

High-resolution, real-time simultaneous 3D surface geometry and temperature measurement

Yatong An and Song Zhang*

School of Mechanical Engineering, Purdue University, West Lafayette, Indiana, 47907, USA

[*szhang15@purdue.edu](mailto:szhang15@purdue.edu)

Abstract: This paper presents a method to simultaneously measure three-dimensional (3D) surface geometry and temperature in real time. Specifically, we developed 1) a holistic approach to calibrate both a structured light system and a thermal camera under exactly the same world coordinate system even though these two sensors do not share the same wavelength; and 2) a computational framework to determine the sub-pixel corresponding temperature for each 3D point as well as discard those occluded points. Since the thermal 2D imaging and 3D visible imaging systems do not share the same spectrum of light, they can perform sensing simultaneously in real time: we developed a hardware system that can achieve real-time 3D geometry and temperature measurement at 26 Hz with 768×960 points per frame.

© 2016 Optical Society of America

OCIS codes: (120.0120) Instrumentation, measurement, and metrology, (120.2650) Fringe analysis.

References and links

1. S. Zhang, *High-speed 3D Imaging with Digital Fringe Projection Technique* (Taylor & Francis (CRC), 2016), 1st ed.
2. K. Skala, T. Lipić, I. Sović, L. Gjenero, and I. Grubišić, "4d thermal imaging system for medical applications," *Periodicum biologorum* **113**, 407–416 (2011).
3. D. Borrmann, A. Nüchter, M. Dakulović, I. Maurović, I. Petrović, D. Osmanković, and J. Velagić, "A mobile robot based system for fully automated thermal 3d mapping," *Adv. Eng. Inform.* **28**, 425–440 (2014).
4. J. Rangel, S. Soldan, and A. Kroll, "3D thermal imaging: Fusion of thermography and depth cameras," presented at the International Conference on Quantitative InfraRed Thermography, Bordeaux, France, 2014.
5. M. A. Akhloufi and B. Verney, "Multimodal registration and fusion for 3d thermal imaging," *Mathematical Problems in Engineering* **2015**, 450101 (2015).
6. U. R. Dhond and J. K. Aggarwal, "Structure from stereo—a review," *IEEE Trans. Systems, Man, and Cybernetics* **19**, 1489–1510 (1989).
7. C. P. Keferstein and M. Marxer, "Testing bench for laser triangulation sensors," *Sensor Review* **18**, 183–187 (1998).
8. A. Kolb, E. Barth, and R. Koch, "Time-of-flight cameras in computer graphics," *Computer Graphics Forum* **29**, 141–159 (2010).
9. C. Filiberto, C. Roberto, P. Dario, and R. Fulvio, "Sensors for 3d imaging: Metric evaluation and calibration of a ccd/cmos time-of-flight camera," *Sensors* **9**, 10080–10096 (2009).
10. J. Salvi, S. Fernandez, T. Pribanic, and X. Llado, "A state of the art in structured light patterns for surface profilometry," *Patt. Recogn.* **43**, 2666–2680 (2010).
11. M. Subbarao and G. Surya, "Depth from defocus: a spatial domain approach," *Int. J. Comput. Vision* **13**, 271–294 (1994).
12. B. Li, N. Karpinsky, and S. Zhang, "Novel calibration method for structured light system with an out-of-focus projector," *Appl. Opt.* **53**, 3415–3426 (2014).

13. S. Zhang and P. S. Huang, "Novel method for structured light system calibration," *Opt. Eng.* **45**, 083601 (2006).
14. S. Coorg and S. Teller, "Real-time occlusion culling for models with large occluders," in "Proceedings of the 1997 symposium on Interactive 3D graphics" (ACM, 1997), pp. 83–ff.
15. G. Vaněček, "Back-face culling applied to collision detection of polyhedra," *JOVA* **5**, 55–63 (1994).
16. J.-S. Hyun and S. Zhang, "Enhanced two-frequency phase-shifting method," *Appl. Opt.* **55**, 4395–4401 (2016).

1. Introduction

Real-time measurement of a 3D geometric shape is vital for numerous applications including manufacturing, medical practices, and more [1]; temperature sensing using a thermal imaging camera is also of great interest to benefit both scientific research and industrial practices [2–5]. We believe that the combination of these two sensing modalities can substantially increase applications.

Static and real-time 3D shape measurement have been extensively studied over the past decades. 3D shape measurement techniques use different principles to achieve different capabilities. In general, 3D shape measurement techniques include stereo vision [6], laser triangulation [7], time of flight [8, 9] (e.g., Microsoft Kinect 2), structured light [10] (e.g., Intel RealSense and Microsoft Kinect 1), as well as shape from focus/defocus [11]. Among these methods, stereo vision and shape from focus/defocus do not require active illumination, and thus are regarded as passive methods. The passive methods can work well if an object surface has rich texture information, yet their accuracy will be compromised if a surface is uniform or has low texture variations. In contrast, those methods requiring active illumination are less sensitive to surface properties since 3D reconstruction is mainly based on the emission sent out from the emitter. Among those active methods, structured light techniques that use digital video projection devices to project computer generated structured patterns are popular because of their flexibility and accuracy.

The active structured light method can work well for both visible and near infrared wavelengths, yet they cannot work for longer wavelengths at which the silicon-based sensing devices fail to operate (e.g., thermal spectrum). Therefore, to the best of our knowledge, there are no systems that can simultaneously measure 3D geometric shape and surface temperature in real time. One major challenge is that the regular camera and the thermal camera do not *see* the same wavelength, thus it is difficult to calibrate these two types of cameras under the same coordinate system. Furthermore, commercially available, relatively inexpensive thermal cameras have low resolutions and large distortions, making the mapping between a thermal camera and a regular camera challenging.

This paper proposes a method to address the aforementioned challenges. To conquer these challenges, the calibration of these two cameras has to be carried out under the same world coordinate system and preferably uses the same calibration target. We propose a method that allows these two types of cameras to *see* the same object features. The basic idea is that we use a heat lamp to shine thermal energy on a black/white circle pattern calibration target. Due to the emissivity differences of black and white areas, the thermal camera and the regular camera can *see* the same calibration features (e.g., circles). By this means, these two cameras can be calibrated under the same world coordinate system using the same calibration target. Since the regular camera and the projector share the same spectrum, the structured light system can be calibrated using the same circle patterns. By coinciding the world coordinate system with the regular camera lens coordinate system, the whole system including the thermal camera is calibrated under the same world coordinate system. Since thermal cameras usually have a much lower resolution and larger distortion than a regular camera, we developed a computational framework to achieve sub-pixel corresponding temperature mapping point for each 3D point, and discard those occluded 3D points that are not visible to the thermal camera. Two different

hardware systems have been developed to verify the performance of the proposed method: 1) a static system that has a resolution of 1280×1024 points per frame; and 2) a real-time system that can achieve simultaneous 3D geometric shape and surface temperature measurement at 26 Hz with a resolution of 768×960 pixels per frame.

2. Principle

2.1. Least-square phase-shifting algorithm

Phase-shifting algorithms are extensively used in optical metrology because of their speed and accuracy. There are many phase-shifting algorithms such as three-step, four-step, five-step, etc. Generally, the more steps used, higher accuracy phase is obtained due to the averaging effect. For an N -step phase-shifting algorithm with equal phase shifts, the intensities of the k th fringe image can be described as:

$$I_k(x, y) = I'(x, y) + I''(x, y) \cos(\phi + 2k\pi/N), \quad (1)$$

where I' is the average intensity, I'' is the intensity modulation, and ϕ is the phase to be solved for. Using a least-square method, we can get

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

Since an inverse tangent function is used, the phase values obtained from this equation only vary from $-\pi$ to $+\pi$. The phase with 2π modulus should be unwrapped to obtain a continuous phase map for 3D shape reconstruction. There are numerous spatial or temporal phase unwrapping algorithms. Essentially those algorithms are trying to determine a fringe order $n(x, y)$ for each pixel and unwrap the phase by adding $2n\pi$. The spatial phase unwrapping typically only provides phase values relative to a point on the phase map, while the temporal phase unwrapping can provide absolute phase information that can be pre-defined. Once the absolute phase is obtained, the phase can be converted to 3D coordinates with a calibrated system, or to carry on unique information for other analysis, e.g., establish mapping between a camera and a projector for system calibration.

2.2. Pinhole camera model

In this research, we use a well-established pinhole model for the regular camera, the thermal camera, as well as the projector. The pinhole model essentially establishes the relationship between a point (x^w, y^w, z^w) in the world coordinate system, (x^c, y^c, z^c) in the camera lens coordinate system, and its imaging point (u, v) on the camera sensor. The linear pinhole model can be written as,

$$s \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \mathbf{A} [\mathbf{R}, \mathbf{T}] \begin{bmatrix} x^w \\ y^w \\ z^w \\ 1 \end{bmatrix}, \quad (3)$$

where s is a scaling factor indicating the depth, \mathbf{R} and \mathbf{T} are the rotation matrix and the translation vector that represent the transformation from the world coordinate system to the camera lens coordinate system; and \mathbf{A} is the intrinsic matrix of the camera describing the projection from the lens coordinate system to the 2D imaging plane. These matrices are usually in the following forms

$$\mathbf{A} = \begin{bmatrix} f_u & \gamma & u_0 \\ 0 & f_v & v_0 \\ 0 & 0 & 1 \end{bmatrix}, \mathbf{R} = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix}, \mathbf{T} = \begin{bmatrix} t_1 \\ t_2 \\ t_3 \end{bmatrix}, \quad (4)$$

where f_u and f_v are the effective focal lengths of the camera lens; (u_0, v_0) is the location of principle point; and γ is the skew factor of u and v axes, which is usually 0 for modern cameras. r_{ij} and t_i represent the rotation and translation from the world coordinate system to the camera lens coordinate system.

The linear model works well for perfectly designed and fabricated lenses, yet most lenses have distortion that the linear model does not represent. Among different kinds of distortions, radial and tangential distortions are the most severe and common. Typically, five coefficients are used to describe radial and tangential distortions as

$$\mathbf{Dist} = [k_1, k_2, p_1, p_2, k_3], \quad (5)$$

where k_1, k_2 , and k_3 describe radial distortions, and p_1 and p_2 describe tangential distortions. Radial distortions can be modeled as,

$$\begin{aligned} u' &= u(1 + k_1 r^2 + k_2 r^4 + k_3 r^6), \\ v' &= v(1 + k_1 r^2 + k_2 r^4 + k_3 r^6), \end{aligned} \quad (6)$$

where $r = \sqrt{u^2 + v^2}$, and (u', v') is the location of pixel (u, v) after radial distortion. Similarly, tangential distortions can be modeled as,

$$\begin{aligned} u' &= u + [2p_1 uv + p_2(r^2 + 2u^2)], \\ v' &= v + [p_1(r^2 + 2v^2) + 2p_2 uv]. \end{aligned} \quad (7)$$

2.3. 3D structured light system calibration

System calibration is intended to estimate the intrinsic and extrinsic matrix of the camera and the projector as well as the geometric relationship between them. Structured light system calibration follows the well-established method described in Reference [12]. In brief, a flat circle pattern board shown in Fig. 1(a) is used. It is placed at different orientations within the calibration volume to capture those 2D images. For each pose, both horizontal and vertical fringe patterns are projected to capture absolute horizontal and vertical absolute phase maps. These phase maps are then used to establish a one-to-one mapping between the camera and the projector [13] and to determine the corresponding points for each feature point on the camera. In this case, the feature points are those circle centers on the calibration board.

We used OpenCV camera calibration toolbox to detect those circle centers for the camera images and then found those corresponding center points for the projector by building the one-to-one mapping through phases. Once those center points are detected, the intrinsic parameters for the camera (\mathbf{A}^r) and the projector (\mathbf{A}^p) are estimated. We then use the stereo-calibration toolbox provided by the OpenCV camera calibration library to estimate the extrinsic parameters: \mathbf{R}^r the rotation matrix for the camera, \mathbf{T}^r the translation vector for the camera, \mathbf{R}^p the rotation matrix for the projector, and \mathbf{T}^p the translation vector for the projector. In this research, we coincide the world coordinate system with the camera lens coordinate system, and thus \mathbf{R}^r is an identity matrix, and \mathbf{T}^r is a zero vector.

As discussed in Reference [13], once the system is calibrated, 3D coordinates (x^w, y^w, z^w) in the world coordinate system can be computed for each camera point by solving the linear

equations

$$s^r \begin{bmatrix} u^r \\ v^r \\ 1 \end{bmatrix} = \mathbf{A}^r [\mathbf{R}^r, \mathbf{T}^r] \begin{bmatrix} x^w \\ y^w \\ z^w \\ 1 \end{bmatrix}, \quad (8)$$

$$s^p \begin{bmatrix} u^p \\ v^p \\ 1 \end{bmatrix} = \mathbf{A}^p [\mathbf{R}^p, \mathbf{T}^p] \begin{bmatrix} x^w \\ y^w \\ z^w \\ 1 \end{bmatrix}, \quad (9)$$

combined with the absolute phase constraint. Here (u^r, v^r) is the camera image coordinates, and (u^p, v^p) is the projector image coordinate. We only use the linear calibration model for our structured light system because we found such a model can achieve sufficient good accuracy.

2.4. Thermal camera calibration

Since a thermal camera is only sensitive to temperature variations, it cannot *see* the color difference in visible images. For example, Figure 1(a) shows the regular image of the circle pattern we used for system calibration, and Figure 1(b) shows image from the thermal camera. To solve this problem, we used a heat lamp (Model: GE Lighting 48037), as shown in Fig. 1(c), to shine thermal energy to the calibration board. Due to different emissivity of black and white areas, the thermal camera can capture circle patterns that are used for structured light system calibration. Figure 1(b) shows the thermal image of the circle patterns captured by a thermal camera without the heat lamp. Figure 1(d) shows the thermal image after turning on the heat lamp. Once the thermal camera can capture circle pattern images, its calibration becomes the well-established regular camera calibration problem.

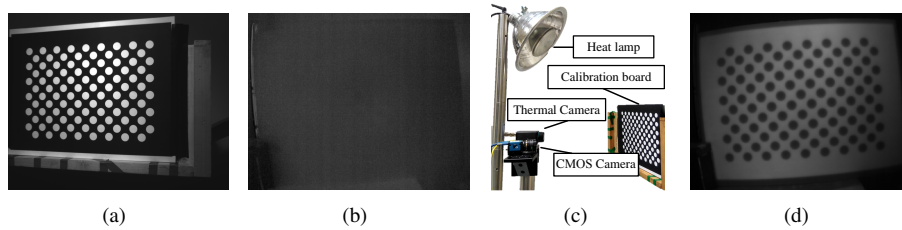


Fig. 1. Stereo calibration between the regular and thermal camera. (a) Calibration board used to calibrate the whole system (image was captured by the regular camera); (b) Image captured by a thermal camera before turning on the heat lamp; (c) System setup to calibrate the thermal and regular camera; (d) Image captured by a thermal camera after turning on the heat lamp.

However, as can be seen in Fig. 1(d), the thermal image has serious distortions. Therefore, a linear calibration model is no longer sufficient for thermal imaging camera calibration, and the nonlinear distortion coefficients **Dist** are considered in our research. Similarly, after capturing a sequence of circle pattern images under different poses, the intrinsic parameter matrix \mathbf{A}^t can be estimated.

Because the thermal camera calibration can use the same calibration target as the regular camera, the stereo calibration can also be carried for the regular camera and thermal camera pair to establish the geometric relationship between these two cameras. Again, we coincide the world coordinate system with the regular camera lens coordinate system and then estimate the rotation matrix \mathbf{R}^t , and the translation vector \mathbf{T}^t for the thermal camera.

2.5. Sub-pixel mapping between structured light system and thermal camera

Since the world coordinate system coincides with the regular camera for both the structured light system calibration and the thermal camera calibration, all these device calibrations are under exactly the same world coordinate system. That is, \mathbf{R}^p and \mathbf{R}^t respectively describe the rotation from the projector coordinate system to the regular camera lens coordinate system, and the rotation from the thermal camera coordinate system to the regular lens coordinate system; \mathbf{T}^p and \mathbf{T}^t respectively describe the translation from the projector coordinate system to the regular camera lens coordinate system, and the translation from the thermal camera coordinate system to the regular lens coordinate system. Therefore, it is straightforward to find the corresponding (u^t, v^t) point for a given 3D point $\mathbf{P}^w = (x^w, y^w, z^w)$ recovered from the structured light system. Mathematically, one can solve the following equation to find the mapped point

$$s^t \begin{bmatrix} u^t \\ v^t \\ 1 \end{bmatrix} = \mathbf{A}^t [\mathbf{R}^t, \mathbf{T}^t] \begin{bmatrix} x^w \\ y^w \\ z^w \\ 1 \end{bmatrix}, \quad (10)$$

assuming a linear calibration model is used, where s^t is a scaling factor.

However, as discussed above, although the structured light system uses a linear model, the thermal camera has to use the nonlinear model to represent its severe distortions. Therefore, the thermal image has to be rectified using the distortion coefficients **Dist** before mapping. In other words, (u^t, v^t) obtained directly from Eq. (10) corresponds to the rectified thermal image point, *not* the actual camera image point with distortions.

Even after rectification, the mapped point does not correspond to the thermal camera image pixel since the thermal camera's resolution is much lower than the regular camera's resolution. Hence, we propose to use a 2D Gaussian model to find the actual sub-pixel mapped thermal image point (or the actual temperature corresponding to that point).

Assume a 3D point is mapped to (u_0, v_0) based on Eq. (10). The Gaussian model provides the weighted average on all neighboring pixel values. Weights are related to the distance between the neighboring pixel and the mapped (u_0, v_0) point, which can be described as

$$f(i, j) = \exp \left(-\frac{(u_{ij} - u_0)^2 + (v_{ij} - v_0)^2}{2\sigma^2} \right), \quad (11)$$

where $f(i, j)$ is the weight function without normalization for a pixel at $[u_{ij}, v_{ij}]$, and σ is the standard deviation. Suppose the window size is $2L \times 2L$, the normalized weight is

$$w(i, j) = \frac{f(i, j)}{\sum_{\text{floor}(u_0)-L+1}^{\text{floor}(u_0)+L} \sum_{\text{floor}(v_0)-L+1}^{\text{floor}(v_0)+L} f(i, j)}, \quad (12)$$

where *floor* takes the nearest integer less than or equal to that element.

Finally, the temperature \mathcal{T} corresponding to the mapped point (u_0, v_0) is computed using

$$\mathcal{T}(u_0, v_0) = \sum_{\text{floor}(u_0)-L+1}^{\text{floor}(u_0)+L} \sum_{\text{floor}(v_0)-L+1}^{\text{floor}(v_0)+L} w(i, j) \mathcal{T}(i, j). \quad (13)$$

2.6. Invisible 3D point culling

Since two cameras have different perspectives, there are areas that can be seen by only one camera but not the other. In other words, some part of the object will be occluded in a specific viewpoint. Figure 2 illustrates one scenario, Curve \widehat{ABC} can be seen by the thermal camera, but not the part of Curve \widehat{DEF} . Therefore, those invisible points to the thermal camera should

be properly handled, otherwise, incorrect temperature mapping can be generated for 3D data points. To accurately detect those areas, we employed both the *occlusion culling* method [14] and the *back-face culling* method [15].

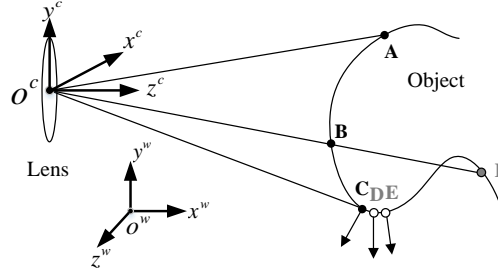


Fig. 2. Illustration of culling. Curve \widehat{ABC} can be seen in the view point of \mathbf{O}^c , but not the part of Curve \widehat{DEF} . Generally speaking, \widehat{EF} can be detected by the occlusion culling algorithm since they are obviously hidden by some other parts, while \widehat{DE} can be detected by the backface culling algorithm since they are on the edge of visible and invisible parts. For better culling results, we combine both occlusion culling and back-face culling methods.

Occlusion culling method finds occluded areas by finding the projected depth information of 3D points to the camera: if two points are corresponding to the same point, the point further away cannot be *seen* by the camera and thus should be regarded as an occluded point and discarded. For example, the B and F illustrated in Fig. 2 correspond to the same point on camera \mathbf{O}^c ; since F is further away from the camera, it should be regarded as occluded and thus discarded.

The occlusion culling method can be easily executed. We use Eq. (10) to map all points $\mathbf{P}^w = (x^w, y^w, z^w)$ on a 3D surface to the thermal image sensor. To quickly locate those occluded points, we create a vector \mathbf{S}_{ij} map for each pixel (i, j) on the thermal image to store projected depth z values and the corresponding 3D point.

$$\mathbf{S}_{ij} = \{z_{p_1}, z_{p_2}, \dots, z_{p_{n_{ij}}}\}, \quad (14)$$

where $z_{p_1}, z_{p_2}, \dots, z_{p_{n_{ij}}}$ are z -values of 3D points that are mapped to (i, j) on the thermal image. We then find the smallest element z_{ij}^{\min} in \mathbf{S}_{ij} ,

$$z_{ij}^{\min} = \min\{\mathbf{S}_{ij}\} \quad (15)$$

and discard all 3D points p_k in \mathbf{S}_{ij} whose z -values satisfy

$$z_{p_k} > z_{ij}^{\min} + th. \quad (16)$$

where th is the predefined threshold. In other words, we discard those points whose z -values are larger than the smallest one by a threshold th . The threshold value is determined from the prior knowledge of the hardware system and the type of object to be measured.

Practically, since the resolutions of two cameras are different, we can set a virtual camera with higher or lower resolution for more accurate culling. Suppose the resolution of a virtual camera is N times of the resolution of a real one. (u^t, v^t) determined from Eq. (10) need to be scaled up by factor of N , i.e.,

$$u_{new} = \text{floor}(u^t \times N), \quad (17)$$

$$v_{new} = \text{floor}(v^t \times N). \quad (18)$$

Instead of creating a vector for each (u, v) , we create a vector for each (u_{new}, v_{new}) , and the conditions to discard a 3D point is the same as Eq. (16).

If the occluded points are far away from the front point, such as those points between \widehat{EF} on Fig. 2, they can be easily detected by the occlusion culling method and discarded. However, this method does not work well for those points that are close to an edge of an object viewed from \mathbf{O}^c , such as those points between \widehat{DE} . This is because the occlusion culling method solely relies on the depth difference to determine which points to be discarded, and if the difference is very small (within the predefined threshold), those occluded points will fail to be detected. To handle such a condition, we propose to use the *back-face culling* method.

Back-face culling method detects occluded points by its surface normal direction. If the point normal's (\mathbf{n}_p) direction has a positive projection component to a camera's view direction, the point is regarded as a back-face point and thus should be discarded. The hollow circled dots D and E on Fig. 2 are regarded as back-face points and should be discarded. To implement the back-face culling method, the normal for each point of the point cloud data generated by the structured light system should be computed. Fortunately, since the point cloud data coming out of a structured light system are naturally aligned with the camera pixel grid, the point normal computation is straightforward: the averaged normals of the triangles formed by combining the surrounding pixels. We compute a point normal by considering the 3×3 neighborhood points of point \mathbf{P} , and normal \mathbf{n}_p is the average of $\mathbf{n}_1, \dots, \mathbf{n}_8$.

From the discussion in Sec. 2.5, the point cloud data coming out of the structured light system is in the world coordinate system that is perfectly aligned with the regular camera lens coordinate system, and thus all point normals should direct towards the regular camera. Since the thermal camera also has the same world coordinate system, the back-face point can be defined as

$$\mathbf{n}_p \cdot (\mathbf{P} - \mathbf{O}^c) > 0. \quad (19)$$

Here \mathbf{P} is the (x^w, y^w, z^w) coordinates of an arbitrary point P on the surface, and \mathbf{O}^c is the 3D coordinates of the second camera lens origin in the world coordinate system.

3. Experiment results

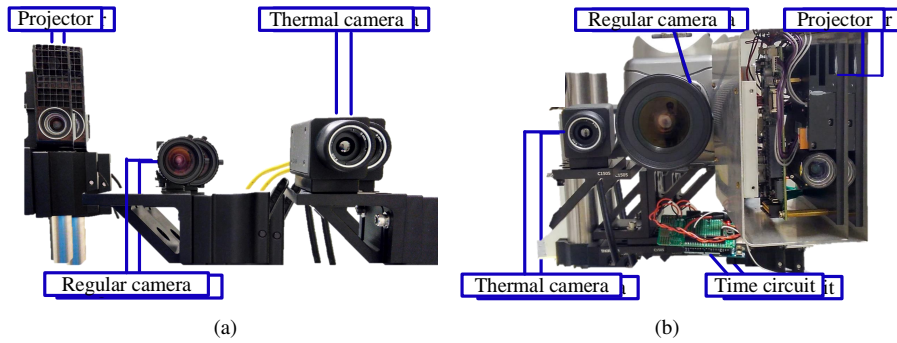


Fig. 3. Experimental system setups. (a) Static object measurement system consists of a DLP projector (DELL M115HD), a CMOS camera (Imaging Source 23UX174) and a thermal camera (FLIR A35); (b) Real-time measurement system consists of a thermal camera (FLIR A35), a high-speed DLP projector (LightCrafer 4500), a high-speed CMOS camera (Vision Research Phantom V9.1), and an external timing generator (Arduino UNO R3).

We developed a hardware system to verify the performance of the proposed method. Figure 3(a) shows the hardware system we developed. The overall system includes a digital-light-

processing (DLP) projector (Model: DELL M115HD), a complementary metal-oxide semiconductor, (CMOS) camera (Model: Imaging Source 23UX174) and a thermal camera (Model: FLIR A35). The resolution of the projector is 1280×800 , the resolution of the CMOS camera is 1280×1024 , and the resolution of the thermal camera is 320×256 . The CMOS camera is attached with 8 mm focal length lens (Model: Computar M0814-MP2). For all 3D shape measurement experiments carried out with this system, we used $N = 9$ phase-shifted fringe patterns with a fringe period of 18 pixels to obtain wrapped phase map. The wrapped phase map is then unwrapped by projecting 7 binary coded patterns to uniquely determine fringe orders for each pixel. The absolute phase is further converted to 3D geometry pixel by pixel.

We first measured a black/white checkerboard heated up by the thermal lamp that was used for system calibration to verify the mapping accuracy. Figure 4 shows the results. Figure 4(a) shows the checkerboard image from the regular CMOS camera; and Figure 4(b) shows the image captured by the thermal camera at the same time. Color shows temperature ranging from 290 to 323 Kelvin (K). We measured the 3D geometry of the checkerboard using the structured light system and then mapped the thermal temperature image onto the 3D geometry. Figure 4(c) shows the 3D geometry of the checkerboard rendered in shaded mode, and Figure 4(d) shows the mapping result. This figure shows that the temperature difference between black and white blocks. Temperature in black blocks are higher than that in white blocks since black is of higher emissivity. This figure also shows that the boundary between black and white blocks are also very clear. Therefore, the mapping was fairly accurate, at least visually.

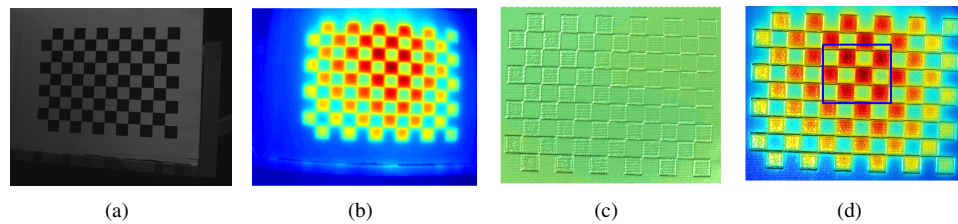


Fig. 4. Mapping example of a checkerboard. (a) Image of the checkerboard captured by the CMOS camera. Its resolution is 1280×1024 ; (b) Image captured by the thermal camera before rectification. Its resolution is 320×256 ; (c) 3D reconstructed geometry; (d) Mapping result. Color represents temperature ranging from 290 to 323 K in both (b) and (d).

To better visualize the mapping quality, we showed close-up views of checker squares in Fig. 5(a), and area around the corners in Fig. 5(b). Again the mapping quality is pretty good. We further analyzed the mapping quality by plotting a vertical slice of the result, as shown in Fig. 5(c). Due to the large contrast of the checkerboard, the 3D shape measurement system created border artifacts (transition from the black to white or from white to black is not smooth). For better comparison, we detrend the depth values using a linear model and shifted them by adding 314 mm. This figure shows that these borders are perfectly aligned with the middle of the temperature changes. In summary, the sub pixel mapping method developed in this research is very accurate even though the thermal camera has a much lower resolution than the regular camera.

Since the checkerboard we used for previous measurement is flat, the occlusion problem is not obvious. We then measured a complex shape 3D statue to further verify the performance of the mapping method and validate the performance of the culling method. Figure 6(a) shows the photograph of the statue captured by the regular camera of the structured light system. Again, the statue was heated up by the thermal lamp, and we captured a temperature image by the thermal camera, as shown in Fig. 6(b). Figure 6(c) shows the 3D reconstruction from the

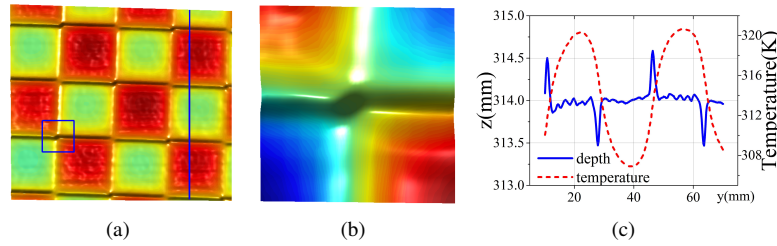


Fig. 5. Zoom-in analysis. (a) Zoomed-in result of the blue rectangle part of Fig. 4(d); (b) Further zoomed-in result of the corner part in (a); (c) Temperature and depth of the cross section in (a). For better comparison, we detrend the depth values using a linear model and shifted them by adding 314 mm. Color represents temperature ranging from 290 to 323 K in (a) and (b).

structured light system. Figure 6(a) and Fig. 6(b) show that these two images are of different poses, which is caused by different viewpoints of the CMOS and thermal cameras. Therefore, we have to properly remove those occlusion 3D points from the thermal camera in order to generate the correct temperature map. Figure 6(d) shows the temperature mapped onto the recovered 3D geometry after applying the culling methods discussed in this paper. Clearly a lot areas are culled out since they cannot be seen by the thermal camera. For those points that can be seen by the thermal camera, the temperature mapping is fairly accurate.

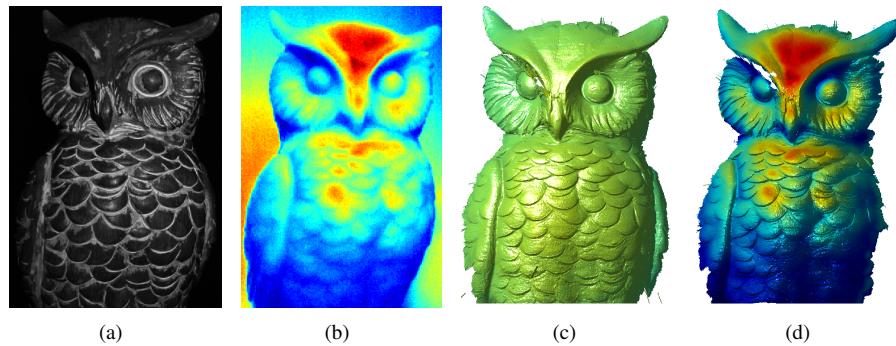


Fig. 6. Mapping example of a 3D object. (a) Photography of the measured object; (b) Image captured by the thermal camera before rectification; (c) 3D reconstructed geometry; (d) Temporal mapping result. Color represents temperature ranging from 292 to 297.5 K in both (b) and (d).

To better visualize the culling effects, Figure 7(a) and Fig. 7(b) respectively shows the zoomed-in view of the 3D geometry and that of the 3D geometry with temperature mapping of the head of the statue. Comparing these two images, we can clearly see that a lot points on the 3D geometry are culled out because the thermal camera cannot see them. To clearly mark those points that are culled out, Fig. 7(c) highlights those culled out points as black. These experiments demonstrated that our proposed mapping and culling methods both perform satisfactorily.

Since the thermal camera and the regular CMOS camera do not see the same spectrum of light, and 3D shape measurement and surface temperature measurement can be done at the same time, making real-time applications possible. To demonstrate this capability, we developed a system that uses the same thermal camera, a high-speed DLP projector (Model: Texas Instru-

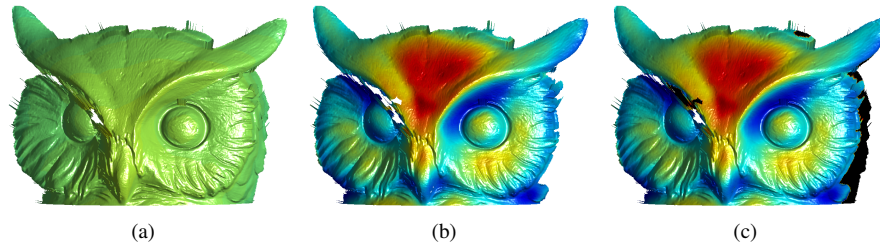


Fig. 7. Zoom-in view of the object showed earlier. (a) The top part of the original 3D geometry; (b) Temporal mapping result; (c) Highlighted points that are culled out as black.

ment LightCrafer 4500), and the high-speed CMOS camera (Model: Vision Research Phantom V9.1). Three devices are synchronized by using an external timing generator (Model: Arduino UNO R3 board with DIP ATmega328P). The whole system is shown in Fig. 3(b). The resolution of the camera is 768×960 and it is fitted with a 24 mm focal length lens (Model: Sigma 24mm /1.8 EX DG Aspherical Macro). For this experiment, the projector projects 912×1140 resolution binary dithered patterns at 780 Hz, and the thermal camera captures at 26 Hz. We used an enhanced two-frequency temporal phase unwrapping algorithm [16] to obtain absolute phases that are further converted to 3D geometry. Since it requires 6 fringe patterns to recover one 3D geometry, the 3D data acquisition speed is actually 130 Hz, which is 5 times of the thermal camera acquisition speed. Then we just pick one in every five frames to do mapping.

To demonstrate the real-time capacity, we measured both hands and human facial expressions using such a system. Figure 8 and Visualization 1, Visualization 2, Visualization 3, Visualization 4 and Visualization 5 are the results of the hand. Figure 8(a) shows one of the fringe images of the hand from the CMOS camera (associated with Visualization 1); and Figure 8(b) shows the image captured by the thermal camera at the same time (associated with Visualization 2). For real time experiments, we employed the enhanced two-wavelength phase shifting method [16] for 3D reconstruction, and mapped the temperature onto the 3D geometry simultaneously. Figure 8(c) shows one frame of the 3D reconstructed geometry (associated with Visualization 3); and Figure 8(d) shows the same frame with temperature mapping (associated with Visualization 4). In both Fig. 8(b) and Fig. 8(d), color represents temperature ranging from 296 to 303 K.

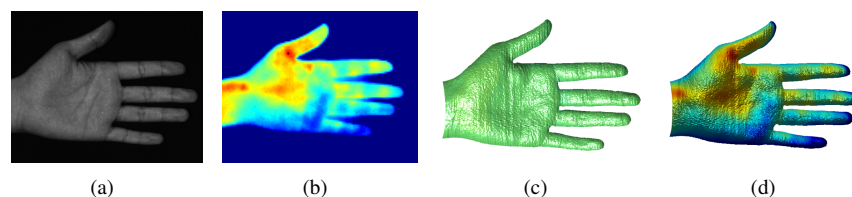


Fig. 8. Example of real time mapping of hand (Visualization 1, Visualization 2, Visualization 3, Visualization 4 and Visualization 5). (a) Photography of the hand to be measured captured by the CMOS camera (associated with Visualization 1); (b) Image captured by the thermal camera at the same time (associated with Visualization 2); (c) One frame of the 3D reconstructed geometry (associated with Visualization 3); (d) The same frame of the temperature mapping result (associated with Visualization 4). Color represents temperature ranging from 296 to 303 K in both (b) and (d).

We also measured human facial expressions. Figure 9 and Visualization 6, Visualization 7,

Visualization 8, Visualization 9 and Visualization 10 show the results. Fig. 9(a) shows the human face from the CMOS camera (associated with Visualization 6); and Figure 9(b) is the image captured by the thermal camera at the same time (associated with Visualization 7). Figure 9(c) shows one frame of the real time 3D reconstruction result (associated with Visualization 8); and Figure 9(d) shows the same frame with temperature mapping (associated with Visualization 9). In both Fig. 9(b) and Fig. 9(d), color represents temperature ranging from 297 to 305 K. This experiment verifies our algorithm's capacity for real time 3D geometric shape measurement and temperature mapping.

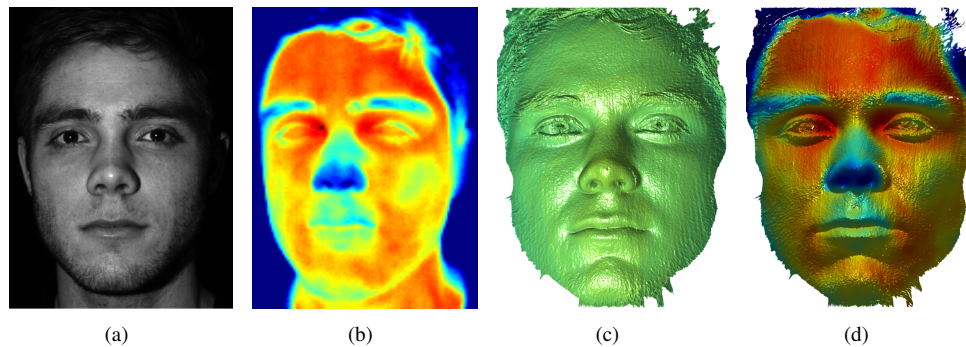


Fig. 9. Example of real time mapping of human face (Visualization 6, Visualization 7, Visualization 8, Visualization 9 and Visualization 10). (a) Photography of the human face captured by the CMOS camera (associated with Visualization 6); (b) Thermal image captured by thermal camera at the same time (associated with Visualization 7); (c) One frame of the 3D reconstructed geometry (associated with Visualization 8); (d) The same frame of temperature mapping result (associated with Visualization 9). Color represents temperature ranging from 297 to 305 K in both (b) and (d).

4. Summary

This paper has presented a high-resolution, real-time simultaneous 3D geometric shape and temperature measurement method. We developed a holistic approach to calibrate both structured light system and thermal camera under exactly the same world coordinate system even though these two sensors do not share the same wavelength; and a computational framework to determine the sub-pixel corresponding temperature for each 3D point as well as discard those occluded points. Experiments verified the accuracy of our algorithm, and we demonstrated that the proposed method can be applied in real time applications.

Acknowledgments

We would like to thank Tyler Bell for proofreading and serving as the model for testing the real time system. Also, we would like to thank Ziping Liu and Chufan Jiang for helping to do real time experiments. Last but not least, we would like to thank Beiwen Li, Jaesang Hyun and Huitaek Yun for the generous suggestions about algorithm development and hardware design.

This study was sponsored by the National Science Foundation (NSF) Directorate for Engineering (100000084) under grant number: CMMI-1521048. The views expressed in this paper are those of the authors and not necessarily those of the NSF.