euspen's 26th International Conference & Exhibition, Krakow, PL, June 2026
www.euspen.eu

Design and calibration of a single-camera near-field photometric stereo head with tiltable illumination

Seunghun Oh[1], Jinsung Son[1], Kihyun Kim[2], Hyo-Young Kim[2], Moon-Gu Lee[3] and Jaehong Shim[2*]

[1] *Department of IT and Semiconductor Convergence Engineering, Graduate School, Tech University of Korea (TU Korea)*
[2] *Department of Mechatronics Engineering, Tech University of Korea (TU Korea)*
[3] *School of Mechanical Engineering, Ajou University, Korea*

**jhshim@tukorea.ac.kr*

Abstract

Inline inspection on fast-changing lines often involves glossy, rough, or multi-material surfaces; compact, high-speed, non-contact sensing is needed without line stops. Structured-light capture, laser-line sensing, and 3D depth sensors provide absolute depth but typically require multi-frame projection or scanning, leave gaps under occlusion or glare, and require additional image-to-point-cloud registration. A photometric stereo system uses a single RGB camera and four angle-adjustable LED panels to emphasise micro-topography for defect detection. The panels are arranged as opposing pairs along the x- and y-axes and controlled 100° of tilt by a z-axis linear-motor crank–slider. After a brief angle search, one image per direction is acquired; including LED switching and exposure, four views take about 0.5 s. A Convergence Radius (CR) directs near-field sources toward a virtual ring on the measurement plane, and the photometric model compensates for distance- and angle-dependent attenuation, stabilising results under small camera-to-part spacing changes. For 3D recovery and metric scale, CR ring calibration projects panel light through a slitted ring artefact and uses the shadows to centre the setup, set the Z scale, and level the reference plane. Outputs are a shape-enhanced image and per-pixel surface-slope maps in the camera image coordinates, enabling pixel-level decisions and direct use in learning pipelines without extra registration. Qualitative tests reveal micro-relief around Ra ≈ 3.2 µm and indicate that angle selection reduces unnecessary captures and computation while enabling fast inline detection of shallow defects on glossy and mixed-material surfaces. The system can be mounted on a robot end effector for multi-position inspection, with demonstrated use on weld seams, painted panels, and moulded or machined parts, detecting scratches, orange peel, and weld-line marks.

Keywords: photometric stereo; inline inspection; near-field lighting; calibration; specular surfaces.

1. Introduction

In fields where surface shape information is central to decision-making—such as precision manufacturing, cultural heritage documentation, and robot vision—the ability to reliably recover fine surface slopes and micro-relief strongly affects the level of quality, safety, and long-term preservation. In practice, however, surface reflectance varies widely (non-Lambertian and highly specular surfaces). As the light sources move closer to the target, the effective lighting direction and irradiance distribution can vary across pixels, and occlusions, shadows, and highlights may occur at the same time, which can easily increase the error of image-based 3D reconstruction. Photometric Stereo (PS) is a representative approach that estimates surface normal and albedo cues from multiple images captured at a fixed viewpoint (a single camera) under varying illumination [1], and it has the advantage that it allows quantitative comparison on both benchmark datasets and real captured data. Under near-field conditions, however, it is difficult to assume distant directional lighting; the solution becomes more sensitive to lighting geometry (position and direction) calibration, and this can increase instability in the normal estimation and depth integration stages [2, 3].

Illumination layouts used in 2D vision inspection can be broadly divided into bright-field and dark-field configurations. In bright-field mode, components with relatively small incidence angles with respect to the camera optical axis are emphasised, which is useful for observing overall surface reflectance. In dark-field mode, components with relatively large incidence angles are emphasised, so fine defects and micro-scattering features become more visible [6]. To enable efficient surface analysis across a wider range of surface properties, this study builds a single-device PS system that supports bright-field/dark-field switching and LED tilt control, and then defines the Convergence Radius (CR) so that the tilt state of four lights can be set intuitively by directly linking it to the target size. Here, CR is defined as "the diameter of the circle formed by the four points where the illumination axes (virtual lines normal to each LED panel passing through its centre) intersect the measurement plane." With this definition, users can reproduce a lighting state by specifying h (the distance from the light centre to the measurement plane) and a target CR, rather than relying on motor counts. In addition, the procedure includes capturing a half-ring calibration target and using the shadow boundaries formed on the ring to calibrate the lighting direction (or effective lighting vector), so that the height scale is consistent during reconstruction. Finally, to mitigate the known weakness of least-squares (LS) based PS to shadows, highlights, noise, and boundary mixing, we propose a robust regression-based PS method that interprets the swept image stack as a spatio-temporal volume (STV) and combines spatial (patch) weighting, edge-aware weighting, temporal outlier rejection, and the Huber loss.

The remainder of this paper is organised as follows. Section 2 presents the methodology, including the CR definition and the mathematical formulation for device control, the half-ring shadow-based calibration, and the STV-based robust PS regression. Section 3 describes the benchmark (DiLiGenT) and real data configuration and the evaluation protocol. Section 4 reports quantitative results (normal angular error and processing time) obtained by applying the proposed method. Section 5 provides the summary, conclusions, and future research directions.

2. Methodology

2.1. System overview and illumination mode

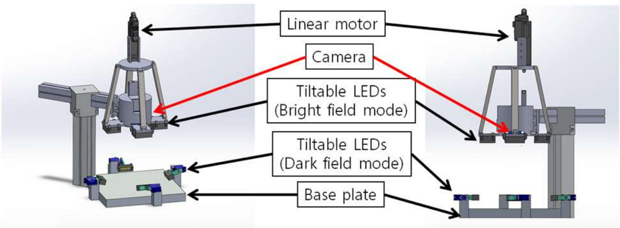

Figure 1. (System composition)

The bright-field illumination module proposed in this study is arranged vertically in four directions (east, west, south, and north) and is mounted to a cylindrical camera module. It performs a tilting motion while remaining fixed to the cylinder, driven by a linear-motor–crank–slider mechanism. The dark-field illumination module is also arranged in four directions (east, west, south, and north) near the measurement plane, and it performs rotational motion driven by a servo motor. Let m_0 denote the motor value at which the illumination axis has a reference direction. For a motor input m, the unit direction vector of the i-th illumination axis is defined as $n_i(m) \in \mathbb{R}^3$(where $\mathbb{R}$ denotes the set of real numbers and $\mathbb{R}^3$ denotes the three-dimensional real vector space). The 3D position of the i-th light centre is denoted by $C_i(m) = [C_{i,x}(m), C_{i,y}(m), C_{i,z}(m)]^\top$. The user-defined parameter h is the distance between the light centre and the measurement plane. When the measurement plane is defined as z=0, $\mathrm{h} = C_{i,z}(m)$, and this value can vary depending on the device configuration. Let τ be a scalar parameter along the axis. Then the line of each illumination axis is expressed as Eq. (1):

$$r_i(\tau) = C_i(m) + \tau n_i(m)$$

· · · (1)

Let $P_i(m)$ be the intersection po int of this line with the measurement plane z=0. When z-componet of$n_i(m)$, denoted by $n_{i,z}(m)$, satisfies $n_{i,z}(m) \neq 0$, the intersection is given by Eq.(2):

$$P_i(m) = C_i(m) - \frac{C_{i,z}(m)}{n_{i,z}(m)} n_i(m)$$

· · · (2)

If $P_{i,xy}(m) = [P_{i,x}(m), P_{i,y}(m)]^\top$, as the 2D coordinate on the measurement plane, the Convergence Radius (CR) is "the diameter of the circle formed by the four intersection points." Let R(m) be the radius of this circle, where R(m) is the radius of the best-fit circle to the four points $P_{i,xy}(m)$. Then CR is defined by Eq. (3):

$$\mathrm{CR(m)} = 2\mathrm{R(m)}$$

· · · (3)

When the device is ideally symmetric and the four intersection points lie on the same radius, (m) can be simplified as $\mathrm{R(m)} = \| P_{i,xy}(\mathrm{m}) \|$. If the symmetry is not perfect, (m) is estimated using a circle fitting procedure on the four points. The practical meaning of this definition is that the tilt state of the lights can be set by directly linking it to the object size (e.g., a target illuminated diameter). In other words, the user can specify a lighting state using the pair (h,) rather than motor counts.

To formulate this as a control problem, given a user-specified target $CR^{\backslash *}$, the motor input m is determined by the following inverse problem:

$$m^{\backslash *} = arg \min_m |CR(m) - CR^{\backslash *}|$$

· · · (4)

2.2. Illumination modes and acquisition protocol

This device provides both bright-field and dark-field modes using tiltable LEDs. In this study, the two modes are selected or combined depending on the surface reflectance characteristics of the object to form an image stack. Let the illumination index be $\mathrm{t} = 1,\ldots,\mathrm{T}$, and let the lighting state turned on at each frame be $\mathcal{L}(\mathrm{t})$. The captured observation image is denoted by $I_t(p)$ (where p is the pixel coordinate). The full image stack $\{I_t\}_{t=1}^{T}$, the mask M, and the per-frame lighting metadata (motor input m, BF/DF mode, and the lighting pattern) are then used as inputs to a unified processing pipeline.

2.3. Shadow-based calibration using half-ring target (CR ring calibration)

This study performs lighting calibration using a half-ring calibration target so that the height scale of the reconstruction is physically consistent. Assuming that the half-ring geometry (size and cross-sectional shape) is known, the shadow boundary produced by illumination i constrains both the lighting direction (effective lighting vector) and the object height scale. To this end, we first perform an illumination sweep. By comparing an "all-lights-on" image with each "single-light-on" image, we separate the additional shadow contribution of the i-th illumination and extract the shadow boundary $\partial \mathrm{S}_i$. The extracted $\partial \mathrm{S}_i$ is corrected using a homography H so that it is mapped consistently from the camera image coordinates to the measurement plane (world coordinates). This is an essential step for handling lighting direction and depth scale within the same coordinate system. Next, the lighting direction (or effective lighting vector) s_i is initialised using a kinematics-based estimate, and an optimisation problem is formulated to either fine-tune s_i or adjust a depth-scale parameter so that the observed shadow boundary on the ring is consistent with the known ring geometry. In other words, the purpose of this procedure is to use the constraints provided by "ring geometry + shadow boundary" to suppress the propagation of residual lighting-direction errors into the final reconstructed height in near-field conditions.

2.4. STV-based robust photometric stereo

In the basic linear model of photometric stereo, the observation at pixel p is approximated by Eq. (5) using an effective lighting vector $s_t \in \mathbb{R}^3$ and an unknown vector $\mathrm{g}(p) \in \mathbb{R}^3$ Here,. $I_t(p)$ represents an effective lighting term that includes near-field distance- and incidence-angle-dependent attenuation Eq (5):

$$I_t(p) \approx \mathrm{s}_t^\top \mathrm{g}(p)$$

$$\cdots (5)$$

where $g(p) = \rho(p)n(p)$, $\rho(p)$ is the albedo, $n(p)$ is the unit surface normal. Global least-squares (LS) initialisation can estimate $g(p)$ efficiently using all frames, but it is vulnerable to shadows, highlights, noise, and boundary mixing effects. The proposed STV–based method interprets the swept image stack $I(x, y, t)$ as a spatio-temporal volume (STV) and updates $g(p)$ robustly by jointly using neighbouring patch information and temporal information.

Let $N(p)$ denote a neighbourhood patch around pixel , and define a patch weight $w_{patch}(p, q)$ for $(q \in N(p))$ that reflects both a spatial weight and an edge-aware weight. Temporal outlier rejection is implemented as a binary weight $w_{temp}(q, t) \in \{0,1\}$,which excludes frames where the observation $I_t(q)$ is judged to be a shadow/highlight outlier. Finally, for the residual $r(p, q, t) = I_t(q) - s_t^{\top}$ we apply the Huber loss $\phi_\delta(\cdot)$ to suppress the influence of large residuals [5], where δ is the Huber threshold parameter. The vector $g(p)$ is then iteratively updated to minimise the robust objective in Eq. (6):

$$\min_{g(p)} \sum_{q \in N(p)} \sum_{t=1}^{T} w_{patch}(p, q) w_{temp}(q, t)\, \phi_\delta\left(I_t(q) - s_t^{\top} g(p)\right) + \lambda \parallel g(p) - \bar{g}(p) \parallel^2 \quad \cdots (6)$$

where $\bar{g}(p)$ is the LS initial estimate and λ is a regularisation coefficient that encourages convergence toward the initial estimate. After the iterations, albedo and normal are separated using Eq. (7):

$$\rho(p) = \parallel g(p) \parallel, \quad n(p) = \frac{g(p)}{\parallel g(p) \parallel} \quad \cdots (7)$$

3.Experiment evaluation

The experiments are organised in two tracks: benchmark data and real captured data. As the benchmark, we use the DiLiGenT dataset [4], which provides, for each object, an aligned image stack, lighting directions (or lighting vectors), an object mask, and a ground-truth normal map, enabling quantitative evaluation using normal angular error. The real captured data are acquired by performing illumination sweeps with the BF/DF and tilt-control device proposed in this study. The swept multi-illumination images are converted into a benchmark-like structure (image stack, lighting metadata, and mask) and processed using the same algorithm pipeline.

The half-ring target based calibration is included as a preprocessing step in the real-data pipeline to ensure consistency of the lighting direction (effective lighting vectors) and the depth scale. This calibration takes, as input, a ring image stack captured under four directional illuminations, separates the per-illumination shadow components (shadow separation), and refines the lighting-direction parameters using constraints from the ring geometry (known radius, thickness, height, etc.) and the observed shadow boundaries. Using the calibrated lighting parameters, we then compute a normalised relative height distribution within the ring region (ring height normalised). This output is used for (i) a qualitative check that the lighting-direction calibration behaves as intended and (ii) setting a reference for lighting parameters and scale to be used in real object reconstruction. Figure 2 summarises the half-ring calibration outputs: (a) ring images under four directional illuminations, (b) per-illumination shadow separation for shadow-boundary isolation, and (c) a normalised ring-height map for checking lighting-calibration consistency and depth-scale behaviour.

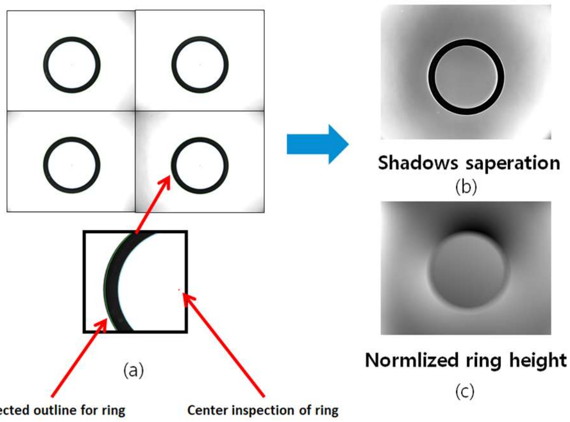

Figure 2. Half-ring target based calibration.

4.Results

Based on the reconstructions on the DiLiGenT dataset, we observed that the proposed STV-based robust regression tends to reduce the normal angular error for many objects compared with global LS initialisation. Over the 10-object average, the mean angular error decreased from 16.150 to 14.412, and the p90 angular error decreased from 31.509 to 28.786. For individual objects, readingPNG showed the largest improvement, with p90 decreasing from 47.031 to 38.373, and p90 reductions were also confirmed for ballPNG and bearPNG. However, for harvestPNG, p90 increased from 60.881 to 63.033, indicating that STV does not consistently reduce upper-quantile errors for all objects (Figure 5 and Table 1).

For real object measurements, a standard surface roughness specimen (12.5 μm R_a) and a laminated-coated copper specimen were captured. In Figure 3, we compare the normal maps produced by the baseline LS method and STVPS, and confirm that distinguishable surface patterns can be represented down to 3.2 μm R_a .We also captured data under both BF and DF conditions and found that DF can provide more effective results for certain surface features. In Figure 4, we present a glossy laminated-coated copper plate (thickness 3 mm) and compare reconstructions before and after applying the ring calibration.

Table 1 . Comparison with DiLigenT benchmark data

object	mean_LS	median_LS	p90_LS	mean_STV	median_STV	p90_STV
ballPNG	4.705	3.108	10.927	**3.451**	**2.629**	**6.536**
catPNG	8.781	6.990	15.159	**8.020**	**6.262**	**13.330**
pot1PNG	10.288	8.419	17.294	**9.401**	**7.113**	**15.691**
bearPNG	9.154	6.894	18.023	**7.571**	**5.517**	**14.389**
pot2PNG	16.708	13.652	33.520	**14.602**	**11.564**	**30.235**
buddhaPNG	15.448	11.202	32.521	**14.466**	**9.762**	**31.896**
gobletPNG	18.868	16.321	36.269	**17.339**	**14.484**	**32.952**
cowPNG	26.794	27.941	43.464	**24.846**	**25.195**	**41.424**
harvestPNG	31.568	26.319	60.882	**29.214**	**22.509**	**63.034**
readingPNG	19.194	11.305	47.032	**15.212**	**8.882**	**38.373**

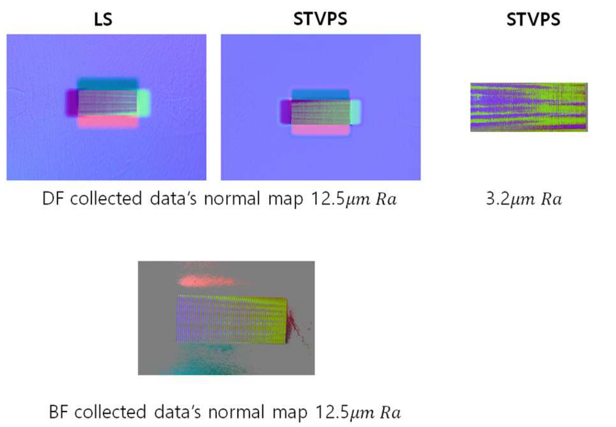

Figure 3. Real object application experiments-standard surface roughness specimen.

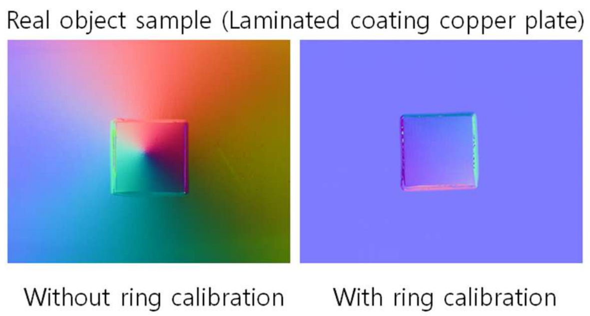

Figure 4. Real object application experiments- Laminated coated copper specimen.

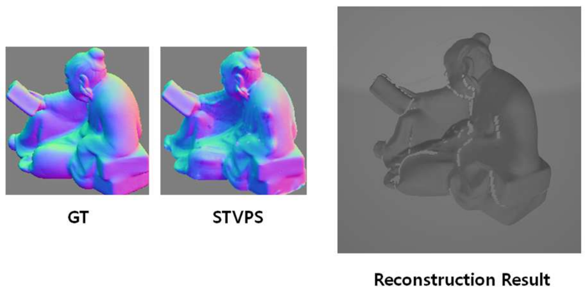

Figure 5. DiligenT benchmark normal map comparison and results

5. Conclution

This study proposed a methodology to address both illumination setting and reconstruction stability in near-field conditions using a single-device photometric stereo system that supports BF/DF switching and LED tilt control. First, the Convergence Radius (CR) was defined as the diameter of the circle formed by the four intersection points on the measurement plane, where the illumination axes of the east–west–south–north lights meet the plane. This definition provides a framework to control the lighting state in an interpretable and reproducible way using (h,CR), rather than motor counts. Second, a procedure was included to calibrate the lighting direction (or effective lighting vector) using shadow boundaries observed on a half-ring target, so that the height scale in reconstruction becomes physically consistent. Third, the instability of least-squares based PS, which is sensitive to shadows, highlights, and noise, was mitigated by introducing an STV-based robust regression that combines patch-based spatial weighting, edge-aware weighting, temporal outlier rejection, and the Huber loss. In the benchmark (DiLiGenT) batch evaluation, an overall tendency for the mean and p90 normal angular errors to decrease was observed.

The conclusions drawn from this work are as follows. In near-field PS, performance degradation is typically not caused by a single factor; instead, it often appears as a combined effect of "sensitivity to lighting-geometry calibration" and "observation outliers (shadows, highlights, and boundary mixing)." Therefore, practical applicability can be improved when device design (illumination mode switching and tilt control), interpretable illumination setting (CR), calibration (half-ring shadows), and robust estimation (STV regression) are integrated into a single, consistent pipeline. In particular, CR allows users to specify a lighting state in an intuitive way that matches the object size, which can support standardised operation and improve reliability for repeated measurements. However, STV introduces higher computational cost, and for some objects the upper-quantile error can increase. For this reason, application conditions and guidelines for parameter selection should be provided together.

Future work can be extended in three directions. First, the processing speed should be improved by reducing the computational cost of the STV iterative regression, for example through GPU acceleration, patch reuse, and multi-resolution strategies. Second, the relationship between CR, motor input, and tilt— i.e., $\mathrm{n}_i(\mathrm{m}), \mathrm{C}_i(\mathrm{m})$ —should be stabilised using measurement–based calibration that accounts for backlash, mechanical play, and elastic deformation, thereby improving control reproducibility for a target $\mathrm{CR}^{\backslash *}$. Third, to report the quantitative impact of half-ring calibration on "height-scale alignment" in real device data, the ring dimensions and the rules for shadow-boundary extraction should be fixed, and, if possible, comparison should be performed using external measurements or repeatability-based metrics.

Acknowledgement

This study was supported by Korea Institute for Advancement of Technology(KIAT) grant funded by the Korea Government(MOTIE) [(P0021527), (RS-2025-24536353) and (RS-2024-00409639) HRD Program for Industrial Innovation]

References

[1] Woodham R J 1980 Photometric method for determining surface orientation from multiple images Opt. Eng. 19 139–144

[2] Mecca R, Wetzler A, Bruckstein A M and Kimmel R 2014 Near Field Photometric Stereo with Point Light Sources SIAM J. Imaging Sci. 7 2732–2770

[3] Logothetis F, Mecca R and Cipolla R 2017 Semi-Calibrated Near Field Photometric Stereo Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR) 941–950

[4] Shi B, Wu Z, Mo Z, Duan D, Yeung S K and Tan P 2016 A Benchmark Dataset and Evaluation for Non-Lambertian and Uncalibrated Photometric Stereo Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR) 3707–3716

[5] Huber P J 1964 Robust Estimation of a Location Parameter Ann. Math. Statist. 35 73–101

[6] Forte P M F, Felgueiras P E R, Ferreira F P, Sousa M A, Nunes-Pereira E J, Bret B P J and Belsley M S 2017 Exploring combined dark and bright field illumination to improve the detection of defects on specular surfaces Opt. Lasers Eng. 88 120–128