Calibration of Multimodal 3D Structured-Light Systems Using Digital Features

EBERTO BENJUMEA,[1] RAÚL VARGAS,[1] FERNANDO QUINTERO,[1] RIGOBERTO JUAREZ-SALAZAR,[2] AND ANDRES G. MARRUGO [1,*]

[1]*Escuela de Ingeniería, Arquitectura & Diseño, Universidad Tecnológica de Bolívar, Cartagena, 130001, Bolivar, Colombia*
[2]*SECIHTI—Instituto Politécnico Nacional, CITEDI, Av. Instituto Politécnico Nacional 1310, Nueva Tijuana, Tijuana, B.C. 22435, Mexico*
[*]*agmarrugo@utb.edu.co*

Abstract: Calibration of multimodal 3D imaging systems that combine structured light with an additional modality typically relies on targets constructed with physical features that must be detectable by all imaging modalities. Such targets can be costly to produce and are prone to fabrication defects that degrade accuracy. Furthermore, reflections, light saturation, and the limited resolution of non-visible-range cameras complicate reliable feature detection. We present a calibration approach that uses digital features generated by a screen, a mirror, and an auxiliary camera—removing the need for specialized targets with physical features. This setup recovers the intrinsic parameters of the visible camera as well as the intrinsic and extrinsic parameters of both the projector and the additional modality camera. To illustrate our method, we employ a thermal camera, though the procedure extends readily to other imaging modalities. Experimental results show the proposed solution achieves a 0.07 mm root-mean-square error in 3D reconstructions, matching conventional techniques. By eliminating the requirement for physical features for targets, this approach reduces costs, avoids fabrication flaws, and simplifies multimodal feature detection.

1. Introduction

Accurate three-dimensional (3D) shape measurement using structured light is a pivotal technique in diverse fields, including industry, cultural heritage, scientific research, and medicine [1–4]. By integrating structured light with thermal, infrared, ultraviolet, or multispectral cameras, multimodal systems broaden their applicability and enrich the data available for both classical and deep learning-based methods [5–7]. Consequently, these combined approaches not only improve feature detection but also allow for the precise spatial localization of multimodal features on a surface.

Reliable 3D measurements in multimodal systems hinge on accurate camera calibration, which typically involves the use of calibration targets [8–10]. Planar targets—commonly checkerboards or circle grids—are especially popular in structured-light configurations due to their simplicity, well-defined geometry, and ease of feature detection [11]. By providing precise reference points for subpixel refinement, these targets facilitate the estimation of each camera's intrinsic parameters (e.g., focal length and lens distortion) as well as extrinsic parameters that define spatial relationships among sensors [11]. Recent efforts continue to refine these procedures, as accurate calibration remains vital for high-precision 3D reconstruction and quality assurance [12].

Despite the benefits of planar targets, applying them to multimodal 3D structured-light calibrations remains troublesome because each modality imposes special acquisition conditions and often requires costly, customized targets. Metallic surfaces, for instance, can produce specular reflections that saturate the visible camera, while low-resolution or spectrally sensitive cameras (e.g., thermal or hyperspectral) struggle to detect features [5, 7, 11, 13]. Disparate fields of view among cameras further complicate calibration, since a target designed for one modality may not appear at suitable resolution in another. Consequently, multi-material or machined targets

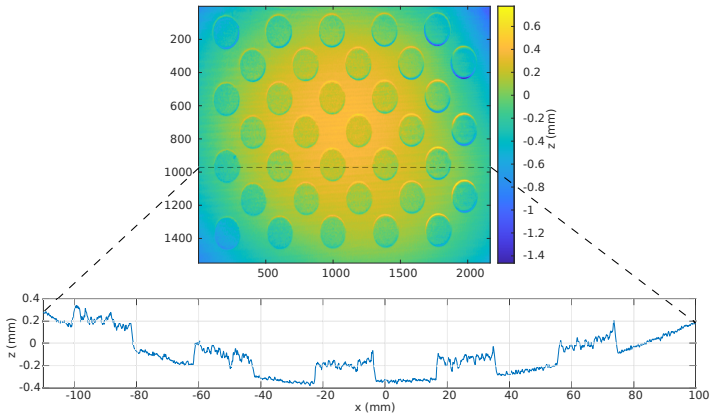


Fig. 1. 3D reconstruction and profile of a thermal calibration target fabricated on a copper circuit board. The measured 0.72 mm depth variation highlights how specialized targets may fail the flatness assumption required for accurate calibration.

still introduce errors if fabrication defects cause the surface to deviate from perfect flatness or warp feature points [14]. Flatness is difficult to achieve, as illustrated in Fig.1, which shows the 3D reconstruction and profile of a thermal-calibration target fabricated on a copper circuit board, revealing a measured 0.72 mm depth variation [15]. Although specialized designs aim to reduce reprojection errors in spectral cameras [14], producing high-precision targets increases manufacturing complexity and cost, limiting their accessibility and effectiveness.

Calibration procedures for multimodal structured-light systems follow two main workflows distinguished by the number of acquisition stages. In a single-stage procedure the same planar target is viewed simultaneously by the visible camera, the projector, and the secondary modality while fringes are projected (Fig. 2a). In a two-stage procedure the structured-light pair (visible camera + projector) is first calibrated with a conventional target and, in a second pass without fringes, the visible and secondary cameras are calibrated as a stereo pair using a second target (Fig. 2b) [16–18]. Both workflows obtain the intrinsic and extrinsic parameters with the standard planar-target calibration method [18], and the world origin is conventionally placed at the visible camera. Because each step depends on a specialized target with physical features, these standard procedures still inherit the cost, flatness, and detection issues outlined above, leaving a persistent gap.

To reduce those issues, several authors have designed custom targets tailored to specific modalities. Examples include heated aluminium plates perforated with circles for thermal–shape systems [19], circuit-board patterns with filled and empty apertures for high-speed thermography [13], and lamp-illuminated asymmetric circle grids to boost thermal contrast [5]. Two-target schemes have also been explored, pairing an LCD-displayed pattern for structured light with a perforated metal plate for the thermal camera [20, 21]. While these designs improve feature visibility, they demand precise machining, heating, or multi-material fabrication and often require many images to counter residual flatness errors. Thus, even the most creative target solutions with physical features remain vulnerable to manufacturing flaws and modality-specific compromises, highlighting the need for a different strategy.

Recent research is moving toward calibration techniques that sidestep the limitations of manufactured boards. Approaches for structured light recover projector parameters with a single image per pose [22] or fuse stereo vision with phase–coordinate mapping for depth-range accuracy [23, 24]. Others place digital feature points on planes extracted from reconstructed surfaces, eliminating physical patterns altogether [25, 26]. Beyond structured light, edge

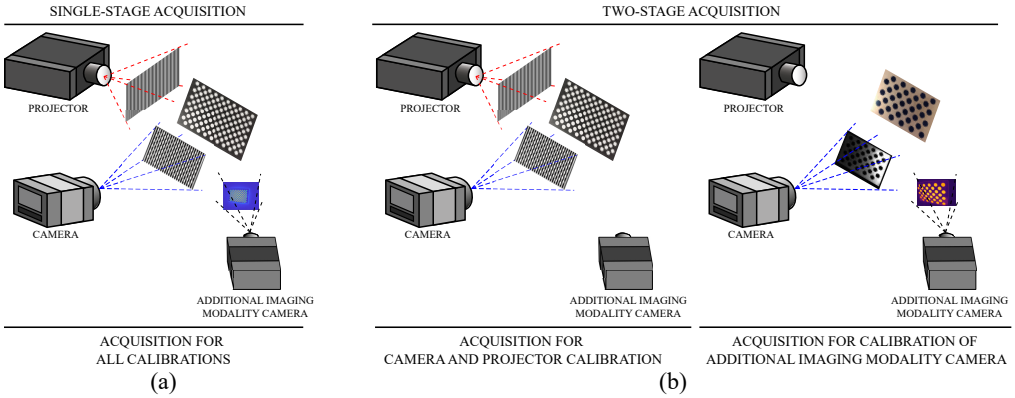


Fig. 2. Conventional calibration procedures. (a) Single-stage acquisition using a visible target in all imaging modalities; (b) Two-stage acquisition using two different targets.

correspondences have calibrated event-camera/LiDAR pairs [27], and two-view geometry has enabled in-situ calibration of refractive systems without external references [28]. By removing fabricated targets with physical features, these methods tackle the cost and flatness issues that plague traditional workflows, yet none has fully addressed multimodal structured-light setups with non-visible cameras. Building on this trend, our work introduces a digital-feature calibration specifically tailored to such multimodal systems. Thus, we can avoid printing or manufacturing sophisticated targets.

We propose a calibration strategy for multimodal structured-light systems that replaces specialized targets with digital features generated by readily available hardware. A tablet or monitor first displays a feature pattern to calibrate the visible and auxiliary cameras; the planar back of a mirror is then imaged to recover projector intrinsics and extrinsics by placing virtual tie points on the mirror's 3D reconstruction; finally, the same mirror corners and virtual points yield the intrinsic and extrinsic parameters of the secondary sensor—illustrated here with a thermal camera but applicable to hyperspectral, SWIR, or other modalities. Experiments with this setup achieve a root-mean-square (RMSE) 3D error of 0.07 mm, matching standard calibration methods while eliminating fabrication costs, flatness defects, and feature detection problems. The result is a cost-effective, accurate, and robust procedure for calibrating a wide range of multimodal structured light imaging systems.

2. Principles

2.1. Stereo-vision model

The pinhole camera model relates a 3-D point $[x^w, y^w, z^w]^T$ in the world coordinate system to its image projection $[u^c, v^c]^T$ in camera c as

$$s^c [u^c, v^c, 1]^T = \mathbf{A}^c \, [\mathbf{R}^c \; \mathbf{t}^c] \, [x^w, y^w, z^w, 1]^T, \tag{1}$$

where s^c is a scale factor, $\mathbf{A}^c$ denotes the camera intrinsic matrix which contains the focal lengths and principal-point coordinates, and $\mathbf{R}^c, \mathbf{t}^c$ express the rigid pose of the world frame with respect to the camera. Any additional device a (auxiliary camera, projector, *etc.*) follows

$$s^a [u^a, v^a, 1]^T = \mathbf{A}^a \, [\mathbf{R}^a \; \mathbf{t}^a] \, [x^w, y^w, z^w, 1]^T, \tag{2}$$

with symbols defined analogously. Intrinsic and extrinsic parameters are usually estimated from multiple views of a planar target using the standard camera calibration procedure [18].

Real optics deviate from the ideal pinhole, primarily through radial and tangential distortion. Normalized undistorted coordinates $[\bar{u}, \bar{v}]^T$ map to distorted coordinates $[\bar{u}_d, \bar{v}_d]^T$ via

$$\begin{bmatrix} \bar{u}_d \\ \bar{v}_d \end{bmatrix} = \left(1 + k_1 r^2 + k_2 r^4 + k_3 r^6\right) \begin{bmatrix} \bar{u} \\ \bar{v} \end{bmatrix} + \begin{bmatrix} 2p_1 \bar{u}\bar{v} + p_2(r^2 + 2\bar{u}^2) \\ 2p_2 \bar{u}\bar{v} + p_1(r^2 + 2\bar{v}^2) \end{bmatrix}, \tag{3}$$

with $r^2 = \bar{u}^2 + \bar{v}^2$ and $[\bar{u}, \bar{v}, 1]^T = \mathbf{A}^{-1}[u, v, 1]^T$. Coefficients $k_{1\text{–}3}$ model radial distortion and p_1, p_2 tangential distortion.

2.2. Phase-shifting algorithm

Phase-shifting fringe projection yields dense, subpixel phase maps for high-resolution 3-D reconstruction [29]. For an N-step sequence with equal phase increments, the k^{th} captured intensity is

$$I = I' + I'' \cos\left(\phi + 2k\pi/N\right), \tag{4}$$

where I' is the average intensity, I'' the modulation, and ϕ the wrapped phase. The phase is obtained from the N intensities as

$$\phi = -\tan^{-1}\left[\frac{I_k \sin(2k\pi/N)}{I_k \cos(2k\pi/N)}\right]. \tag{5}$$

Phase unwrapping removes the 2π discontinuities by adding an integer multiple κ of 2π at each pixel,

$$\Phi = \phi + 2\pi\kappa, \tag{6}$$

yielding the continuous phase Φ used for depth retrieval [30].

3. Method

The proposed method calibrates a multimodal structured-light system without physical features by using a digital screen, the back side of a mirror, and an auxiliary camera. This method involves two stages of image acquisition. In the first acquisition stage two visible cameras capture a calibration pattern shown on a digital screen, providing data for a standard stereo calibration that yields the camera pair's intrinsics and pose (Fig. 3). In the second stage, an uncalibrated projector projects phase-shifted fringes onto the back of a planar mirror while the visible cameras capture the fringe sequence and the additional-modality camera capture one image per pose (Fig. 3). Once the acquisition stages are completed and the visible cameras have been calibrated, the data from the second stage is processed: phase correspondence allows a dense 3-D reconstruction of the mirror; a best-fit plane is extracted, virtual feature points are defined on that plane, and their distorted projections are computed on the main camera sensor (Fig. 4a–f). These points are then re-projected into the projector frame to solve for the projector's intrinsic and extrinsic parameters, all expressed in the stereo coordinate system. The mirror corners seen by the visible and auxiliary cameras, together with the computed 3-D plane, supply the data needed to estimate the additional imaging modality camera's intrinsics. The extrinsic calibration of this camera is performed using the virtual feature points. These points are estimated on the other imaging modality camera sensor through a homography between its sensor and the main camera sensor (Fig. 4g). The following subsections describe each processing step in detail.

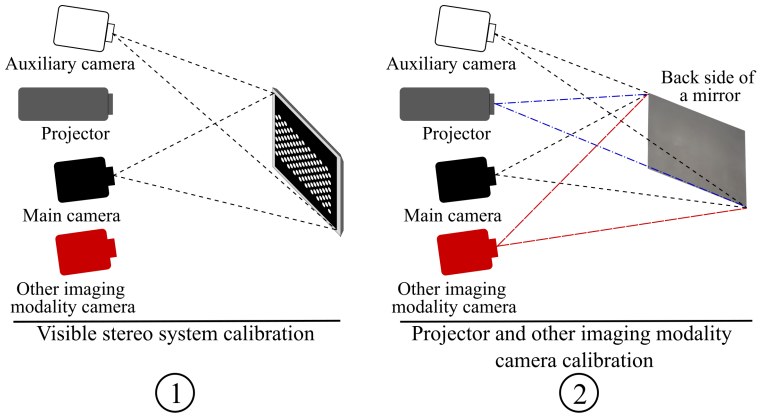


Fig. 3. Acquisition stages for the proposed method. First, visible cameras acquire images of a pattern displayed in a screen for stereo calibration. Second, the projector projects fringes onto the back side of a mirror while the visible cameras capture images of each projected pattern. The additional-modality camera acquires a single image per pose.

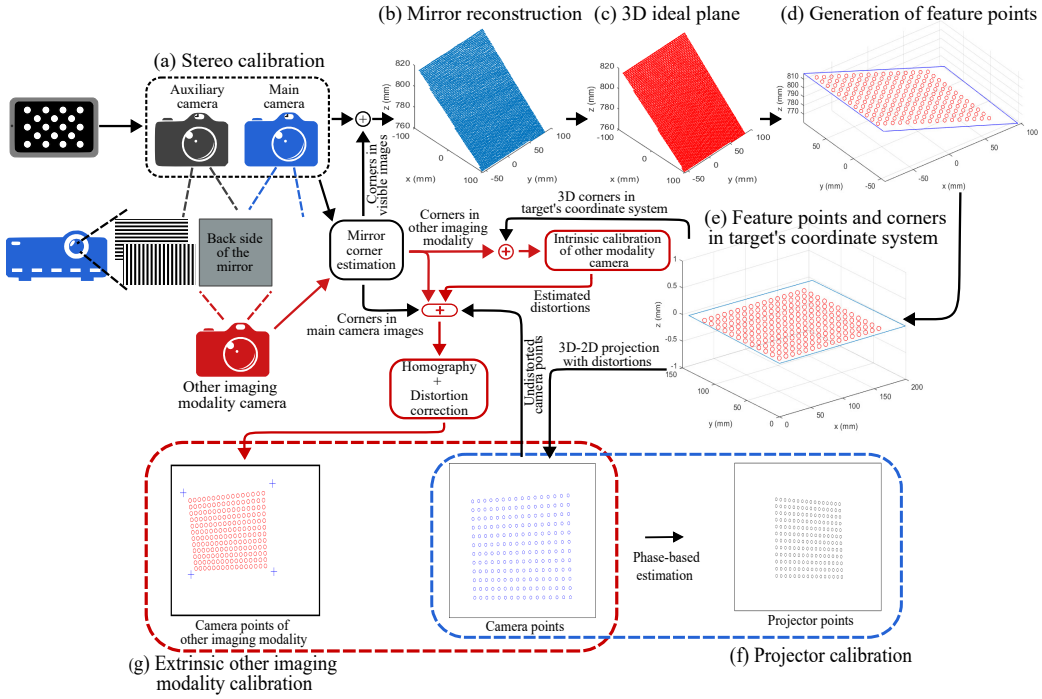


Fig. 4. Proposed digital-feature calibration method. (a) Stereo calibration is performed using a digital pattern. (b) Via fringe projection the mirror surface is reconstructed, and (c) a plane is fitted. (d) Digital feature points are defined and (e) projected onto the main camera and projector for calibration. (f) Projector points are estimated via phase data. (g) The additional-modality camera is calibrated using mirror corners and digital features. The method enables full system calibration without requiring physical features.

3.1. Visible stereo system calibration

The first stage calibrates the stereo pair formed by the main and auxiliary visible cameras. Both sensors capture multiple views of a calibration pattern displayed on an LCD screen (Fig. 4a). A standard stereo calibration procedure yields each camera matrix, both sets of distortion coefficients, and the rigid transform between the cameras [18]. The auxiliary camera is removed once the entire multimodal system has been calibrated.

3.2. Projector calibration

Projector calibration models the back side of the mirror as a virtual planar target. Virtual feature points placed on that plane are projected onto the main-camera and projector image planes, and their correspondences supply the input to a conventional stereo calibration procedure [31].

3.2.1. Target plane estimation

Vertical and horizontal fringe patterns are projected onto the back side of the mirror, and both visible cameras record the corresponding phase sequences. Phase values at selected pixels on the mirror region of the main-camera image are computed with subpixel accuracy, then interpolated in the auxiliary-camera image to establish point correspondences. Distortion is removed with Eq. (3), and the pinhole models of Eq. (1) and Eq. (2) triangulate each pair to create a dense 3-D point cloud of the mirror surface (Fig. 4b).

A least-squares fit determines the plane that best represents these points,

$$a\,x + b\,y + z + c = 0, \tag{7}$$

where $[a, b, c]^T$ is the plane normal expressed in world coordinates. Adopting this plane as the model coordinate system flattens local surface defects (Fig. 4c).

3.2.2. Generation of feature points

Points on the target plane are transformed from the world coordinate system (main camera) to the model coordinate system (target) with z values equal to zero ($z = 0$), following the method described by Zhang [25]. Now, we have a surface where the feature points will be placed. For this, we need defining a mirror vertex as model coordinate system center. Also, feature points should be located inside the surface perimeter so that they extend over the majority of its area, thus allowing better sampling on the camera and projector sensors. Therefore, we need to find the surface vertex to estimate the surface dimensions. The mirror vertex are detected using image processing.

To find the corners of the point cloud (mirror corners), we binarize the texture image and find the contourn of the mirror. We developed an algorithm to estimate the vertices, starting from the initial corner points identified using the method proposed by Matt [32]. Then, the Harris–Stephens algorithm is applied to the thresholded mirror image to obtain a set of vertex candidates. Additionally, contour points are extracted as another set of vertex candidates. Next, the candidates are refined based on their Euclidean distance to the initial corner estimates. Finally, the vertex farthest from the centroid of the mirror is selected as the final corner.

Next, the feature points are defined within the model coordinate system on the surface (Fig. 4e). These points are then transformed into the world coordinate system. Figure 4d represents this process. To obtain their projection onto the main camera sensor, the feature points are multiplied by the intrinsic matrix, following Eq.1, and subsequently corrected for distortion.

We estimate the phase values along the horizontal and vertical directions at these camera points using cubic interpolation based on the phase values of the nearest pixels. Finally, the corresponding projector points are obtained using the following equation

$$[u^p, v^p] = \frac{P}{2\pi}[\phi_u, \phi_v] \quad , \tag{8}$$

where $[\phi_u, \phi_v]$ are the continuous phases along the u- and v-directions, P indicates the pitch of the projected fringes, and $[u^p, v^p]$ are the projector points. Figure 4f illustrates the projection of the feature points onto the camera and projector images.

Repeating the acquisition over several poses yields enough point pairs to run a standard stereo calibration and recover the projector's intrinsic and extrinsic parameters together with its distortion coefficients.

3.3. Other imaging-modality calibration

The proposed procedure is applicable to other imaging modalities because it requires only the detection of mirror corners; a thermal camera serves here as an example. For each pose the thermal camera records one image of the mirror, and the calibration proceeds in two steps.

3.3.1. Intrinsic calibration

Mirror vertices are first located in the thermal image, aided by contrast enhancement and sharpening. These vertices, together with the corresponding vertices already extracted from the texture image, form the corner pairs needed for a standard calibration routine, which returns the thermal camera matrix and its distortion parameters (Fig. 4).

3.3.2. Extrinsic calibration

Accurate alignment relies on the virtual feature points introduced in Section 3.2.2. A homography computed from the undistorted mirror corners maps the feature points from the main-camera sensor to the thermal-camera sensor. The estimated homography is then applied to the undistorted camera points to determine the corresponding feature points in the thermal camera image. The thermal distortion model is then applied to obtain their final pixel coordinates. Fig. 4g illustrates the mapped points on the thermal image. A final stereo calibration between the main and thermal cameras, using these correspondences and the known 3-D feature coordinates, yields the thermal camera pose with respect to the world frame. The required homography for each image pair is defined as follows:

$$\begin{bmatrix} u^t \\ v^t \\ 1 \end{bmatrix} = H \begin{bmatrix} u^c \\ v^c \\ 1 \end{bmatrix} = \begin{bmatrix} h_1 & h_2 & h_3 \\ h_4 & h_5 & h_6 \\ h_7 & h_8 & h_9 \end{bmatrix} \begin{bmatrix} u^c \\ v^c \\ 1 \end{bmatrix} , \tag{9}$$

where $[u^c, v^c, 1]^T$ represents the homogeneous coordinates of the undistorted corners in the visible camera image; $[u^t, v^t, 1]^T$ denotes the homogeneous coordinates of the undistorted corresponding points in the thermal camera image; H is the homography matrix, and $h_1, h_2, \ldots, h_9$ are its elements. The homography matrix is estimated using singular value decomposition [33].

4. Experimental results

For the experiments, we implemented our multimodal calibration approach using a structured light system and a thermal camera. Figure 5 shows the developed multimodal system, which consists of a Dell M115HD DLP projector (1280 × 800 pixels), two Basler acA1300-200um USB 3.0 grayscale cameras (1280 × 1024 pixels) with 12 mm-focal length lenses, an Infiray P2 thermal camera (256 × 192 pixels) with a thermal resolution of 0.04 °C, and an Arduino Uno for synchronization. The system was fine-tuned to achieve a reconstruction cross-sectional area of 200 mm x 250 mm at a distance of around 550 mm. The OpenCV stereo vision system calibration package was used to estimate the intrinsic and extrinsic parameters of the cameras and the projector. The intrinsic parameters of each optical device, once estimated, were fixed for all subsequent extrinsic calibrations.

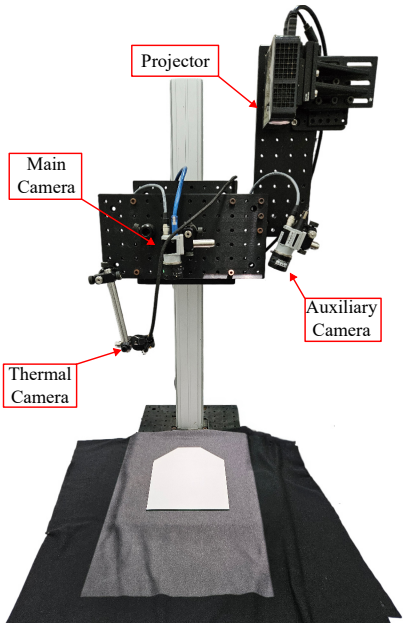


Fig. 5. Proposed setup equipped with two visible cameras, projector and thermal camera for multimodal calibration.

A HUAWEI MediaPad T3 10 tablet was used to display a 6×19 white circle dots to calibrate the stereo system composed by the main and auxiliary cameras. The distance between circle center is 20 mm. We calibrated the stereo cameras with 91 poses and obtained 0.16 as overall reprojection mean error.

Target plane estimation involved using 18-step phase-shifted fringe patterns and 6 gray-coded binary patterns for extracting phase maps in horizontal and vertical orientations. We captured 31 poses of the back side of an irregular hexagon-shaped mirror with a longest base of 20.1 cm and a shortest base of 14.9 cm. This mirror is shown in Fig. 5. We estimated approximately 1200 matching pairs on the phase maps by cubic interpolation. These points were used to reconstruct the mirror surface using the camera parameters estimated previously. Figure 6a shows the mirror surface reconstructed for a pose. Since the mirror surface reconstructed is not perfectly planar, we estimated the ideal 3D plane (See Fig. 6b). This ideal plane was the plane of our target where we placed our feature points. For each pose, we estimated the transformation between the main camera coordinate system and target coordinate system. Thus, the plane points are transformed to the target coordinate system, as Fig. 6c shows.

To generate feature points, we defined a set with 270 points in the target coordinate system adjusted to its surface. The step between the feature points was defined as 10 mm. Figure 7a shows the feature points generated in the target coordinate system for the same pose shown in Fig. 6. To calibrate the structured light system, it is necessary to obtain the projection of the feature points on the camera and projector sensors. Thus, we transformed the points from the target coordinate system to the camera coordinate system, as shown in Fig. 7b. Subsequently, the feature points were projected onto the camera sensor using Equation 1 taking into account the distortions (see Fig. 7c). The points on the projector sensor were estimated using the phase in the camera points and Eq. 8. In the above example, the points on the projector sensor are shown in Fig. 7d. Structured light calibration was performed using the standard stereo calibration procedure with 67 poses, resulting in a mean reprojection error of 0.06 pixels.

The intrinsic calibration of the thermal camera was successfully performed using the afore-

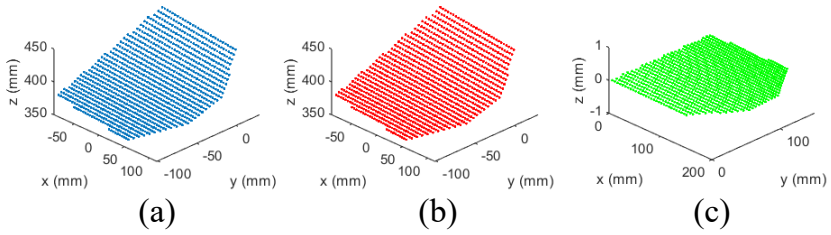


Fig. 6. Target plane estimation. (a) Mirror surface reconstructed using stereo calibration data; (b) Ideal 3D plane of the mirror with function $-0.0853x+0.5504y+422.8925 = z$; (c) Plane points in the target coordinate system.

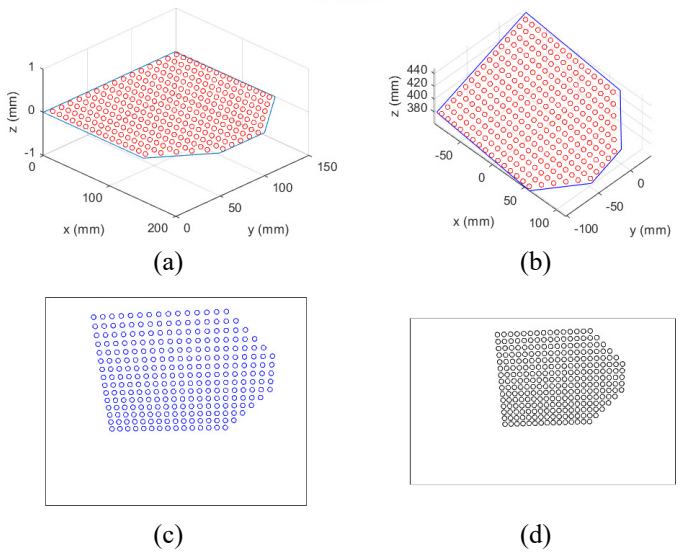
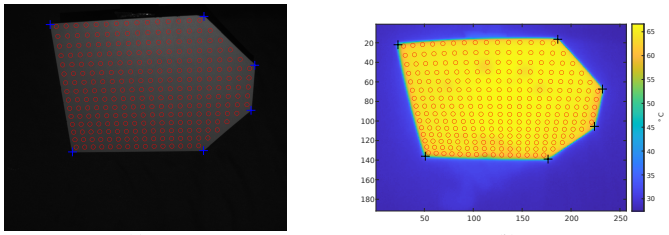


Fig. 7. Feature points. (a) Feature points in target coordinate system; (b) Feature points in main camera coordinate system; (c) Feature points on the camera image; (d) Feature points on the projector image.

mentioned calibration package. This process utilized the mirror corners detected in the thermal images and the ideal 3D reconstructions of the corners in the target coordinate system as input data.

Later, the corners in the visible and thermal images were used to estimate the homography between these images. We present an example of thermal feature point estimation in Fig. 8. In this figure, the mirror corners are marked by the blue crosses. The feature points in the visible image are presented in Fig. 8a. These points were used with the homography to estimate the feature points in the thermal image. Figure 8b shows the feature points estimated in the thermal image. Finally, an extrinsic calibration was performed using the standard stereo calibration procedure and 59 poses, with a mean reprojection error equal to 0.46 pixels. A few poses used in the structured light calibration were not utilized in the thermal camera extrinsic calibration due to corner detection errors in this image modality.



(a) (b)

Fig. 8. Feature points (red circles) and mirror corners (crosses) in thermal camera calibration. (a) Feature points defined in visible camera image; (b) Feature points estimated in thermal camera image using a homography with the mirror corners.

4.1. Comparison with a conventional multimodal structured light and thermography calibration

Conventional calibration based on two-stage acquisition with two targets with physical features was used to calibrate the same system without an auxiliary camera. The structured light calibration employed an asymmetrical circular target composed of 8×21 white circles, spaced 16 mm apart from center to center, set against a black background. The target was placed in 84 unique positions, and at each position, we projected and recorded 18-step phase-shifted fringe patterns and 7 gray-coded binary patterns in both vertical and horizontal directions. We synchronized the image capture with a trigger signal from the projector using an HDMI splitter. The calibration resulted in a mean reprojection error of 0.12 pixels. The thermal camera calibration was performed using a copper-printed circuit board with vinyl cutouts forming a 7×9 asymmetrical pattern of black circles as a target (target in Fig. 10a). To enhance feature point detection with a thermal camera, the target was heated to 70°C, using a 3D printer's heat bed. The target's non-vinyl sections acted as thermal mirrors, reflecting the surrounding temperature to avoid detection issues from thermal reflections. Images were captured with visible and thermal cameras for each calibration pose, and adaptive contrast enhancement was applied to improve feature point detection. The system was calibrated with 84 image pairs, achieving a mean reprojection error of 0.3 pixels.

We evaluated the accuracy of both the proposed and conventional calibration methods by reconstructing the 3D geometry of an object with known dimensions. For this purpose, we acquired multimodal images of a sphere, which has a nominal radius of 20 mm. The radius

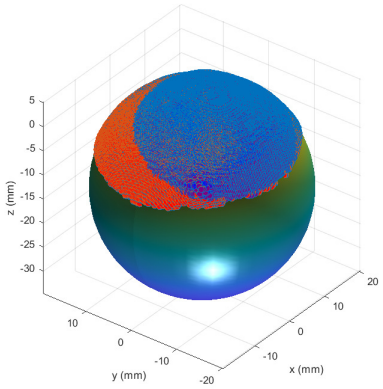
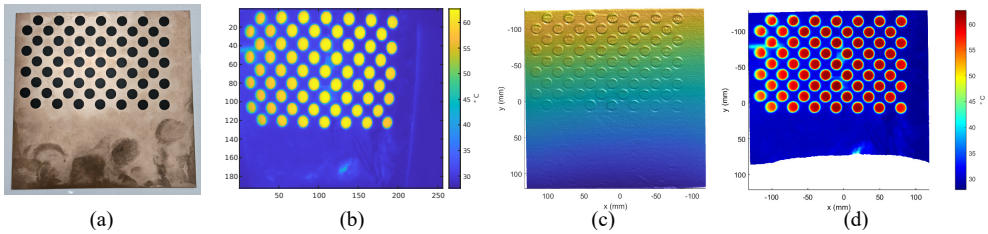
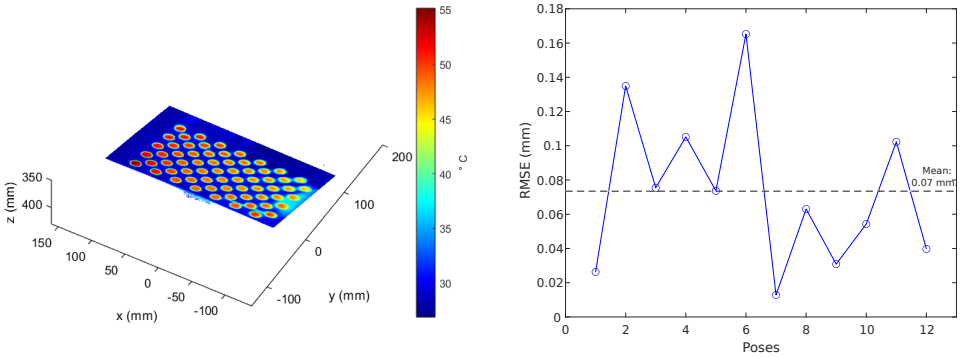


Fig. 9. 3D reconstructions of a 40 mm sphere with the conventional (red) and proposed (blue) methods.



(a) (b) (c) (d)

Fig. 10. Images of the target used for thermal camera calibration in a specific pose across different imaging modalities, employing the calibration parameters estimated by our method: (a) Image of the target, fabricated on a copper-printed circuit board with vinyl cutouts forming a 7×9 asymmetrical pattern of black circles, (b) Thermal image, (c) 3D reconstruction of the target, and (d) Multimodal image combining both modalities.



(a) (b)

Fig. 11. RMSE between the 3D surfaces obtained using the conventional and the proposed methods: (a) Multimodal image from the target in a pose using our approach, (b) RMSE for each pose between the methods. Its mean value is 0.07 mm

estimated from the 3D reconstruction using the conventional method was 19.7969 mm, while the proposed method yielded 19.8058 mm. Figure 9 presents the reconstructed sphere using both methods, along with a reference sphere of radius 19.8 mm. The root-mean-square error (RMSE) of the 3D points was 0.0731 mm for the conventional method and 0.0713 mm for the proposed method. These results confirm that both approaches produce similarly accurate reconstructions consistent with the physical dimensions of the object.

We further compared the performance of both methods in reconstructing a physical multimodal target under varying acquisition poses. Figure 10a shows the target used in this evaluation. Thirteen different poses were acquired. Using calibration parameters estimated by the proposed method, Fig. 10b-d present, respectively, the thermal image, the 3D reconstruction, and the multimodal composite image for a representative pose. During acquisition, the target was thermally stimulated, and temperatures ranged from 30°C to 60°C. The black circles appear hotter than the metallic background, which acts as a thermal mirror reflecting ambient temperatures. The sharp boundaries of the circles in the thermal image indicate accurate thermal-to-geometric mapping.

For each pose, the depth profiles reconstructed with both calibration methods were centered by subtracting their mean, then detrended using a second-order polynomial, and finally compared

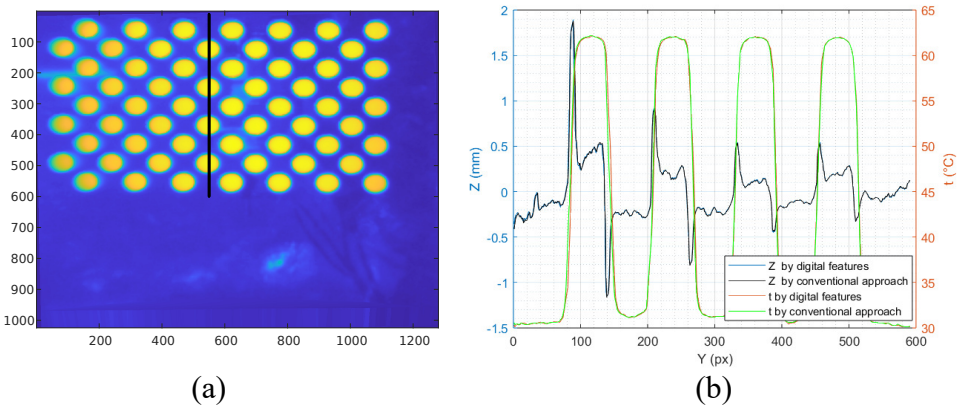


Fig. 12. Mapping validation on a multimodal target. (a) Thermal texture mapped with the digital-feature calibration; the vertical line marks the extracted profile. (b) Temperature and depth profiles along that line. The mean of each profile was subtracted and the depth data were detrended with a second-order polynomial.

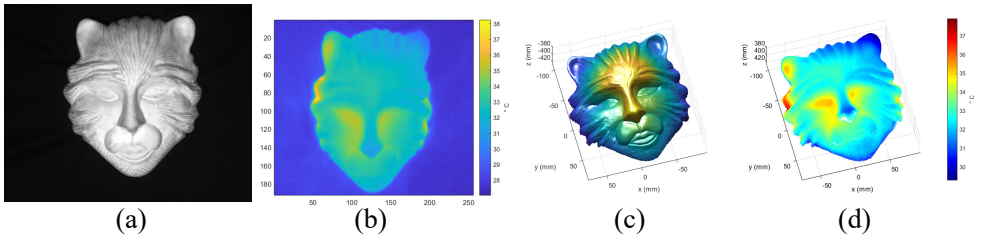


Fig. 13. Multimodal image of a 3D object. (a) Photography of the measured object; (b) Image captured by the thermal camera; (c) 3D reconstructed geometry; (d) Temperature mapping result.

using RMSE. Figure 11a shows one representative multimodal reconstruction, while Fig. 11b plots the RMSE values across all poses. The average RMSE was 0.07 mm, demonstrating that the proposed calibration method achieves the same level of accuracy as the conventional approach.

To further validate the calibration, Fig. 12a displays the thermal texture mapped onto the target using the proposed method. The circle pattern is clearly visible, and the vertical line indicates the location of the extracted profile. Figure 12b shows the corresponding depth and temperature profiles. After detrending, the depth data from both methods align closely, with minor discrepancies near high-contrast edges likely caused by intensity-induced artifacts. The temperature curves also show strong agreement, with corresponding periodic peaks. These results confirm that the proposed method delivers thermal and geometric accuracy equivalent to conventional calibration, even when using a lower-resolution thermal camera.

Finally, we applied the proposed method to a real-world object with complex geometry. Figure 13a shows a grayscale visible image of a ceramic mask, illustrating its surface detail. The thermal image in Fig. 13b reveals temperature variations across the object, with warmer areas shown in yellow and cooler ones in blue. The 3D reconstruction using the proposed calibration is shown in Fig. 13c, capturing fine surface contours. Figure 13d combines depth and temperature into a 3D thermal map. This fusion of modalities demonstrates the method's effectiveness in accurately mapping thermal data onto complex surfaces.

5. Conclusion

We presented a calibration method for multimodal structured light and thermography systems based on digital features, eliminating the need for physical features in targets. The proposed approach accurately calibrates both the structured light system and an additional imaging modality, such as a thermal camera. It effectively overcomes common challenges in multimodal calibration, including thermal reflections, light saturation, fabrication defects, and optical nonlinearity. Importantly, it matches the accuracy of conventional methods while maintaining system simplicity.

Funding. This work has been funded by Universidad Tecnológica de Bolívar (UTB).

Acknowledgments. Eberto Benjumea thanks MinCiencias and Sistema General de Regalías (Programa de Becas de Excelencia) for a PhD scholarship. Fernando Quintero and Raúl Vargas thank Universidad Tecnológica de Bolívar (UTB) for post-graduate scholarships.

Disclosures. The authors declare no conflicts of interest.

Data availability. Data underlying the results presented in this paper are not publicly available at this time but may be obtained from the authors upon reasonable request.

References

1. A. G. Marrugo, F. Gao, and S. Zhang, "State-of-the-art active optical techniques for three-dimensional surface metrology: a review [invited]," J. Opt. Soc. Am. A **37**, B60–B77 (2020).
2. R. Juarez-Salazar, G. A. Rodriguez-Reveles, S. Esquivel-Hernandez, and V. H. Diaz-Ramirez, "Three-dimensional spatial point computation in fringe projection profilometry," Opt. Lasers Eng. **164** (2023).
3. J. Meza, S. H. Contreras-Ortiz, L. A. R. Perez, and A. G. Marrugo, "Three-dimensional multimodal medical imaging system based on freehand ultrasound and structured light," Opt. Eng. **60**, 054106 (2021). Publisher: SPIE.
4. B. Li, "High-speed 3d optical sensing for manufacturing research and industrial sensing applications," Trans. on Energy Syst. Eng. Appl. **3**, 1–12 (2022).
5. Y. An and S. Zhang, "High-resolution, real-time simultaneous 3D surface geometry and temperature measurement," Opt. Express **24**, 14552–14563 (2016). Publisher: Optica Publishing Group.
6. R. Y. Jablonski, C. A. Osnes, B. S. Khambay, *et al.*, "An in-vitro study to assess the feasibility, validity and precision of capturing oncology facial defects with multimodal image fusion," The Surg. **16**, 265–270 (2018).
7. R. Ramm, P. de Dios Cruz, S. Heist, *et al.*, "Fusion of Multimodal Imaging and 3D Digitization Using Photogrammetry," Sensors **24**, 2290 (2024). Number: 7 Publisher: Multidisciplinary Digital Publishing Institute.
8. R. Juarez-Salazar, J. Zheng, and V. H. Diaz-Ramirez, "Distorted pinhole camera modeling and calibration," Appl. Opt. **59**, 11310–11318 (2020). Publisher: Optica Publishing Group.
9. D. M. McClatchy, E. J. Rizzo, J. Meganck, *et al.*, "Calibration and analysis of a multimodal micro-CT and structured light imaging system for the evaluation of excised breast tissue," Phys. Med. & Biol. **62**, 8983 (2017). Publisher: IOP Publishing.
10. S. Feng, C. Zuo, L. Zhang, *et al.*, "Calibration of fringe projection profilometry: A comparative review," Opt. lasers engineering **143**, 106622 (2021).
11. Z. Qiu, J. Martínez-Sánchez, P. Arias-Sánchez, and R. Rashdi, "External multi-modal imaging sensor calibration for sensor fusion: A review," Inf. Fusion **97**, 101806 (2023).
12. S. Sreeharan, H. Wang, K. Hirakawa, and B. Li, "Bayesian calibration of digital fringe projection systems considering both aleatoric and epistemic uncertainties," Opt. Lasers Eng. **193**, 109098 (2025).
13. M. Landmann, S. Heist, P. Dietrich, *et al.*, "High-speed 3D thermography," Opt. Lasers Eng. **121**, 448–455 (2019).
14. A. ElSheikh, B. A. Abu-Nabah, M. O. Hamdan, and G.-Y. Tian, "Infrared Camera Geometric Calibration: A Review and a Precise Thermal Radiation Checkerboard Target," Sensors **23**, 3479 (2023).
15. E. Benjumea, R. Vargas, R. Juarez-Salazar, and A. G. Marrugo, "Toward a target-free calibration of a multimodal structured light and thermal imaging system," in *Dimensional Optical Metrology and Inspection for Practical Applications XIII,* vol. 13038 (SPIE, 2024), pp. 54–59.
16. J.-Y. Bouguet, "Camera Calibration Toolbox for Matlab," (2022). Language: en.
17. J. J. A. Taimal, B. B. Cortes, and A. D. R. Girón, "Software Tool for the Extrinsic Calibration of Infrared and RGBD Cameras Applied to Thermographic Inspection," Ingeniería **28**, e18145–e18145 (2023). Number: 1.
18. Z. Zhang, "A flexible new technique for camera calibration," IEEE Trans. on Pattern Anal. Mach. Intell. **22**, 1330–1334 (2000). Conference Name: IEEE Transactions on Pattern Analysis and Machine Intelligence.
19. S. Barone, A. Paoli, and A. V. Razionale, "Assessment of chronic wounds by three-dimensional optical imaging based on integrating geometrical, chromatic, and thermal data," Proc. Inst. Mech. Eng. Part H, J. Eng. Med. **225**, 181–193 (2011).

20. M. Rosenberger, C. Zhang, Y. Zhang, and G. Notni, "3D high-resolution multimodal imaging system for real-time applications," in *Dimensional Optical Metrology and Inspection for Practical Applications IX,* vol. 11397 (SPIE, 2020), pp. 21–30.
21. R. Yang and Y. Chen, "Design of a 3-D Infrared Imaging System Using Structured Light," IEEE Trans. on Instrum. Meas. **60**, 608–617 (2011). Conference Name: IEEE Transactions on Instrumentation and Measurement.
22. R. Juarez-Salazar, S. Esquivel-Hernandez, and V. H. Diaz-Ramirez, "Are camera, projector, and camera–projector calibrations different?" Appl. Opt. **62**, 5999–6006 (2023).
23. R. Vargas, A. G. Marrugo, S. Zhang, and L. A. Romero, "Hybrid calibration procedure for fringe projection profilometry based on stereo vision and polynomial fitting," Appl. Opt. **59**, D163–D169 (2020).
24. R. Vargas, L. A. Romero, S. Zhang, and A. G. Marrugo, "Pixel-wise rational model for a structured light system," Opt. Lett. **48**, 2712–2715 (2023).
25. S. Zhang, "Flexible structured light system calibration method with all digital features," Opt. Express **31**, 17076–17086 (2023). Publisher: Optica Publishing Group.
26. Y. Yang, Y.-H. Liao, I. Bortins, *et al.*, "Unidirectional structured light system calibration with auxiliary camera and projector," Opt. Lasers Eng. **175**, 107984 (2024).
27. W. Xing, S. Lin, L. Yang, and J. Pan, "Target-Free Extrinsic Calibration of Event-LiDAR Dyad Using Edge Correspondences," IEEE Robotics Autom. Lett. **8**, 4020–4027 (2023).
28. B. Elnashef and S. Filin, "Target-free calibration of flat refractive imaging systems using two-view geometry," Opt. Lasers Eng. **150**, 106856 (2022).
29. C. Zuo, S. Feng, L. Huang, *et al.*, "Phase shifting algorithms for fringe projection profilometry: A review," Opt. Lasers Eng. **109**, 23–59 (2018).
30. S. Zhang, "Absolute phase retrieval methods for digital fringe projection profilometry: A review," Opt. Lasers Eng. **107**, 28–37 (2018).
31. S. Zhang and P. S. Huang, "Novel method for structured light system calibration," Opt. Eng. **45**, 083601–083601 (2006).
32. M. J, "Find vertices in image of convex polygon," MATLAB Central File Exchange (2025). [Online]. Available: https://www.mathworks.com/matlabcentral/fileexchange/74181-find-vertices-in-image-of-convex-polygon. [Accessed: Feb. 18, 2025].
33. Y. Luo, X. Wang, Y. Liao, *et al.*, "A Review of Homography Estimation: Advances and Challenges," Electronics **12**, 4977 (2023).