In addition, we neglected noise in the process of MLA-D image generation mainly due to the fact that the images are acquired under controlled illumination with high signal-to-noise ratio in practice [cf. Fig. 1(b)].

As described in Ref. 22, reference pixel windows were compared with images taken with six closest cameras to calculate the NCC value, which further worked as photo-consistency measure. In contrast to conventional single-lens cameras, MLA-Ds provide multitude EIs covering the same area of interest from (slightly) different views. Hence, a greater number of EIs can be chosen as reference or comparing images and further analyzed to calculate similarity measures. Rather than using the NCC value directly, the proposed measure as described herein combines these inherent characteristics of MLA-Ds by calculating the rate of occasions when the NCC value reaches a maximum at $d = 0$. Statistically it uses more information from the neighboring views. However, as $R(d)$ changes along the virtual optic ray, pairwise comparison of pixel windows from multitude EIs yields very high computational complexity. The calculation of similarity between one pair of pixel windows is independent from the others. Hence, the calculation of photo-consistency could be decomposed into several sub-tasks and could be further accelerated using parallel computing techniques in the future.

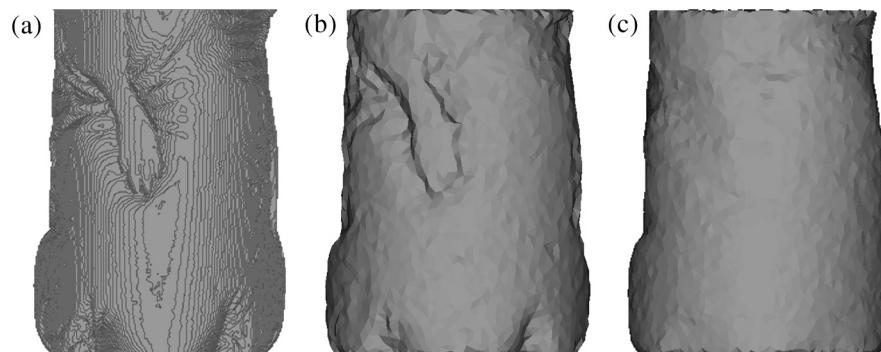


Fig. 13 Comparison of the rendered surface for the phantom (a), reconstructed surface by the proposed method (b), and the back-projection method (c) from 36 views using setup 2.

Increasing the number of projectional views does improve the reconstruction results, as seen in Fig. 10; cf. also Ref. 22. The performance of the two setups for 3-D surface reconstruction differs significantly. Because pixel size of setup 2 is about seven times smaller than that of setup 1, setup 2 features a much higher space sampling rate than that of setup 1. Hence, higher-frequency information can be resolved which could potentially make pixel window comparison even more robust. Moreover, the EIs of setup 2 possess higher spatial resolution, such that a bigger pixel window could be used when the similarity measure is calculated. These two advantages are helpful in suppressing false-positive cases in the presence of noise.³³

In conclusion, this article verifies the feasibility of reconstructing 3-D surfaces from multiview data solely by using MLA-D data within the research field of *in vivo* small animal imaging. We presented a new imaging instrumentation design with respect to the alignment of MLA-Ds and light sources. Given this conceptual imaging system setup, a corresponding algorithm for 3-D surface reconstruction is presented and investigated. The results indicate that the proposed method does have the ability to resolve concave areas and to achieve a high accuracy, even when using a significantly low number of projection views, without the requirement of any additional hardware.

Acknowledgments

The first author gratefully acknowledges the financial support from the China Scholarship Council (CSC).

References

- J. Peter et al., "Development and initial results of a tomographic dual-modality positron/optical small animal imager," *IEEE Trans. Nucl. Sci.* **54**(5), 1553–1560 (2007).
- J. Peter, "Synchrotron optical *in vivo* imaging employing microlens array optics: a complete framework," *Proc. SPIE* **8574**, 857402 (2013).
- D. Unholtz et al., "Image formation with a microlens-based optical detector: a three-dimensional mapping approach," *Appl. Opt.* **48**(10), D273–D279 (2009).
- L. Cao et al., "Iterative reconstruction of projection images from a microlens-based optical detector," *Opt. Express* **19**(13), 11932–11943 (2011).
- L. Cao and J. Peter, "Bayesian reconstruction strategy of fluorescence-mediated tomography using an integrated spect-ct-ot system," *Phys. Med. Biol.* **55**(9), 2693–2708 (2010).
- L. Cao, M. Breithaupt, and J. Peter, "Geometrical co-calibration of a tomographic optical system with CT for intrinsically co-registered imaging," *Phys. Med. Biol.* **55**(6), 1591–1606 (2010).
- B. Brooksby et al., "Imaging breast adipose and fibroglandular tissue molecular signatures by using hybrid MRI-guided near-infrared spectral tomography," *Proc. Natl. Acad. Sci. U. S. A.* **103**(23), 8828–8833 (2006).
- H. R. Basevi et al., "Simultaneous multiple view high resolution surface geometry acquisition using structured light and mirrors," *Opt. Express* **21**(6), 7222–7239 (2013).
- H. Meyer et al., "Noncontact optical imaging in mice with full angular coverage and automatic surface extraction," *Appl. Opt.* **46**(17), 3617–3627 (2007).
- G. Hu et al., "Fluorescent optical imaging of small animals using filtered back-projection 3d surface reconstruction method," in *Int. Conf. BioMedical Engineering and Informatics, 2008. BMEI 2008*, Vol. 2, pp. 76–80, IEEE, Sanya, Hainan, China (2008).
- N. Deliolanis et al., "Free-space fluorescence molecular tomography utilizing 360° geometry projections," *Opt. Lett.* **32**(4), 382–384 (2007).
- F. Stuker, J. Ripoll, and M. Rudin, "Fluorescence molecular tomography: principles and potential for pharmaceutical research," *Pharmaceutics* **3**(2), 229–274 (2011).
- J. Peter et al., "Instrumentation setup for simultaneous measurement of optical and positron labeled probes in mice," in *2011 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSSMIC)*, pp. 2306–2308, IEEE, Valencia, Spain (2011).
- X. Jiang et al., "Evaluation of shape recognition abilities for micro-lens array based optical detectors by a dedicated simulation framework," *Proc. SPIE* **8573**, 85730Q (2013).
- J. Jung et al., "Reconstruction of three-dimensional occluded object using optical flow and triangular mesh reconstruction in integral imaging," *Opt. Express* **18**(25), 26373–26387 (2010).
- R. Horisaki et al., "Three-dimensional information acquisition using a compound imaging system," *Opt. Rev.* **14**(5), 347–350 (2007).
- M. Cho et al., "Three-dimensional optical sensing and visualization using integral imaging," *Proc. IEEE* **99**(4), 556–575 (2011).
- M. Levoy, "Light fields and computational imaging," *Computer* **39**(8), 46–55 (2006).
- X. Xiao et al., "Advances in three-dimensional integral imaging: sensing, display, and applications [invited]," *Appl. Opt.* **52**(4), 546–560 (2013).
- G. Passalis et al., "Enhanced reconstruction of three-dimensional shape and texture from integral photography images," *Appl. Opt.* **46**(22), 5311–5320 (2007).
- S. Seitz et al., "A comparison and evaluation of multi-view stereo reconstruction algorithms," in *2006 IEEE Computer Society Conf. Computer Vision and Pattern Recognition*, Vol. 1, pp. 519–528, IEEE, New York, NY (2006).
- G. Vogiatzis et al., "Multiview stereo via volumetric graph-cuts and occlusion robust photo-consistency," *IEEE Trans. Pattern Anal. Mach. Intell.* **29**(12), 2241–2246 (2007).
- Y. Furukawa and J. Ponce, "Accurate, dense, and robust multiview stereopsis," *IEEE Trans. Pattern Anal. Mach. Intell.* **32**(8), 1362–1376 (2010).
- M. Pharr and G. Humphreys, *Physically Based Rendering: From Theory to Implementation*, Morgan Kaufmann, Burlington, Massachusetts (2010).
- L. Cao et al., "Geometric co-calibration of an integrated small animal spect-ct system for intrinsically co-registered imaging," *IEEE Trans. Nucl. Sci.* **56**(5), 2759–2768 (2009).
- X. Chen et al., "Generalized free-space diffuse photon transport model based on the influence analysis of a camera lens diaphragm," *Appl. Opt.* **49**(29), 5654–5664 (2010).
- J. A. Guggenheim et al., "Multi-modal molecular diffuse optical tomography system for small animal imaging," *Meas. Sci. Technol.* **24**(10), 105405 (2013).
- J. Kleinberg and E. Tardos, "Approximation algorithms for classification problems with pairwise relationships: metric labeling and Markov random fields," *J. ACM* **49**(5), 616–639 (2002).
- Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," *IEEE Trans. Pattern Anal. Mach. Intell.* **23**(11), 1222–1239 (2001).
- K. N. Kutulakos and S. M. Seitz, "A theory of shape by space carving," *Int. J. Comput. Vis.* **38**(3), 199–218 (2000).
- J. Banks, M. Bennamoun, and P. Corke, "Non-parametric techniques for fast and robust stereo matching," in *Proc. IEEE TENCON'97. IEEE Region 10 Annual Conference. Speech and Image Technologies for Computing and Telecommunications*, Vol. 1, pp. 365–368, IEEE, Brisbane, Queensland, Australia (1997).
- Y. Boykov and V. Kolmogorov, "An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision," *IEEE Trans. Pattern Anal. Mach. Intell.* **26**(9), 1124–1137 (2004).
- N. D. Campbell et al., "Using multiple hypotheses to improve depth-maps for multi-view stereo," in *Computer Vision—ECCV 2008*, pp. 766–779, Springer, Marseille, France (2008).

Biographies of the authors are not available.