Micro-Baseline Structured Light

Vishwanath Saragadam*, Jian Wang, Mohit Gupta, and Shree Nayar
NYC Research Lab, Snap Inc.
vishwanathsrv@cmu.edu, {jwang4, snayar}@snap.com, mohitg@cs.wisc.edu

Abstract

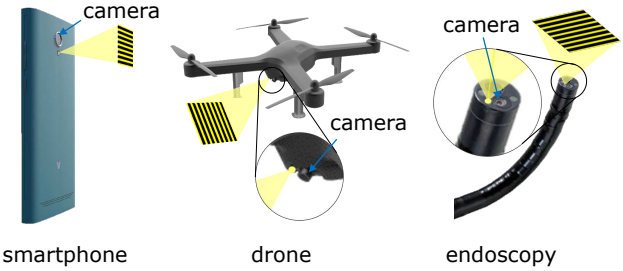
We propose Micro-baseline Structured Light (MSL), a novel 3D imaging approach designed for small form-factor devices such as cell-phones and miniature robots. MSL operates with small projector-camera baseline and low-cost projection hardware, and can recover scene depths with computationally lightweight algorithms. The main observation is that a small baseline leads to small disparities, enabling a first-order approximation of the non-linear SL image formation model. This leads to the key theoretical result of the paper: the MSL equation, a linearized version of SL image formation. MSL equation is under-constrained due to two unknowns (depth and albedo) at each pixel, but can be efficiently solved using a local least squares approach. We analyze the performance of MSL in terms of various system parameters such as projected pattern and baseline, and provide guidelines for optimizing performance. Armed with these insights, we build a prototype to experimentally examine the theory and its practicality.

1. Introduction

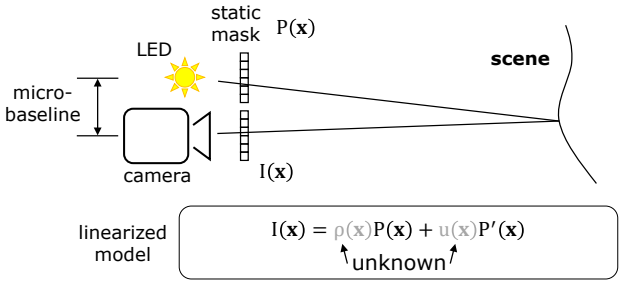
Structured light (SL) is one of the most widely used 3D imaging techniques. Due to simple hardware and high depth resolution, SL is suitable for diverse applications, such as 3D modeling, biometrics, gaming, and user interfaces. A typical SL system consists of a projector that illuminates the scene of interest with coded patterns, and a camera that images the scene. Depth at each scene point is computed via triangulation, after determining correspondences between projector and camera pixels.

Like all triangulation-based methods, SL requires a large baseline between projector-camera pair for reliable depth estimation [15, 31]. However, there are several emerging applications that increasingly rely on small form-factor devices. Imagine a cell-phone based augmented reality (AR) application building a 3D model of its surroundings [19, 11, 29], or a micro-drone [22] identifying obstacles such as tree twigs while flying through a dense forest, or a 3D camera on a thin endoscope [28] aiding a complex surgery inside the human body (Figure 1 (a)). Due to their small size, these devices cannot accommodate large baselines, making it challenging for existing SL approaches to recover accurate scene geometry.

In addition to small size, many of these devices are further constrained by low computation power, as well as low device complexity. For example, while it may be possible



(**a**) Challenges faced by structured light on constrained devices: limited baseline, low device complexity, low computational power



(**b**) Proposed micro-baseline structured light

Figure 1: Structured light on constrained devices. Small form-factor devices have severe constraints such as (i) small projector-camera baseline, (ii) low-cost hardware with limited capabilities, and (iii) low computational power. (b) We propose Micro-baseline Structured light, a novel approach that linearizes the SL image formation model. This enables fast and accurate depth recovery with low complexity projection devices, and lightweight computations.

*This work was done at Snap's NYC Research Lab and supported by Snap Inc. Vishwanath is with the Department of Electrical and Computer Engineering at Carnegie Mellon University, and Mohit Gupta is with the Department of Computer Science at the University of Wisconsin-Madison.

to have a simple, single-pattern projector on these devices that uses a static mask or a diffractive optical element (Figure 1 (a)), it may be considerably more challenging to place a full projector that can dynamically change the projected patterns. This precludes a large family of SL techniques, called multi-shot SL [23, 3], that require projecting several coded patterns sequentially. However, approaches that use a single pattern often require complex correspondence matching or learning algorithms, that may be prohibitively expensive given a limited computational budget and lack of sufficient data.

We propose a novel SL approach called Micro-baseline Structured Light (MSL), which is tailored to such highly constrained devices, thereby opening up the possibility of deploying SL on small, low-power and low-complexity devices. MSL works under the constraint of a small (micro) projector-camera baseline, as shown in Figure 1 (b), and is based on the following observation: small baseline leads to small disparities between projector and camera pixels. Our key theoretical insight is that with small disparities, the structured light image formation model, which is otherwise non-linear in the unknowns (depths and albedos), can be *linearized via a first order approximation*. This leads to the derivation of a new linear SL constraint, the *micro-baseline structured light (MSL) equation*, which relates scene albedos and depths to the measured intensities.

Theoretical and practical performance analysis: The MSL equation is under-constrained with two unknowns (depth and albedo) at each pixel and hence is impossible to solve without additional constraints. We design a local least-squares approach, which assumes local constancy of depth and albedo and expresses the MSL equation as a linear system with a 2×2 matrix, called the *MSL matrix*. Depth is recovered by inverting the 2×2 MSL matrix for each pixel, which can be done in a computationally efficient manner. The depth recovery performance of MSL can be characterized in terms of the properties of the MSL matrix, which, surprisingly, depend only on the projected pattern. Based on this observation, we provide practical guidelines for designing the projected pattern for the MSL matrix to be reliably invertible.

Real-world applicability and limitations: In order to relax the restrictive assumption of locally constant shape and albedo made for theoretical analysis, we design a practical MSL approach that captures *two images* - one with projected pattern, and another by turning off the projector. The no-pattern image is used as a guide image [10] to recover depth for scenes with complex texture, while maintaining low computational and hardware complexity so that the projector can still be implemented with a single, static mask. However, due to the extra image, MSL is not strictly a single-shot technique and may suffer from high-speed motion artifacts.

Scope: MSL is specifically tailored to, and optimized for constrained devices with small baseline, low-power, and low complexity; it is not meant to be a general-purpose SL technique that can replace existing approaches. Indeed, in unconstrained scenarios (large baseline, availability of data or high computing power, ability to project multiple patterns), existing SL techniques will achieve better performance as compared to MSL. However, on devices with strong constraints, MSL offers a light-weight solution while achieving good accuracy.

2. Related Work

Structured light coding techniques: Broadly, SL techniques can be classified into multi-shot and single-shot methods [25]. Multi-shot techniques such as light striping [2], Gray coding [23], and sinusoidal phase-shifting [3] estimate shape by projecting multiple patterns in quick succession. These techniques can recover high-precision depths with computationally simple decoding algorithms, but require complex projection devices (e.g., LCD, DMD) that can dynamically change the projected patterns, making them unsuitable for dynamic scenes and low-complexity devices such as cell-phones.

Single-shot techniques project only a single pattern, and rely on coding of projector correspondences in intensity [32], color [8, 13], or in a local neighborhood [9, 20, 14]. Single pattern techniques are well suited for dynamic scenes; however, these techniques often use computationally complex decoding algorithms, requiring dedicated hardware [1] for real-time performance. There are single-shot methods with relatively simple decoding (e.g., Fourier Transform Profilometry (FTP) [30]), but they make strong assumptions on the scene's texture and depths.

Real-time SL systems: There are approaches for performing high-speed (1000 fps) SL, either with high-cost high-speed cameras [12] that cannot be ported to mobile setups, or more recently, with learning-based approaches such as HyperDepth [24] and UltraStereo [7]. With sufficient data, and dedicated hardware such as Kinect [1], these methods have shown to be fast and accurate. Our goal is different. We aim to develop a method with a simple, analytical, closed-form decoding approach that leverages a differential formulation of conventional SL equation under small baseline constraint. An interesting future research direction is to augment MSL with data-driven techniques to potentially further increase accuracy and speed.

3. Structured Light Preliminaries

We start by describing the image formation model for an SL system, in order to understand the role of the projector-camera baseline in a structured light system.

Image formation model. Consider a projector-camera pair as shown in Fig. 1 (b). We assume a rectified projector-camera configuration, where the projector and camera centers are shifted horizontally by B units. We further assume that both the projector and camera have the same spatial resolution and focal length f. These assumptions are made only for ease of exposition; the presented analysis and the techniques are valid for general configurations and system parameters.

Let $P(x, y)$ be the projected pattern. The image intensity captured at a camera pixel (x, y) is given by:

$$I(x, y) = A(x, y) + \rho(x, y)P\left(x + u(x, y), y\right), \quad (1)$$

where $A(x, y)$ is contribution from ambient light, $\rho(x, y)$ is the reflectivity term that encapsulates scene texture and BRDF, projector brightness and intensity fall-off, and $u(x, y) = \frac{Bf}{z}$ is the disparity term that encodes scene depths z. Note that Eq. (1) is a single, non-linear equation in three unknowns: A, ρ, and disparity u.

Estimating u: The disparity u is computed by finding correspondences between projector and camera pixels. Multi-shot techniques estimate correspondences by uniquely coding each projector column. In contrast, single-shot techniques encode the correspondences in local neighborhood, intensity or frequency of a single projected pattern, but rely on global reasoning and complex optimization algorithms to decode accurately.

In the next section, we design a technique that requires projection of a single pattern (but capture of two images), and yet, is computationally cheap so that it can be efficiently implemented on power limited systems. Further, although conventional SL systems use as large a baseline as possible, the proposed technique is tailored to small form-factor devices that allow only a small (micro) baseline between projector and camera.

4. Micro-baseline Structured Light

We now consider an SL system with a small projector-camera baseline B, for example, that allowed by a cell-phone or a micro-drone geometry. Since the disparity $u = \frac{Bf}{z}$ at a pixel is proportional to B, given a scene depth z, and camera focal length f, a small B results in small disparities. Our *key observation* is that in such an SL system with small disparities, the image formation model (Eq. 1) can be linearized via a Taylor first order approximation:

$$\begin{aligned} & I(x, y) \approx A(x, y) + \rho(x, y)(P(x, y) + u(x, y)P'(x, y)) \\ \implies & I(x, y) \approx A(x, y) + \rho(x, y)P(x, y) + \widetilde{u}(x, y)P'(x, y), \end{aligned} \quad (2)$$

where $P'(x, y) = \frac{\partial P}{\partial x}(x, y)$ and $\widetilde{u}(x, y) = \rho(x, y)u(x, y)$.

Non-pattern image: Eq. (2) is a linear equation, albeit in three unknowns. To reduce the number of unknowns, we capture an extra image $I_{\text{no pattern}}$ with no projected pattern, by switching off the projector. This extra image measures the effect of ambient light, i.e., $I_{\text{no pattern}}(x, y) = A(x,y)$. We then subtract $I_{\text{no pattern}}$ from the pattern image $I_{\text{pattern}}(x, y)$ to remove the ambient component. The resulting difference image $I(x, y) = I_{\text{pattern}}(x, y) - I_{\text{no pattern}}(x, y)$ is given as:

$$\boxed{I(x, y) \approx \rho(x, y)P(x, y) + \widetilde{u}(x, y)P'(x, y).} \quad (3)$$

We call Eq. (3) the Micro-baseline Structured Light (**MSL**) constraint. Instead of the original non-linear SL equation, the MSL constraint is linear in $\widetilde{u}(x, y)$, and $\rho(x, y)$, and hence the projector-camera correspondences can be solved efficiently with low computational complexity. While removal of ambient light requires an extra image, it requires only turning off the projector. This maintains the simplicity of the projector that still needs to project only a single, static pattern. Henceforth, we use Eq. (3) for upcoming analyses.

Local least squares. While Eq. (3) is linear, the two unknowns, $\rho(x, y)$ and $\widetilde{u}(x, y)$ make it an under-constrained linear system. One way to further regularize it is to assume that both albedo and disparity are constant in a small neighborhood[1] which increases the number of equations per unknown. Consider a window of $n \times n$ pixels in the camera image. By assuming a constant albedo ρ_0 and disparity u_0 within this window, the modified MSL constraint equations for pixels $(x_l, y_m),\ 1 \leq l, m \leq n$ can be written as:

$$I(x_l, y_m) = \rho_0 P(x_l, y_m) + \rho_0 u_0 P'(x_l, y_m). \quad (4)$$

the set of linear equations have a unique solution for $n > 1$ and can be obtained using linear least squares.

We now provide a simple algorithm to compute albedo and disparity within the window. Let $\mathbf{i}_c = [I(x_1, y_1), \cdots, I(x_n, y_n)]^\top$ be the vector of camera measurements. Similarly, let $\mathbf{p} = [P(x_1, y_1), \cdots, P(x_n, y_n)]^\top$ be the vector of projector intensities, and $\mathbf{p}_x = [P'(x_1, y_1), \cdots, P'(x_n, y_n)]^\top$ be vector of derivative of pattern along x-axis. The set of equations within the window can then be written as,

$$\mathbf{i}_c = \rho_0 \mathbf{p} + \widetilde{u}_0 \mathbf{p}_x = \underbrace{\begin{pmatrix} \mathbf{p} & \mathbf{p}_x \end{pmatrix}}_{A} \begin{pmatrix} \rho_0 \\ \widetilde{u}_0 \end{pmatrix}, \quad (5)$$

where $\widetilde{u}_0 = \rho_0 u_0$ and $A = \begin{pmatrix} \mathbf{p} & \mathbf{p}_x \end{pmatrix}$. Using a least squares approach requires multiplication by $A^\top$ on both

[1] These assumptions are made for ease of analysis, and will be relaxed later in the paper.

sides, which gives us, $A^\top \mathbf{i}_c = A^\top A \begin{pmatrix} \rho_0 \\ \widetilde{u}_0 \end{pmatrix}$,

$$\begin{pmatrix} \mathbf{p}^\top \mathbf{i}_c \\ \mathbf{p}_x^\top \mathbf{i}_c \end{pmatrix} = \underbrace{\begin{pmatrix} \mathbf{p}^\top \mathbf{p} & \mathbf{p}_x^\top \mathbf{p} \\ \mathbf{p}^\top \mathbf{p}_x & \mathbf{p}_x^\top \mathbf{p}_x \end{pmatrix}}_{M_{\text{MSL}}} \begin{pmatrix} \rho_0 \\ \widetilde{u}_0 \end{pmatrix} \tag{6}$$

We call M_{MSL} the Micro-baseline Structured Light matrix, or **MSL matrix** in short. Eq. (6) states that with the proposed approach, estimation of disparities requires simply inverting a 2×2 matrix at every pixel, which is computationally cheap, and can be easily parallelized.

Relation to differential methods. The above analysis bears similarities to recent differential approaches designed for photometric stereo [5] and light-field-based motion estimation [18]. These approaches also linearize an otherwise *hard-to-solve* non-linear problem, resulting in tractable analysis and solutions. In the same spirit, MSL can be considered as a differential version of SL.

Relation to optical flow. It is also worth noting that the MSL matrix is similar to the *structure tensor* in Lucas-Kanade tracker [16]. Similar linearization of disparity/optical flow and formulating a 2×2 matrix have been explored before in the context of stereo vision [6, 21]. A key difference between structure tensor and the MSL matrix is that the MSL matrix depends only on the projected pattern and its derivative. Hence, invertiblity of the MSL matrix can be analyzed *only in terms of the properties of the projected pattern*, and not the scene.

5. Invertibility of MSL Matrix

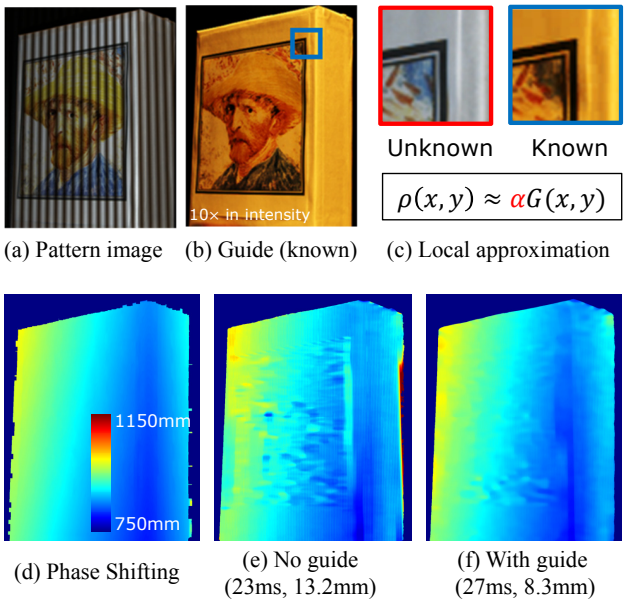
We now discuss the requirements for the MSL matrix to be invertible and well-conditioned. To analyze its invertibility, we look at its determinant,

$$\begin{aligned} \Delta_{\text{MSL}} &= (\mathbf{p}^\top \mathbf{p})(\mathbf{p}_x^\top \mathbf{p}_x) - (\mathbf{p}^\top \mathbf{p}_x)^2 \\ &= \|\mathbf{p}\|^2 \|\mathbf{p}_x\|^2 (1 - \cos^2(\theta)), \end{aligned} \tag{7}$$

where θ is the angle between $\mathbf{p}$ and $\mathbf{p}_x$. Eq. (6) is solvable if and only if $\Delta_{\text{MSL}} \neq 0$. This is ensured by displaying a pattern that is not uniformly zero ($\mathbf{p} \neq 0$), not a constant, ($\mathbf{p}_x \neq 0$), or not any pattern such that $\theta \neq 0, \pi$. The condition $\theta = 0$ or π implies $\mathbf{p} \propto \mathbf{p}_x$, which is only satisfied by exponential patterns[2].

Proposition 5.1 (Necessary and sufficient condition for invertibility)**.** *The MSL matrix is invertible iff the projected pattern is not horizontally constant or an exponential pattern of the form* $P(x, y) = c_1 e^{c_2 x}$.

[2]See supplementary material for derivation.



(a) Pattern image (b) Guide (known) (c) Local approximation

(d) Phase Shifting (e) No guide (23ms, 13.2mm) (f) With guide (27ms, 8.3mm)

Figure 2: Handing high-frequency texture. Most real scenes have textured objects. To account for high frequency texture in scene, the guided MSL approach uses a (b) no-pattern image of the scene for robust depth recovery. We model the unknown object texture as a simple linear scaling of the guide image, shown in (c), which leads to significant improvement in depth estimation, while maintaining the low decoding complexity.

This proposition states that by projecting a pattern that is not a constant or an exponential function theoretically guarantees that the MSL equation has a solution. Next, we discuss the *stability* of the solution, an important consideration in the presence of noise.

Invertibility with noisy measurements. In order to stably invert the MSL matrix in the presence of noise, its two eigen values must be maximized. The two eigen values are,

$$\lambda_1 = \|\mathbf{p}\| \sqrt{1 - \cos^2(\theta)}, \lambda_2 = \|\mathbf{p}_x\| \sqrt{1 - \cos^2(\theta)}, \tag{8}$$

and are directly proportional to $\|\mathbf{p}\|$, $\|\mathbf{p}_x\|$ and $1 - \cos^2(\theta)$. Intuitively $\|\mathbf{p}\|$ is proportional to the projector's brightness and $\|\mathbf{p}_x\|$ is proportional to local gradients of the pattern. While simultaneously maximizing both quantities is a complex optimization problem, we observe that the third term, $1 - \cos^2(\theta)$ is maximized when $\theta = 90°$, which is satisfied by continuous periodic signals. Therefore, stable solution to the MSL equation is achieved when the projected pattern is periodic. The pattern period may not align with analysis window. However, in practice, as shown in our experiments, depth estimates are robust to small misalignment.

6. Handling Texture Edges

The analysis so far assumes that both albedo and disparity within a small window are constant. While this simplifies the analysis, these assumptions are severely restrictive, as most real-world scenes consist of textured objects. To handle such realistic scenes, we propose a simple modification to MSL that leverages the no-pattern image captured for removing the ambient component (Section 4). The no-pattern image acts as a *guide image* $G(x, y)$ to further regularizes MSL decoding, especially in the presence of high-frequency scene texture, as illustrated in Figure 2. We call this approach *guided MSL*.

Guided MSL: Instead of locally smooth albedo, we assume that the albedo is locally a scaled version of the guide image. Such an assumption holds when the window is small and the objects are not highly specular. Specifically, we assume the following relationship between albedo and the guide image, $\rho(x_l, y_m) \approx \alpha_0 G(x_l, y_m)$ within the analysis window. Then, the MSL equation is modified as,

$$I(x_l, y_m) = \alpha_0 P_1(x_l, y_m) + \widetilde{u}_0 P_2(x_l, y_m), \qquad (9)$$

where $P_1(x_l, y_m) = G(x_l, y_m)P(x_l, y_m)$ and $P_2 = G(x_l, y_m)P'(x_l, y_m)$. The above equation is same as standard MSL except with different expressions for pattern and derivative and hence can be solved equally efficiently. This is similar to guided filtering [10] where the image to be filtered is locally expressed as a linear scaling of the guided image, plus an offset. To keep computation simple, we assume the albedo is only a scaled version of the guided image. Figure 2 illustrates the advantage of guided MSL over standard MSL by computing depth of a highly textured object. Guided MSL considerably improves the accuracy of MSL-based depth recovery, with practically no computational overhead, thus expanding the scope of the proposed approaches. Henceforth, all our results are computed using the guided MSL approach.

7. Practical Considerations for MSL

We now discuss several practical design choices for an MSL system, including designing the projected pattern, and various system parameters, including the baseline.

7.1. Choice of Projected Pattern

As discussed in Section 5, a periodic pattern guarantees that the MSL matrix is invertible, albeit under the assumption of locally constant depths (disparity). In this section, we aim to design patterns that are applicable to a broader range of scene geometries. Instead of assuming locally constant depths, we model the local geometry in a small window with locally planar depth variation, which is approximated well by linear disparity. Can we design patterns that accurately estimate depth for locally linear disparity?

It can be shown that the solution of MSL equation is unbiased for locally linear geometry *iff* $|P'(x,y)| = f(y)$, i.e., the pattern has a constant derivative[3]. The only such pattern is the intensity ramp, which is not desirable due to a small derivative. Instead, we choose a *piece-wise linear* pattern within the window, which ensures approximately constant derivative in the window, and has large magnitude of local gradients. A periodic symmetric triangle is such a pattern, which is continuous and has a constant derivative almost everywhere. All the results shown henceforth in the paper are with a triangular pattern; comparisons with other patterns are shown in the supplementary material.

7.2. Choice of System Parameters

Next, we evaluate the effect of various MSL system parameters (baseline, pattern period and window size) on depth accuracy. Since capturing real data with a wide range of parameters is infeasible, we instead evaluate via simulations using several scenes from the Middlebury dataset [26, 27]. For all our simulations, we chose a system configuration that closely resembled a mobile phone platform. Specifically, focal length of camera and projector was set to $f = 25$mm, and scenes were simulated over a depth range of $100 - 2000$mm. We also added readout and photon noise to understand the effect of noise on performance of MSL. Based on these simulations, we provide guidelines for choosing parameters for achieving high performance, which we then demonstrate via real experiments.

Window size. The size of window dictates the accuracy and spatial resolution of estimated depth map. This is similar to the effect of window size in block-matching based stereo techniques. To quantify the effect of window size, Figure 3 shows a plot of accuracy for varying window size for a $20px$ period. Intuitively, a smaller window does not capture sufficient spatial information, leading to ambiguity in correspondence estimation. In contrast, a larger window leads to spatial smoothening, leading to loss in resolution. As noted in section 5, the appropriate window size is close to the pattern period, and hence a window size of $20px$ leads to highest accuracy across all scenes.

Pattern period. For a fixed baseline, a smaller period accurately captures small depth variations, but limits the depth range. In contrast, a larger period enables a larger depth range, but with a lower depth resolution. Specifically, given a desired depth range d_{min} and d_{max}, the disparity range is given by $\Delta = Bf\left(\frac{1}{d_{\text{min}}} - \frac{1}{d_{\text{max}}}\right)$. To ensure unambiguous solution and local linear approximation to hold, the period

[3] See supplementary material for further details.

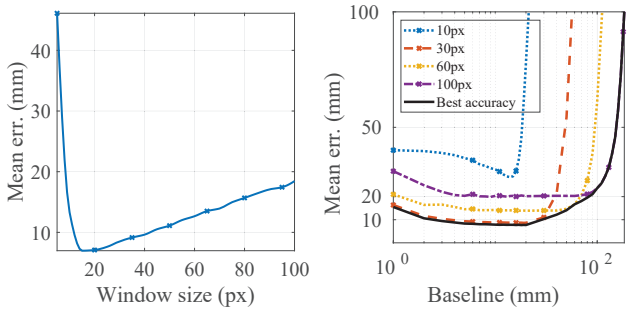


Figure 3: Accuracy of MSL vs. window size and pattern period. (a) Performance (depth errors) of MSL for varying window size, for a $20px$ pattern period. A window size comparable to pattern period achieves the lowest error. (b) Performance for various pattern periods as a function of baseline. The optimal period increases with increasing baseline. We choose the appropriate period for each baseline, giving us the best accuracy curve (black).

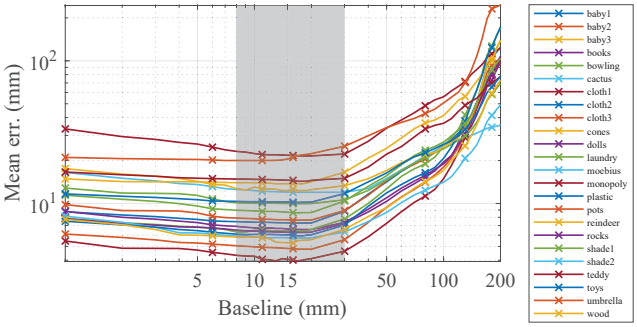
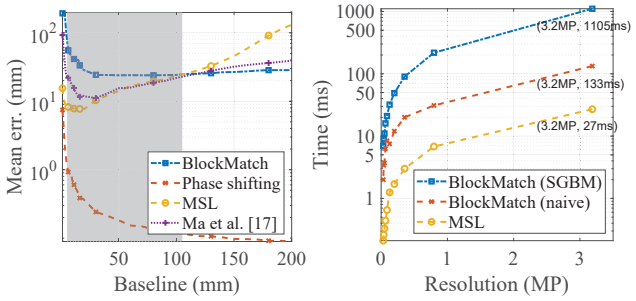


Figure 4: Accuracy of MSL vs. baseline. The plot shows accuracy of MSL vs. baseline using simulations on a few example scenes from the Middlebury dataset. Across scenes, MSL achieves the best performance with a baseline from 8 to 30mm.

n needs to satisfy $n \geq 2\Delta$. Figure 3 (b) illustrates the accuracy as a function of baseline for some representative pattern periods. Evidently, the period corresponding to the smallest error increases with increasing baseline.

Choosing an appropriate baseline. A small baseline ensures that the first order approximation holds, but suffers from triangulation error[31]. On the other hand, a large baseline requires a large window and hence local constancy assumption may not hold. Figure 4 shows simulation of accuracy as a function of baseline. For this analysis, given a baseline, we choose the pattern period that achieves the best accuracy for that baseline. We observe that MSL consistently achieves highest accuracy between $8 - 30$mm across a diverse set of examples. In practice, the exact choice of parameters depends on several additional factors, such as allowable resolution of projector, and defocus of camera and projector. We found that a baseline of 15mm led to the most accurate results, and hence our lab prototype was configured with this baseline (see Figure 6).



(a) Error vs baseline for various methods

(b) Runtime vs resolution on an Android phone

Figure 5: When should MSL be used? (a) Multi-pattern approaches such as phase-shifting consistently outperform MSL. Single pattern approaches (e.g., BlockMatch) can also outperform MSL for larger baselines; MSL achieves higher accuracy only for baseline smaller than 100mm, which can be considered the operating regime for MSL. (b) MSL is an order of magnitude faster than BlockMatch for a range of image sizes. Here runtime is tested on Android phone Pixel 2 XL. Therefore, MSL is suited for devices constrained by small baseline, low computing power, and inability to project multiple patterns.

7.3. When should MSL be used?

Under what device constraints is MSL more appropriate than existing SL techniques? MSL is targeted at platforms with restricted form factor, low hardware complexity and computational resources and hence *should not* be seen as an all-scenario, general-purpose alternative to existing ranging hardware. For example, if a system is capable of projecting multiple patterns, then phase shifting [3] works accurately even with a narrow baseline, as shown in Figure 5.

Similarly, if a system has sufficient computational resources and/or large baseline, existing single-shot techniques [20, 8, 33, 1, 7, 24] can achieve higher accuracy than MSL. Further, if the system is equipped with two cameras, one can rely on precise stereo-matching techniques [17] to obtain correspondences, albeit with heavy computational requirements. However when the device under consideration is small with limited hardware and computational capabilities, MSL promises a light weight solution. Figure 5 illustrates that MSL is more accurate than a block-matching for baselines smaller than 100mm, while being significantly faster. While the specific numbers depend on the exact configuration, MSL is suitable when baseline is small and only a single pattern can be projected.

8. Experiments

Hardware setup. Our setup consisted of a 1280×720 DLP projector (AAXA technologies) with $f = 8mm$ and a 2048×1536 machine vision camera (Basler acA2040-120uc) with $f = 12mm$. The differing focal lengths and

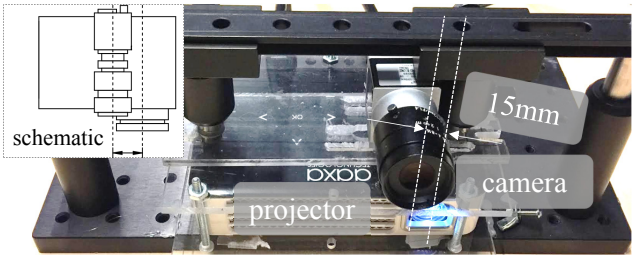


Figure 6: Hardware setup. We placed a machine vision camera on top of a DLP projector with a 15mm baseline in the horizontal direction. While the vertical baseline is large due to mechanical constraint, only the horizontal baseline matters since we use a vertically symmetric pattern.

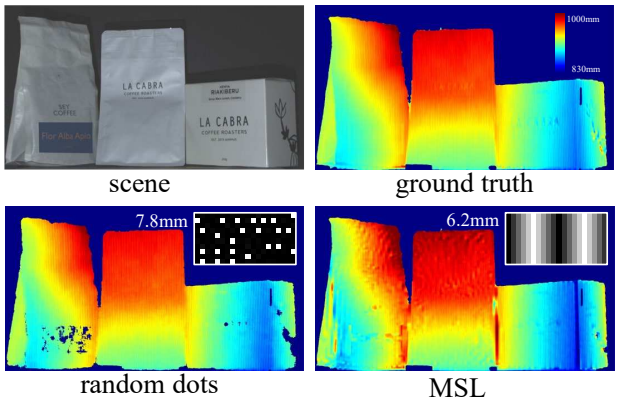
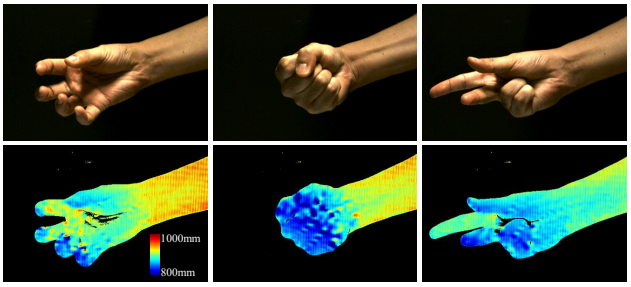


Figure 7: Comparison with single-pattern methods. We compare MSL against popular single-shot techniques that rely on pseudo-random patterns. Single-pattern techniques rely on search strategies with global reasoning and complex optimization. MSL can achieve comparable performance as existing single-shot approaches, while being considerably faster (27 ms vs. 133ms).

pixel sizes resulted in $\sim 2.5\times$ magnification of projector pattern in the camera image. The camera was placed above the projector with a horizontal baseline of $15mm$, as shown in Figure 6. The system also has a baseline along the vertical direction, which could not be avoided due to mechanical constraints. However, since we project a vertically symmetric pattern, only the horizontal baseline and disparities are considered; the vertical baseline does not affect the computation of horizontal disparities.

Ground truth. We captured ground truth depth information using phase shifting codes at five frequencies, corresponding to pattern periods of 1280px, 100px, 50px, 20px and 10px. Lower frequencies were used to unwrap higher frequency phases, which enabled a sub-pixel accurate estimate of disparity.

Run-time comparison on a cellphone. To evaluate real-time capabilities, we compared MSL with a stereo block matching algorithm with micro-baseline, by projecting random dots pattern. The results are visualized in Figure 7. Note that the projected pattern, as well as the decoding strategy are not optimized for narrow-baseline; our focus here is a comparison of timing complexity and not accuracy. Figure 5 (b) shows a comparison between runtime for varying image resolutions on an Android device, Google Pixel 2 XL, with existing matching-based approaches such as block matching and semi global method (SGBM) using OpenCV [4] implementation. The running time for a 3MP image for block matching and semi-global technique is 133ms and 1s respectively. In contrast, MSL is considerably faster at 27ms, suggesting the suitablity of MSL for mobile platforms.



(a) Frame 179 (b) Frame 285 (c) Frame 437

Figure 8: Video rate depth with MSL. Low computational and acquisition complexity of MSL enables capturing depth at video rates. We show select frames of a 15fps 3D imaging of hand gestures that was computed using the proposed technique.

Video sequence. One benefit of a light-weight SL technique is the ability to compute depth at video rate. To test this, we captured a sequence of images at 30fps for video rate 3D imaging. Alternate frames were captured without any pattern to use as guide image. The system outputs depth video and pattern-free video at 15fps which is computation-free (no need for pattern-scene separation) and is often useful for augmented reality. We show three representative depth frames in Figure 8. Note that the depth changes are clearly visible in various hand gestures. More importantly, the computational overhead for estimating depth is sufficiently small that it can be output in real-time, making MSL a compelling technique for mobile systems.

Experimental evaluation. Figure 9 shows MSL-based 3D recovery results on several scenes with varying geometric and texture complexity. All experiments were captured with triangular pattern of different periods to show various scenarios where MSL can be used. The first row shows results with planar objects of various texture complexity. Mannequin scene demonstrates MSL for a non-planar scene with limited texture. Note how the 3D model shows curves on the forehead as well as cheeks. Finally, the bas-relief scene shows the accuracy for small depth range but high spatial complexity. The depth map for the bas-

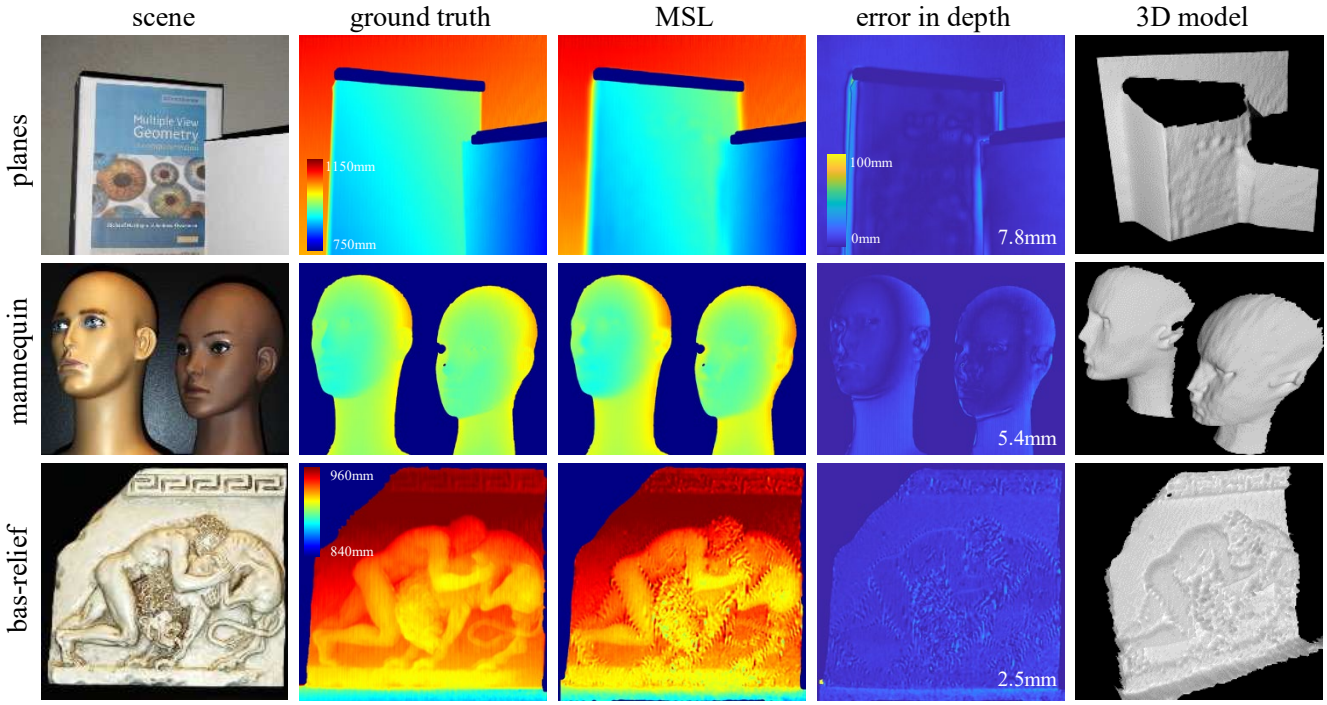
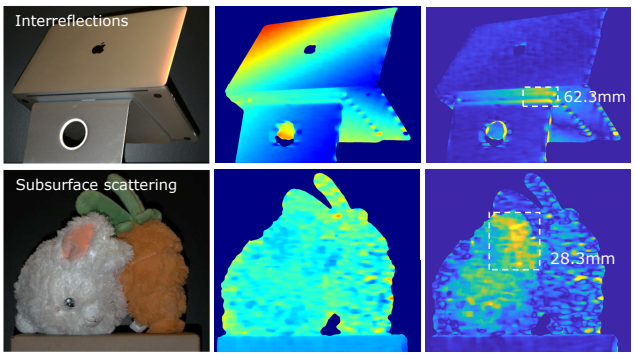


Figure 9: 3D imaging experiments with MSL. We demonstrate the performance of MSL on several scenes with varying complexity - a set of planes (top row), mannequin faces (middle row), and a bas-relief structure (bottom row). Scenes were captured with triangular patterns, with periods $20px$, $10px$ and $6px$, respectively. MSL accurately recovers depth across a wide range of experiments including planar and curved surfaces with textures and small depth variations.

relief scene was computed by displaying a pattern with $6px$ period which lead to a high spatial resolution. Note the accurate reconstruction of the fighter's thigh in the 3D model. In all cases, the error in depth is less than 8mm.

Failure cases. Since MSL is a local, windowed estimation technique, the computed depth at depth edges are smoothened, leading to adhesion to object boundaries (see planes scene in Figure 9). Performance similary degrades for highly textured objects, and complex geometry such as fine structures due to violation of local constancy assumption. Second, guided MSL assumes that the albedo within the window is a scaled version of image under ambient illumination. This assumption does not hold if spectrum of ambient illumination, projector illumination or reflectance, or surface normals have large variations, which leads to artifacts. Third, MSL relies on intensity to disparity with subpixel precision but is vulnerable to indirect illumination, and hence fails to perform well under interreflections or subsurface scattering (See Figure 10).



(a) Scene (b) Depth by MSL (c) Depth error

Figure 10: Effect of global illumination on MSL. Global illumination effects such as scattering and interreflections can lead to error in depth estimates. (Top) Interreflections between the laptop and the stand results in incorrect depths (62.3mm error). (Bottom) Depth errors due to subsurface scattering and complex texture.

9. Discussion

We propose a new SL technique designed to operate under the constraints of a narrow baseline, simple, low-cost hardware and low computing power. By linearizing the projector-camera correspondence equation, we showed that depth can be estimated efficiently using a local least-squares approach. We provided theoretical and practical guidelines for designing the projected patterns. MSL enables depth computation with limited hardware, making it an ideal technique for range imaging on cellphones, drones, micro-robots and endoscopes.

References

[1] Microsoft kinect. https://en.wikipedia.org/wiki/Kinect. Accessed: 2018-11-16. 2, 6

[2] Gerald J Agin and Thomas O Binford. Computer description of curved objects. *IEEE Trans. Computers*, (4):439–449, 1976. 2

[3] Dirk Bergmann. New approach for automatic surface reconstruction with coded light. In *Remote Sensing and Reconstruction for Three-dimensional Objects and Scenes*, volume 2572, pages 2–10. International Society for Optics and Photonics, 1995. 2, 6

[4] G. Bradski. The OpenCV Library. *Dr. Dobb's J. Software Tools*, 2000. 7

[5] Manmohan Chandraker, Jiamin Bai, and Ravi Ramamoorthi. On differential photometric reconstruction for unknown, isotropic brdfs. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 35(12):2941–2955, 2013. 4

[6] Julie Delon and Bernard Rougé. Small baseline stereovision. *J. Mathematical Imaging and Vision*, 28(3):209–223, 2007. 4

[7] Sean Ryan Fanello, Julien Valentin, Christoph Rhemann, Adarsh Kowdle, Vladimir Tankovich, Philip Davidson, and Shahram Izadi. Ultrastereo: Efficient learning-based matching for active stereo systems. In *IEEE Intl. Conf.Comp. Vision and Pattern Recognition (CVPR)*, 2017. 2, 6

[8] Zheng Jason Geng. Rainbow three-dimensional camera: New concept of high-speed three-dimensional vision systems. *Optical Engineering*, 35(2):376–384, 1996. 2, 6

[9] Paul M Griffin, Lakshmi S Narasimhan, and Soung R Yee. Generation of uniquely encoded light patterns for range data acquisition. *Pattern Recognition*, 25(6):609–616, 1992. 2

[10] Kaiming He, Jian Sun, and Xiaoou Tang. Guided image filtering. In *IEEE European Conf. Comp. Vision (ECCV)*, 2010. 2, 5

[11] Anders Henrysson, Mark Billinghurst, and Mark Ollila. Face to face collaborative ar on mobile phones. In *IEEE/ACM intl. Symposium on Mixed and Augmented Reality*. IEEE Comp. Society, 2005. 1

[12] Peisen S Huang, Chengping Zhang, and Fu-Pen Chiang. High-speed 3d shape measurement based on digital fringe projection. *Optical engineering*, 42(1):163–169, 2003. 2

[13] Changsoo Je, Sang Wook Lee, and Rae-Hong Park. High-contrast color-stripe pattern for rapid structured-light range imaging. In *IEEE European Conf. Comp. Vision (ECCV)*, pages 95–107. Springer, 2004. 2

[14] Hiroshi Kawasaki, Ryo Furukawa, Ryusuke Sagawa, and Yasushi Yagi. Dynamic scene shape reconstruction using a single structured light pattern. In *IEEE Intl. Conf. Comp. Vision and Pattern Recognition (CVPR)*, 2008. 2

[15] Jianyang Liu and Youfu Li. Performance analysis of 3-d shape measurement algorithm with a short baseline projector-camera system. *Robotics and Biomimetics*, 1(1):1, 2014. 1

[16] Bruce D Lucas and Takeo Kanade. An iterative image registration technique with an application to stereo vision. *Proc. of Imaging Understanding Workshop*, 1981. 4

[17] Ning Ma, Peng-fei Sun, Yu-bo Men, Chao-guang Men, and Xiang Li. A subpixel matching method for stereovision of narrow baseline remotely sensed imagery. *Math. Problems in Engineering*, 2017. 6

[18] Sizhuo Ma, Brandon Smith, and Mohit Gupta. 3d scene flow from 4d light field gradients. In *IEEE European Conf. Comp. Vision (ECCV)*, 2018. 4

[19] Mathias Mohring, Christian Lessig, and Oliver Bimber. Video see-through ar on consumer cell-phones. In *IEEE/ACM Intl. Symposium on Mixed and Augmented Reality*, 2004. 1

[20] Raymond A Morano, Cengizhan Ozturk, Robert Conn, Stephen Dubin, Stanley Zietz, and J Nissano. Structured light using pseudorandom codes. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 20(3):322–327, 1998. 2, 6

[21] Gareth Llewellyn Keith Morgan, Jian Guo Liu, and Hongshi Yan. Precise subpixel disparity measurement from very narrow baseline stereo. *IEEE Trans. Geoscience and Remote Sensing*, 48(9):3424–3433, 2010. 4

[22] Tomoyuki Mori and Sebastian Scherer. First results in detecting and avoiding frontal obstacles from a monocular camera for micro unmanned aerial vehicles. In *IEEE Intl. Conf. Robotics and Automation*, 2013. 1

[23] Jeffrey L Posdamer and MD Altschuler. Surface measurement by space-encoded projected beam systems. *Computer Graphics and Image Processing*, 18(1):1–17, 1982. 2

[24] Sean Ryan Fanello, Christoph Rhemann, Vladimir Tankovich, Adarsh Kowdle, Sergio Orts Escolano, David Kim, and Shahram Izadi. Hyperdepth: Learning depth from structured light without matching. In *IEEE Intl. Conf. Comp. Vision and Pattern Recognition (CVPR)*, 2016. 2, 6

[25] 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. 2

[26] Daniel Scharstein and Richard Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *Intl. J. Comp. Vision*, 47(1-3):7–42, 2002. 5

[27] Daniel Scharstein and Richard Szeliski. High-accuracy stereo depth maps using structured light. In *IEEE Int. Conf. Comp. Vision and Pattern Recognition (CVPR)*, 2003. 5

[28] Christoph Schmalz, Frank Forster, Anton Schick, and Elli Angelopoulou. An endoscopic 3d scanner based on structured light. *Medical Image Analysis*, 16(5):1063–1072, 2012. 1

[29] Gabriel Takacs, Vijay Chandrasekhar, Natasha Gelfand, Yingen Xiong, Wei-Chao Chen, Thanos Bismpigiannis, Radek Grzeszczuk, Kari Pulli, and Bernd Girod. Outdoors augmented reality on mobile phone using loxel-based visual feature organization. In *ACM Intl. Conf. Multimedia Information Retrieval*, 2008. 1

[30] Mitsuo Takeda and Kazuhiro Mutoh. Fourier transform profilometry for the automatic measurement of 3-d object shapes. *Appl. Optics*, 22(24):3977–3982, 1983. 2

[31] Marjan Trobina. Error model of a coded-light range sensor. *Technical Report*, 1995. 1, 6

[32] Piet Vuylsteke and André Oosterlinck. Range image acquisition with a single binary-encoded light pattern. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 12(2):148–164, 1990. 2

[33] Li Zhang, Brian Curless, and Steven M Seitz. Rapid shape acquisition using color structured light and multi-pass dynamic programming. In *IEEE Intl. Symposium on 3D Data Processing Visualization and Transmission*, 2002. 6