Real-time mobile 3D temperature mapping

Moghadam, Peyman; Vidas, Stephen; Sridharan, Sridha

2014

Vidas, S., Moghadam, P., & Sridharan, S. (2014). Real-time mobile 3D temperature mapping. IEEE sensors journal, 15(2), 1145-1152.

<https://hdl.handle.net/10356/106987>

<https://doi.org/10.1109/JSEN.2014.2360709>

© 2014 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works. The published version is available at: [<http://dx.doi.org/10.1109/JSEN.2014.2360709>].

Downloaded on 26 Apr 2023 20:03:48 SGT

Real-time Mobile 3D Temperature Mapping

Stephen Vidas, *Member, IEEE*, Peyman Moghadam, *Member, IEEE*, and Sridha Sridharan, *Senior Member, IEEE*

Abstract—The ability to measure surface temperature and represent it on a metrically accurate 3D model has proven applications in many areas such as medical imaging, building energy auditing, and search and rescue. A system is proposed that enables this task to be performed with a handheld sensor, and for the first time with results able to be visualized and analyzed in real-time. A device comprising a thermal-infrared camera and range sensor is calibrated geometrically and used for data capture. The device is localized using a combination of ICP and video-based pose estimation from the thermal-infrared video footage which is shown to reduce the occurrence of failure modes. Furthermore, the problem of misregistration which can introduce severe distortions in assigned surface temperatures is avoided through the use of a risk-averse neighborhood weighting mechanism. Results demonstrate that the system is more stable and accurate than previous approaches, and can be used to accurately model complex objects and environments for practical tasks.

Index Terms—3D mapping, handheld, HeatWave, localization, mobile, real-time, temperature mapping, thermography

I. INTRODUCTION

Thermal-infrared cameras allow the passive and noncontact measurement of surface radiation which can be used to accurately estimate temperatures and model temperature distributions in environments. By representing this temperature information on an accurate 3D model, estimates can become both more certain and more useful. The increase in certainty is primarily achieved through the combining of measurements from multiple viewpoints to reduce noise. The increase in utility is largely due to the ability to combine both temperature and geometric information together, for example to enable the calculation of energy loss as a function of both temperature and area.

Methods for measuring and representing surface temperature on 3D models have been presented in areas such as medical imaging [1], disaster response [2] and building energy auditing [3]–[5], with this work focusing on the latter.

The proposed system enables the task of 3D thermography to be performed using a single handheld device. In contrast to previous work, the 3D model with incorporated temperature information (an example shown in Figure 1) can be processed and visualized to users as it is captured in real-time. Temporal changes in surface temperature can also be detected in real-time because of the high framerate of the thermal-infrared camera (approximately 80 frames per second).

The device used for data capture is called HeatWave and is comprised of a thermal-infrared camera, color camera and a

Fig. 1: A generated 3D temperature model of an additive manufacturing system using the HeatWave device. The colorbar indicates the mapping between colors and true temperature values.

range sensor [4], [5]. Localization of the sensor in 3D space for each captured frame is crucial for accurate results, and is achieved through the consideration of pose estimates obtained by separate algorithms that apply to the range and thermal-infrared data. This utilization of both sensor modalities is shown to reduce the occurrence of failure modes which prevent more primitive systems from functioning effectively.

Furthermore, the problem of misregistration which can cause severe inaccuracies in assigned surface temperatures is avoided through the use of a risk-averse neighborhood weighting mechanism. Results demonstrate that the system is more stable and accurate than previous approaches, and can be used to accurately model complex objects and environments for practical tasks.

II. BACKGROUND

Estimating surface temperature using 2D thermal-infrared imagery is a well-researched area, with sensors capable of detecting most of the radiation emitted from objects at typical terrestrial temperatures (e.g. around -20 to 100 degrees Celsius). The Stefan-Boltzmann Equation can then be used to convert the radiation measurement to a temperature estimate, although to maximize accuracy other environmental factors such as air temperature and possible reflections should be considered.

Low-cost, handheld technologies for obtaining accurate 3D information from the environment are also available on the market, many of which are based on technology developed by PrimeSense [6]. These RGB-D or range devices have the capacity to be incorporated into a 3D reconstruction framework to achieve dense 3D models beyond what is possible from a single view. The most successful of these frameworks is the

Stephen Vidas is with the School of Electrical and Electronic Engineering, NTU, Singapore, e-mail: svidas@ntu.edu.sg.

Peyman Moghadam is with Autonomous Systems, CSIRO, Australia, e-mail: peyman.moghadam@csiro.au.

Sridha Sridharan is with the Science and Engineering Faculty, QUT, Australia, e-mail: sridha.sridharan@qut.edu.au.

KinectFusion algorithm, which forms a key component of the 3D temperature mapping solution proposed in this paper [7].

For the task of 3D model generation using a range sensor, effective estimation of the 6 Degree of Freedom (DoF) pose of the sensor is crucial. With an accurate estimate of pose, the estimation of 3D structure is relatively straightforward.

The most common method for estimating the 6 DoF pose involves the use of the Iterative Closest Point (ICP) algorithm. This is generally used to align each successive range-image to either the previous frame, or to a global model that is progressively optimized [7]. ICP attempts to estimate a geometric transformation that minimizes the Euclidean distance d (1) between corresponding points in the target and source models.

$$d = \mathbf{q} - \mathbf{p} = \sqrt{(q_x - p_x)^2 + (q_y - p_y)^2 + (q_z - p_z)^2} \quad (1)$$

Here $\mathbf{q}$ is the 3D point in the source model and $\mathbf{p}$ is the 3D point in the target model.

However, ICP-based methods have many well-documented failure modes which are fatal to systems relying solely on this method of pose estimation [8], [9]. As a result, the integration of additional methods for estimating pose is an active area of research. Visual odometry depends solely on video data from the integrated color camera and not on range data, and has been proven to enhance the stability and accuracy of range-based mapping systems [10]. Alternatively, approaches which directly consider both range and color data when performing alignment have also achieved good results [11]. It should be noted that in poor lighting conditions, including with no lighting at all, neither of these methods can offer any advantage over a simple ICP-only approach.

As an alternative to 3D modeling based on range output, methods also exist for reconstructing both the motion of the device and portions of 3D surfaces using only 2D video data. Classical methods for performing this can be described as feature-based, in that they depend on tracking local points of interest (such as corners and other stable and identifiable points) across consecutive frames [12]. With good feature tracking results, an estimate can be made of the initial shift in translation and rotation early in the captured data sequence [13]. The 3D locations of tracked 2D features can then be estimated using one of many available triangulation methods that utilize known (or estimated) camera poses. These estimated point locations can then be used to estimate the pose of new camera views, and the cycle can continue.

III. RELATED WORK

Attention to the problem of automatic 3D temperature mapping has recently increased, particularly in the medical, energy efficiency and robotics research communities. A brief review of alternative 3D thermography approaches for building energy auditing is provided by [4]. It broadly classifies the major approaches as being centered around LiDAR technologies [14], robotic platforms [3] or image-only configurations [15]. Each of these approaches has disadvantages, however, and so an alternative method which utilizes a relatively cheap range

Fig. 2: HeatWave 3D thermography device

sensor is suggested by [4]. This method forms the basis of comparison for the proposed work.

Research into the integration of non-visible imaging sensors into 3D mapping systems has so far received relatively little attention. What work has been done in this area has demonstrated that 3D thermography incorporating cheap range sensors has powerful advantages over existing alternatives [4], [5], [16]. These include a reduced cost and development time, reduced setup and data processing time, a flexibility in the motion of the device that allows complex environments to be traversed and mapped, and a resilience to difficult lighting conditions.

The use of the thermal-infrared imaging modality as a means for localization has only recently become an area of active research. Monocular, geometric SLAM using only thermal-infrared video has been demonstrated to be possible, albeit fraught with danger due to commonly occurring failure modes [17], [18]. However, it has been demonstrated that the thermal modality has a complementary nature when combined with conventional localization techniques operating on the visible-spectrum [19]. This suggests that these failure modes tend not to coincide with those of existing methods, and therefore that utilizing the thermal-infrared modality when it is available has the potential to improve the performance of 3D mapping approaches.

IV. SYSTEM

The proposed system utilizes HeatWave: a mobile handheld 3D thermography device consisting of a thermal-infrared camera and an RGB-D range sensor, rigidly attached together in an ergonomic form factor as shown in Figure 2 [4], [5]. The RGB-D sensor itself contains an RGB camera, near-infrared camera and near-infrared projector. The near infrared projector (actuator) and camera (sensor) pair is used to extract depth information from the scene, while the RGB image is registered onto this output. This allows color and range measurements to be made simultaneously for surfaces in the shared field of view of the cameras.

Figure 3 shows the interaction between the algorithms that comprise the proposed system. The environment is explored by the operator using the HeatWave device to record or stream data. Each video feed is processed by a separate pose-estimation algorithm that is specialized for that data type. For the Video-Based Pose Estimation (VBPE) algorithms,

Fig. 3: System Diagram. The Iterative Closest Point (ICP) and Video-Based Pose Estimation (VBPE) algorithms are used to mutually enhance the corresponding pose estimates. Many of the algorithms are integrated into a single “Mapper” framework to make use of shared memory on the host computer.

estimates of pose from the depth (ICP-based) pose estimation algorithm (if available) are used for motion prediction. In turn, VPBE results are used to guide the ICP algorithm in processing new depth frames. The final pose estimate is then used to update the 3D model, both in terms of its physical structure, and its surface properties. To update the physical structure, an implementation of Volumetric Range Image Processing (VRIP) is used which optimizes a 3D representation of the scene as represented by voxels. The assignment of colors and temperatures to the surface of the model is done by separate raycasters that take advantage of additional motion, geometric and temporal information to enhance the accuracy of the estimates. The multi-modal representation of the 3D surfaces can then be displayed to the user.

An example of the real-time visualization available to the user is shown in Figure 4. All software follows the protocols of the open-source Robotics Operating System (ROS) which enables the utilization of both new and improved versions of tools and methods as they are developed and shared by the community.

Calibration of the proposed mobile thermography system involved determining both the intrinsic properties of the individual sensors, and the extrinsic spatial relationship between sensors. For the process of extrinsic calibration, the mask-based approach of [20] was used, together with a method of determining the temporal offset between the sensors first proposed in [16]. Similarly, intrinsic calibration was performed using the same mask-based approach. Candidate calibration frames were selected using a combination of the Enhanced-Monte-Carlo approach and a spatially weighted sampling mechanism designed to reduce the risk of overfitting [21].

No emissivity calibration was performed for the thermal camera, and the emissivity of all surfaces was assumed to be constant at a value of 1.00. However, the existence of the detailed 3D model as a result of the proposed algorithm provides the opportunity to segment and label different surfaces, enabling unique emissivity estimates to be assigned for different materials. Post-processing could then be used to further enhance the estimate of surface temperatures based on

this information.

V. POSE ESTIMATION

In the proposed approach, the basis for obtaining an estimate of the pose of the range sensor remains the use of ICP and VRIP to align each depth image with the constantly updated 3D voxel representation. However, due to commonly occurring failure modes when using this approach, an additional Video-Based Pose Estimation (VBPE) method is used simultaneously. This video-based localization method is initialized by and closely coupled with the ICP-based localization method, but is capable of operating independently for some time if the ICP-based method fails or produces less satisfactory pose estimates. Estimates of sensor pose from the VBPE algorithm can then be used to guide the ICP algorithm as it attempts to recover.

A. Video-Based Pose Estimation

The Video-Based Pose Estimation (VBPE) approach employed was an extension of an algorithm originally published in [18]. For the thermal-infrared video, 14-bit monochromatic images were streamed from the camera which were quantized with a difference of one bit representing a difference of 0.1 degrees Celsius.

Feature tracking was performed using GFTT [22] and a sparse iterative version of the Lucas-Kanade optical flow algorithm [12]. For the special case of the thermal-infrared camera, the problem of Non-Uniformity Corrections (periodic interruptions of thermal data for recalibration) was addressed using the homography-based recovery mechanism proposed by [18].

The localization thread subscribes to ICP-based estimates of the RGB-D camera pose, and combines these with a constant velocity assumption to predict the pose of the thermal-infrared camera for each received thermal-infrared image. To make this prediction, linear interpolation is first applied to the RGB-D camera positions and orientations to find the RGB-D camera pose for times corresponding to captured thermal-infrared

Fig. 4: Real-time visualization of a 3D surface temperature model generated by the proposed approach. The open-source visualization software ‘rviz’ was used, which does not have the in-built capacity to display spatial scale or a temperature legend. However, necessary information for these features is inherent in the model, and easily accessible by custom software.

frames. The known and fixed spatial transformation that relates the physical position and orientation of the RGB-D camera to the thermal-infrared camera when the HeatWave device is stationary is then applied. This motion prediction model was found in general to achieve a valid initial estimate for later refinement. However, in cases where no valid ICP-based pose estimates are received, the 3D pose estimation algorithm simply applies the constant velocity assumption to its own previous two pose estimates in order to form a prediction.

Once an initial estimate of the pose of a new image has been received, the Perspective-n-Point (PnP) algorithm using keyframes is used to further refine the estimate. Following this, sparse bundle adjustment [23] also using the keyframes is applied, which under favorable conditions can reduce the pixel reprojection error to below 1.0 [24].

After the round of bundle adjustment, 2D features which appear in the current image but have not yet had their 3D locations estimated are triangulated if there are sufficient (more than 3) viewpoints. The procedure for point triangulation involves a clustering of estimates based on multiple viewpoint pairs. All candidate triangulations are considered, with the final triangulation estimate calculated as the mean of all triangulations in the largest cluster. Clusters are formed by grouping all points that are separated by one another by less than 0.3m. This threshold was selected empirically to work effectively with the chosen range sensor, and the typical distances to surfaces in the scene that are encountered when performing 3D temperature scans. Each individual triangula-

tion was performed using the Iterative Linear Least Squares approach [25].

Keyframes were favored over using all frames for the purpose of improving the robustness of the estimation, reducing the time to convergence, and limiting the increase in memory requirement and processing time as more frames are acquired. Each time a frame is rejected as a keyframe, or an existing keyframe is replaced by a new candidate, the memory allocated for storing the 2D and 3D position estimates for features that were no longer present in the keyframes is freed.

In selecting a keyframe, first it was determined if there was sufficient connectivity to existing keyframes in terms of a significant number of common tracked features. A minimum of 8 features shared between the candidate and the existing set of keyframes was found empirically to work well.

Given a candidate keyframe that had passed the connectivity test, the final keyframe selection algorithm shown in Figure 5 was used to update the keyframe set. The minimum required and ideal number of keyframes were set as 5 and 15 respectively, which were found empirically to strike a good balance between efficiency and stability.

B. Pose Estimate Fusion

An efficient method was employed to utilize pose estimates from the VBPE algorithm as a means for supporting the primary ICP-based pose estimation framework. An advantage of this method is that it can enable the successful operation of the 3D mapping system even when computational limitations

Fig. 5: Keyframe Selection Procedure. The abbreviation ‘KF’ is used for Keyframe.

result in lower range-image processing framerates than what would normally be necessary. This is because VBPE is much less computationally expensive than ICP.

In general, the previous pose estimate acquired from ICP was used as an initial starting point for performing the ICP algorithm upon a subsequent range image. Under small motions and with ideal scenes containing a large variety of unrepeated 3D texture, this starting point generally leads to convergence to the global optimal solution, and results in good system performance. However, with large motions between frames or non-ideal scenes, the ICP algorithm may fail to accurately converge from this initial starting estimate. To combat this, under certain circumstances the pose estimate from the VBPE algorithm may be preferred for initialization. For the proposed system, these circumstances were defined as those where the confidence of the pose estimate from the VBPE algorithm exceeded that of the previous ICP result.

The confidence scores for pose estimates from VBPE were defined as either 0 or 1 by using empirically determined thresholds for a number of key metrics. Therefore, if any metric violated its threshold, the VBPE estimate would not be considered valid for initializing the subsequent ICP procedure. The metrics are described in Table I.

The confidence for ICP-based pose estimates was defined as either 0.5 or 1. A confidence of 0.5 was allocated if the ICP-based pose estimate was based on a range image received more than 0.2 seconds in the past. Otherwise, a full confidence of 1.0 was assigned. The fusion can therefore be summarized as using VBPE to initialize ICP only if the VBPE result was within the limits of a set of defined thresholds, and the last ICP procedure was performed more than 0.2 seconds ago.

The opportunity exists for a more sophisticated mechanism for preferring one pose estimate initialization method over the other, however, this has been reserved for future work as discussed in Section VIII.

Fig. 6: Pixel-threaded volumetric raycasting for occlusion handling. The illustrated method is implemented individually for each depth image pixel in parallel on the computer GPU.

VI. TEMPERATURE ESTIMATION

For estimating the surface temperatures on the generated 3D model, the weighted raycasting method of [4], [5] was used. However, two key improvements are proposed:

- A transition from voxel-based to pixel-based GPU-threading, to more effectively handle occlusion.
- A novel, risk-averse neighborhood weighting mechanism to reduce errors associated with registration inaccuracy.

This combination of Occlusion Handling and Misregistration Avoidance is from here on referred to as OHMA.

A. Occlusion Handling

The conventional method of VRIP used in works such as [7] and [4] assumes an environment without significant occlusion. Voxels are tested for intersected rays originating from the image in a layer-by-layer fashion. As a result, in many environments, colors or temperatures assigned to foreground objects may also be incorrectly assigned to background surfaces directly behind them. To address this, a pixel-based approach was employed which is outlined in Figure 6. This approach searches for voxels in the paths of rays corresponding to each pixel at progressively larger distances from the camera, and terminates the search once an intersecting voxel has been found. The search distance starts at 0, and increments by the value d_{step} up to a maximum of the depth of the working volume d_{max} . For the proposed system, d_{step} was set to equal the depth of a single voxel (approximately 6mm), 512 times smaller than d_{max} , which was typically set to 3m. The calculation of the 3D point location corresponding to a pixel and a distance is done using conversion from projective camera space to metric world space, and is shown in (2).

$$x_m = z_m \frac{x_p - c_x}{f_x}, \quad y_m = z_m \frac{y_p - c_y}{f_y} \quad (2)$$

Here, the subscript m defines the metric space while p is used for camera space (where the x and y co-ordinates correspond to the column and row of the pixel in the original image). c_x and c_y are the co-ordinates of the camera center,

TABLE I: Summary of VBPE metrics and their corresponding thresholds

Metric	Description	Threshold
BA Error	Reprojection Error (in pixels) of 3D points onto images after bundle adjustment.	3.00 (max)
Translation Shift	Translation distance (in m) between pose estimated by PnP algorithm and final BA result.	0.50 (max)
Rotation Shift	Orientation distance (in deg) between pose estimated by PnP algorithm and final BA result.	90 (max)
Inlier Proportion	Proportion of 3D points reprojected within 5.0 pixels of their image location by PnP.	0.75 (min)

and f_x and f_y are the focal lengths. These camera parameters are determined using the camera calibration procedure discussed in Section IV.

To convert from the metric coordinates with respect to the image (C) to those with respect to the 3D model (W), the known translation and orientation of the current pose can be applied as shown in (3).

$$\begin{bmatrix} X_W \\ Y_W \\ Z_W \end{bmatrix} = \mathbf{R} \begin{bmatrix} X_C \\ Y_C \\ Z_C \end{bmatrix} + \mathbf{t} \quad (3)$$

Here, $\mathbf{R}$ is the rotation matrix and $\mathbf{t}$ is the translation vector which describe the pose of the camera relative to the origin (the initial starting pose).

The algorithm can be interpreted as projecting a set of rays (one from each pixel) toward the volume, and incrementing the distance from the camera for each of these rays. As soon as a surface is encountered, temperature or color assignment is performed, and the remaining extent of the ray is no longer traversed. The outcome is that surfaces intersected by the ray that are behind other surfaces are not considered valid candidates for new temperature or color assignments, and therefore the problem of occlusion is theoretically avoided. Unfortunately, practical inaccuracies in pose estimation and calibration mean that occlusions can still have harmful effects. However, a method for reducing these is proposed in Section VI-B.

B. Misregistration Avoidance

Due to potential errors in calibration or pose-estimation, for many images the pose-based registration to the existing 3D model will contain some error. This will cause some spatial misassignments of pixel values to voxels. In many cases this will have a negligible effect, because the voxels corresponding to adjacent pixels in the image may exist on the same physical surface and have an identical or similar temperature. However, in some cases gross errors can occur, for example as a temperature that belongs to a foreground object is incorrectly assigned to a surface in the background. These gross errors occur for image regions that correspond to the neighborhoods of depth discontinuities from the perspective of the thermal-infrared camera.

To reduce and potentially avoid gross errors due to misregistration, a cautionary edge filter has been added to the filter suite proposed by [4], [5]. This edge filter implements a novel method for assigning reduced confidence levels for rays that are at more risk of being assigned to incorrect surfaces. This ensures that unreliable estimates of surface temperature based on high-risk rays are only used in the absence of more

Fig. 7: Cautionary Edge Filter.

confident estimates that have been made from more favorable camera poses. The method is outlined in Figure 7.

In this algorithm, a maximum neighborhood size a of 10 was selected, which dictates the maximum distance away (in pixels) that sample depths will be from the pixel of interest. The ‘edge-score’ assigned to a pixel is effectively the maximum difference in depth between the pixel of interest and any of its sampled neighbor pixels, up to an empirically determined maximum d_{max} of 10cm. Furthermore, if the sampled pixel is nearer to the camera than the pixel of interest, the edge-score for the pixel is also set to d_{max} . After completing a survey of all neighbor pixels, the edge-score is linearly mapped to a weighting, with d_{max} mapping to a weighting of zero and a difference of zero mapping to a weighting of one.

However, in storing the weighting in memory as an 8-bit integer, weightings of zero are converted to the minimum nonzero value of 1. This ensures that a zero confidence level is reserved for a voxel that does not have a single intersecting ray, while enabling tentative temperature assignments where none previously exist. A demonstration in the effectiveness of this approach is provided in Section VII-B.

VII. EVALUATION

For the evaluation, the current HeatWave device which includes an ASUS Xtion Pro Live as the RGB-D sensor was used. It has a framerate of 30 frames per second and outputs images at a resolution of 640 by 480 pixels.

The thermal-infrared camera integrated as part of the HeatWave device is an Optris PI450. This model outputs images at up to 80 frames per second with a resolution of 384 by 288 pixels. Pixel depth is 14-bit, with a single bit representing an approximate temperature change of 0.1 degree Celsius, although this may vary with different firmware versions. The Noise-Equivalent Differential Temperature is quoted by the manufacturer to be 0.4 degrees Celsius.

Using a modern laptop equipped with a GeForce GTX 680M graphics card, a spatial resolution of approximately 4mm for an 8 cubic meter volume was achieved for each result. Experiments were designed to determine the following:

- (A) Is the system is more stable than the previous approach?
- (B) Can the system achieve more accurate results than the previous approach?

A. Experiment in Stability

This experiment aimed to test the capacity of the thermal-infrared video-based localization method to enable the 3D thermography system to continue mapping effectively during failure of the conventional ICP-based localization.

A sequence of 30 seconds was captured of an industrial HVAC environment. The sequence was digitally manipulated to alternately permit and exclude 0.5 second blocks of range data. This was to simulate circumstances in which the range sensor is physically occluded, or data from the range sensor is unavailable such as due to distant (out-of-range) surfaces or a lack of 3D texture in the scene.

Figure 8 compares the translation pose estimates for the case of using only ICP-based pose estimates (ICP-only) and of using the video-based localization to provide pose estimates while the ICP method cannot (video-supported).

From the figure, it can be seen that the proposed video-supported method enables errors to be kept minimal (under 2.0 cm) for a significantly longer duration than what is possible using the conventional ICP-only method. This is because when range data is interrupted, a reasonable estimate of pose is able to be maintained using video-based localization, so that ICP can have a good estimate for re-registering once valid range images become available again. However, there are cases (such as close to the end of the sequence) where circumstances mean that even with video-based support, pose estimation is unable to be recovered after an interruption to the range data.

B. Experiment in Accuracy

In order to evaluate the effectiveness of the proposed Occlusion Handling and Misregistration Avoidance (OHMA) framework, a simple desktop data sequence of approximately 30 seconds duration was collected. This sequence focused on a single laptop, and contained views of the laptop that were both unoccluded and occluded by other surfaces, as shown in Figure 9.

Fig. 8: Comparison of pose translation estimate error between video-supported and conventional approaches.

Fig. 9: Two images from the test sequence

Fig. 10: Comparison of resulting models

The resulting, rendered 3D models are shown in Figure 10. It can be seen that the proposed OHMA framework avoids gross temperature assignment errors where foreground temperatures are incorrectly assigned to background objects. The phenomenon of this gross misassignment is best illustrated with video, however, it can be summarized as the accumulation of smaller misassignments as the object of interest becomes increasingly occluded by a foreground surface.

VIII. CONCLUSIONS AND FUTURE WORK

The first handheld 3D thermography system capable of performing in real-time has been presented. Video-based localization using thermal-infrared video has been shown to offer benefits in improving the stability of the system beyond what can be achieved using ICP only. This can be especially useful for reducing the occurrence of failure modes which have been problematic for previous 3D thermography approaches.

A novel Occlusion Handling and Misregistration Avoidance method has been proposed and demonstrated to be effective in improving the accuracy of final surface temperature models beyond what was previously achievable.

Potential avenues of development for future generations of the 3D thermography system include:

- Exploring machine learning to optimize the confidence ratings of pose estimates from different modalities, which will enable a more intelligent fusion of data from different modalities to optimize stability.
- Use of a particle filter or Extended-Kalman Filter (EKF) to improve the accuracy by incorporating pose estimates from multiple modalities.
- Incorporating an environmental radiation model to optimize surface temperature estimates and enable the estimate of additional surface properties such as emissivity.
- Extending the method to operate effectively at different spatial scales, and to incorporate additional information such as thermorefectance, for applications such as analyzing the cooling of microelectronics.

ACKNOWLEDGMENT

The authors gratefully acknowledge funding of the project by the CSIRO Realtime Perception project, CSIRO Event-Driven Mobile Node project and CSIRO TCP SSN. The institutional support of both CSIRO and QUT is also greatly appreciated.

REFERENCES

- [1] B. Lahiri, S. Bagavathiappan, T. Jayakumar, and J. Philip, "Medical applications of infrared thermography: A review," *Infrared Physics & Technology*, vol. 55, no. 4, pp. 221–235, Jul. 2012. 1
- [2] K. Nagatani, K. Otake, and K. Yoshida, "Three-dimensional Thermography Mapping for Mobile Rescue Robots," *Field and Service Robotics*, vol. 92, pp. 49–63, 2014. 1
- [3] D. Borrmann, A. Nuchter, M. Dakulovic, I. Maurovic, I. Petrovic, D. Osmankovic, and J. Velagic, "The Project ThermalMapper – Thermal 3D mapping of indoor environments for saving energy," *Proceedings of the 10th International IFAC Symposium on Robot Control (SYROCO)*, vol. 10, 2012. 1, 2
- [4] S. Vidas and P. Moghadam, "HeatWave: A handheld 3D thermography system for energy auditing," *Energy and Buildings*, vol. 66, pp. 445–460, Nov. 2013. 1, 2, 5, 6
- [5] P. Moghadam and S. Vidas, "Heatwave: the next generation of thermography devices," in *Thermosense: Thermal Infrared Applications XXXVI*, vol. 9105. Proc. SPIE, 2014. 1, 2, 5, 6
- [6] S. Meister, S. Izadi, and P. Kohli, "When can we use KinectFusion for ground truth acquisition?" *Workshop on Color-Depth Camera Fusion in Robotics, IROS*, pp. 3–8, 2012. 1
- [7] S. Izadi, R. A. Newcombe, D. Kim, O. Hilliges, D. Molyneaux, S. Hodges, P. Kohli, J. Shotton, A. J. Davison, and A. Fitzgibbon, "KinectFusion : Real-Time Dynamic 3D Surface Reconstruction and Interaction," *ACM SIGGRAPH*, 2011. 2, 5
- [8] M. Dou, L. Guan, J.-M. Frahm, and H. Fuchs, "Exploring High-Level Plane Primitives for Indoor 3D Reconstruction with a Hand-held RGB-D Camera," *ACCV 2012 Workshops*, pp. 94–108, 2013. 2
- [9] L. Douadi, M.-J. Aldon, and A. Crosnier, "Pair-wise Registration of 3D/Color Data Sets with ICP," *2006 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 663–668, Oct. 2006. 2
- [10] T. Whelan, H. Johannsson, M. Kaess, J. Leonard, and J. McDonald, "Robust real-time visual odometry for dense RGB-D mapping," *IEEE International Conference on Robotics and Automation (ICRA)*, pp. 5724–5731, 2013. 2
- [11] B. Peasley and S. Birchfield, "Replacing Projective Data Association with Lucas-Kanade for KinectFusion," *2013 IEEE International Conference on Robotics and Automation*, pp. 638–645, May 2013. 2
- [12] J. Bouguet, "Pyramidal Implementation of the Affine Lucas Kanade Feature Tracker," *Intel Corporation*, 2001. 2, 3
- [13] G. Klein and D. Murray, "Parallel Tracking and Mapping for Small AR Workspaces," *2007 6th IEEE and ACM International Symposium on Mixed and Augmented Reality*, pp. 225–234, Nov. 2007. 2
- [14] D. González-Aguilera, P. Rodríguez-Gonzálvez, J. Armesto, and S. Lagüela, "Novel approach to 3D thermography and energy efficiency evaluation," *Energy and Buildings*, vol. 54, pp. 436–443, Nov. 2012. 2
- [15] Y. Ham and M. Golparvar-Fard, "Rapid 3D Energy Performance Modeling of Existing Buildings Using Thermal and Digital Imagery," *Construction Research Congress 2012*, pp. 991–1000, May 2012. 2
- [16] S. Vidas, P. Moghadam, and M. Bosse, "3D thermal mapping of building interiors using an RGB-D and thermal camera," *2013 IEEE International Conference on Robotics and Automation*, pp. 2311–2318, May 2013. 2, 3
- [17] A. Gonzalez, M. Devy, and J. Sola Ortega, "SLAM visuel monoculaire par caméra infrarouge," *ORASIS*, 2011. 2
- [18] S. Vidas and S. Sridharan, "Hand-held monocular SLAM in thermal-infrared," *International Conference on Control, Automation, Robotics and Vision*, pp. 859–864, 2012. 2, 3
- [19] W. Maddern and S. Vidas, "Towards robust night and day place recognition using visible and thermal imaging," *RSS: Beyond laser and vision*, 2012. 2
- [20] S. Vidas and R. Lakemond, "A mask-based approach for the geometric calibration of thermal-infrared cameras," *IEEE Transactions on Instrumentation and Measurement*, pp. 1625–1635, 2012. 3
- [21] J.-n. Ouellet and P. Hebert, "Developing assistant tools for geometric camera calibration: assessing the quality of input images," *Proceedings of the 17th International Conference on Pattern Recognition*, vol. 4, pp. 80–83, 2004. 3
- [22] J. Shi and C. Tomasi, "Good Features to Track," *Computer Vision and Pattern Recognition (CVPR)*, pp. 593–600, 1994. 3
- [23] C. Engels, H. Stewénius, and D. Nistér, "Bundle adjustment rules," *Photogrammetric computer vision*, 2006. 4
- [24] K. Konolige, "Sparse Sparse Bundle Adjustment," *Proceedings of the British Machine Vision Conference 2010*, pp. 102.1–102.11, 2010. 4
- [25] R. I. Hartley and P. Sturm, "Triangulation," *Computer Vision and Image Understanding*, vol. 68, no. 2, pp. 146–157, Nov. 1997. 4