

# A Mask-Based Approach for the Geometric Calibration of Thermal-Infrared Cameras

Stephen Vidas, *Member, IEEE*, Ruan Lakemond, *Member, IEEE*, Simon Denman, *Member, IEEE*, Clinton Fookes, *Member, IEEE*, Sridha Sridharan, *Senior Member, IEEE*, and Tim Wark, *Senior Member, IEEE*

**Abstract**—Accurate and efficient thermal-infrared (IR) camera calibration is important for advancing computer vision research within the thermal modality. This paper presents an approach for geometrically calibrating individual and multiple cameras in both the thermal and visible modalities. The proposed technique can be used to correct for lens distortion and to simultaneously reference both visible and thermal-IR cameras to a single coordinate frame. The most popular existing approach for the geometric calibration of thermal cameras uses a printed chessboard heated by a flood lamp and is comparatively inaccurate and difficult to execute. Additionally, software toolkits provided for calibration either are unsuitable for this task or require substantial manual intervention. A new geometric mask with high thermal contrast and not requiring a flood lamp is presented as an alternative calibration pattern. Calibration points on the pattern are then accurately located using a clustering-based algorithm which utilizes the maximally stable extremal region detector. This algorithm is integrated into an automatic end-to-end system for calibrating single or multiple cameras. The evaluation shows that using the proposed mask achieves a mean reprojection error up to 78% lower than that using a heated chessboard. The effectiveness of the approach is further demonstrated by using it to calibrate two multiple-camera multiple-modality setups. Source code and binaries for the developed software are provided on the project Web site.

**Index Terms**—Calibration, cameras, geometry, infrared (IR) image sensors, IR imaging.

Manuscript received July 4, 2011; revised September 22, 2011; accepted November 14, 2011. Date of publication February 3, 2012; date of current version May 11, 2012. This work was supported in part by a Queensland Government Smart Futures PhD Scholarship and in part by a Commonwealth Scientific and Industrial Research Organisation Office of the Chief Executive Scholarship. The Associate Editor coordinating the review process for this paper was Dr. Zheng Liu.

S. Vidas is with the Speech, Audio, Image and Video Technologies Group, Information Security Institute, Queensland University of Technology, Brisbane, Qld. 4000, Australia, and also with the Autonomous Systems Laboratory, Information and Communication Technologies Centre, Commonwealth Scientific and Industrial Research Organisation, Queensland Centre for Advanced Technologies, Pullenvale, Qld. 4069, Australia (e-mail: stephen.vidas@qut.edu.au; steve.vidas@csiro.au).

R. Lakemond, S. Denman, C. Fookes, and S. Sridharan are with the Speech, Audio, Image and Video Technologies Group, Information Security Institute, Queensland University of Technology, Brisbane, Qld. 4000, Australia (e-mail: r.lakemond@qut.edu.au; s.denman@qut.edu.au; c.fookes@qut.edu.au; s.sridharan@qut.edu.au).

T. Wark is with the Autonomous Systems Laboratory, Information and Communication Technologies Centre, Commonwealth Scientific and Industrial Research Organisation, Queensland Centre for Advanced Technologies, Pullenvale, Qld. 4069, Australia (e-mail: t.wark@csiro.au).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TIM.2012.2182851

## I. INTRODUCTION

ACCURATE geometric camera calibration must be performed for both single- and multiple-device imaging systems as a precursor to many existing computer vision algorithms. Geometric camera calibration is concerned with determining the camera parameters which map a 3-D scene onto a 2-D image and is a separate problem to spectral or intensity calibration. Although spectral calibration is a useful process for many applications, it is, in many cases, not needed for the effective implementation of computer vision algorithms and is not included in the scope of this paper.

Algorithms that greatly benefit from accurate geometric calibration include those for feature matching, localization, and 3-D reconstruction. Such algorithms form critical components in systems applied in domains such as surveillance, medical assessment, and industrial inspection. The different characteristics of thermal-infrared (IR) cameras make them superior or complementary to conventional visible-spectrum cameras for many applications. For example, thermal-IR cameras are more robust to poor lighting or lighting changes and also to dust and fog [1]. However, these differences also affect the process of calibration and, therefore, the accuracy of results.

Multiple-camera setups involving combinations of thermal and other cameras are an exciting platform for computer vision research and applications. These setups allow investigations into fusion techniques to enhance algorithm performance. The effectiveness of these systems has been demonstrated for many tasks such as hazard awareness [2], thermal analysis [3], and driver activity analysis [4]. To advance research in this area, it is important that an accurate and efficient geometric calibration procedure exists that is able to accommodate both modalities simultaneously. Such calibration can be used to spatially undistort and rectify imagery, as well as determine the relative poses of the cameras.

The purpose of the proposed approach is to provide a convenient but accurate method of geometric camera calibration for both single- and multiple-camera systems involving thermal-IR cameras. Existing approaches for this task have many limitations in terms of their accuracy, convenience, and flexibility. The proposed system is an adaptation of Zhang's method [5] and is available in [6]. The key modifications are the use of a new easily prepared calibration pattern, the use of a new automatic pattern-finding algorithm, and the implementation of automatic frame selection. The proposed approach is structured as follows.

- 1) A calibration pattern is manufactured and prepared.

- 2) An input video or image sequence of the proposed calibration pattern at different angles is provided by the user for each camera (or all cameras together).
- 3) Calibration points in each frame are located automatically using the proposed pattern-finding algorithm.
- 4) A near-optimal subset of valid frames is selected using an implementation of the enhanced Monte Carlo method (MCM) approach [7].
- 5) An optimization algorithm is used to fit a distortion model (for intrinsic calibration) or a six-parameter pose model (for extrinsic calibration) to the data.

The proposed method makes calibration of thermal cameras and multimodality camera setups easier and more accurate than the current most popular approach. It can be used in a number of scenarios where competing methods are unfeasible, such as when additional powered equipment is unavailable or when camera systems involve just one or two cameras.

This paper is structured as follows. In Section II, a background into the process of camera calibration and existing methods is provided. Section III presents the new proposed calibration pattern. In Section IV, the proposed pattern-finding algorithm is outlined. Section V presents the proposed calibration system implementation. In Section VI, the multiple-camera multiple-modality setups used for the evaluation are presented. Section VII contains an evaluation of the contributions outlined in this paper. This paper is concluded in Section VIII.

## II. BACKGROUND

### A. Spectral Calibration

Although this paper deals with the problem of geometric or spatial calibration, the related problem of spectral or intensity calibration (also referred to as radiometric or photometric calibration) deserves some attention. Spectral calibration refers to the process of forming a quantitatively accurate mapping between the pixel intensity values of an image and the actual power incident on the imaging sensor. This can be relevant for visible-spectrum cameras but is particularly useful for thermal-IR cameras as it enables accurate estimates of surface temperatures to be made from the image data [2]. Accurate spectral calibration results can benefit areas as diverse as remote sensing, robotics, astronomy, and energy-efficient design.

Methods for performing spectral calibration of thermal-IR cameras are outlined in [8] and [9]. Both methods utilize a blackbody surface for which temperature is known in order to determine the relationship between intensity level and thermal-IR radiance. It should be noted that actual thermal-IR radiance is dependent not only on surface temperature but also on surface emissivity and reflectance. As a result, temperature estimates become less accurate as the properties of the material in the image differ more greatly from those which are assumed.

### B. Camera Models

For most computer vision applications, the pinhole camera model (1) is assumed. In this model, the image-plane coordinates  $\mathbf{x}$  of a physical point  $\mathbf{X}$  can be found by multiplying by

the matrix  $\mathbf{C}$  representing the camera's internal parameters and the matrix  $\mathbf{P}$  representing the camera's translation and rotation relative to the world coordinate frame [10]

$$\mathbf{x} = \mathbf{CPX}$$

$$\begin{bmatrix} u \\ v \\ w \end{bmatrix} = \begin{bmatrix} f_x & 0 & c_x \\ s & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ w \end{bmatrix}. \quad (1)$$

Here,  $\mathbf{x}$  represents the homogeneous coordinates of a point in the image.  $f_x$  and  $f_y$  are the focal lengths (in pixels), and  $c_x$  and  $c_y$  are the center coordinates of the camera.  $r_{mn}$  and  $t_m$  are the coefficients of rotation and translation of the camera, and  $\mathbf{X}$  represents the real-world coordinates of the corresponding point. The skew parameter  $s$  is typically negligible in modern imaging devices and is therefore often set to equal zero.

The pinhole model assumes zero distortion in the image, which is rarely the case. Therefore, distortion models are often used to correct for the effects of lens distortion on the image. Distortion models provide a mapping between each pixel in the original image and its corresponding location in the ideal undistorted image. In the past, the conventional model outlined in [11] has been overwhelmingly used; however, the more recent "rational function" (RF) model shown in (2) has been shown to have superior accuracy for modeling severe distortion, such as for cameras with wide-angle lenses [12]. The RF model does not specialize to any particular lens geometry, and its simplicity means that it is easy to estimate using standard calibration procedures. In the RF model, six radial coefficients ( $k_n$ ) and two tangential coefficients ( $p_n$ ) are used to describe the distortion

$$\begin{aligned} u'' &= u' \frac{1 + k_1 r^2 + k_2 r^4 + k_3 r^6}{1 + k_4 r^2 + k_5 r^4 + k_6 r^6} + 2p_1 u' v' + p_2 (r^2 + 2u'^2) \\ v'' &= v' \frac{1 + k_1 r^2 + k_2 r^4 + k_3 r^6}{1 + k_4 r^2 + k_5 r^4 + k_6 r^6} + p_1 (r^2 + 2v'^2) + 2p_2 u' v'. \end{aligned} \quad (2)$$

Here,  $(u'', v'')$  represents the normalized and distorted coordinate corresponding to a normalized and undistorted coordinate  $(u', v')$ , and  $r$  is the Euclidean distance between the normalized undistorted coordinate and the image center. Equation (3) defines the relationship between the normalized coordinates and image coordinates  $(u, v)$

$$(u', v') = \left( \frac{(u - c_x)}{f_x}, \frac{(v - c_y)}{f_y} \right). \quad (3)$$

The accuracy of a lens distortion model is highly dependent on the provided calibration footage. The footage must contain a sufficient variety of scales and positions of the calibration pattern within the camera's field of view. Additionally, the calibration points within the pattern must have high acutance in order to be located accurately [13].

### C. Previous Work

A proven approach for the accurate calibration of the intrinsic and extrinsic geometric parameters of cameras involves the use

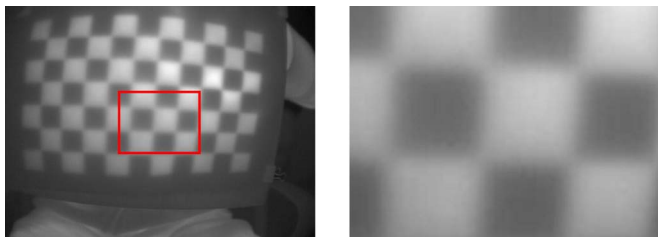


Fig. 1. (Left) Appearance of the heated chessboard in the thermal-IR modality. (Right) Magnification of the highlighted subregion.

of a planar calibration pattern or a 3-D calibration object. The known geometry of the pattern or object is exploited in order to determine how the camera views the world [14]. A sequence of images of the calibration pattern or object from different angles is typically required by most calibration algorithms [5]. The accuracy (in pixels) of the calibration model is measured by mean reprojection error (MRE) [15]. Initially, estimates of the 2-D image coordinates of calibration points are extracted from an image. The 3-D pose of the pattern is then estimated from these 2-D coordinates. The estimated 3-D locations of calibration points are then “projected” back onto the image, with the average Euclidean distance between the original 2-D estimates and the projections forming the MRE.

The most popular calibration pattern is a regular printed chessboard. This pattern has been demonstrated to be highly effective and convenient for the task of calibrating visible-spectrum cameras [5]. The calibration points (corners where the squares join) are easily located in the image with high precision due to the high image contrast. However, without additional preparation, the calibration points on a chessboard cannot be reliably located in a thermal-IR image. This is because the pattern will typically have a near-uniform temperature and thermal radiance and will therefore appear with low contrast in the image. A popular method for enabling the use of a calibration chessboard for thermal camera calibration involves heating the pattern through exposure to a flood lamp [4], [16]. This results in the appearance of the pattern in the thermal modality resembling its appearance in the visible modality. However, this approach struggles to get the crisp corners required for accurate calibration. As an example, Fig. 1 shows perhaps the best quality calibration image that could be produced using the heated chessboard method. An alternative approach which appears to achieve better results involves printing the chessboard pattern onto a specially manufactured printed circuit board [17]. Another approach involves using a suspended wire grid, heated with a heat gun [3]. The cost and required equipment and preparation associated with these more sophisticated approaches were motivating factors for the development of the proposed approach.

Another interesting approach to the calibration of thermal-IR cameras is the adaptation by Johnson and Bajcsy [18] of the method first presented in [19]. In the original method, a freely moving bright spot is used as the calibration object [19]. This bright spot can be generated by a laser pointer, a flashlight, or other alternatives. The adapted method uses a pole with an open-bulb flashlight attached to the end, which is clearly detectable to both the thermal-IR and visible-spectrum cameras

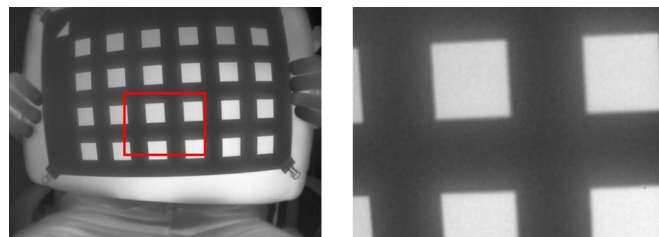


Fig. 2. (Left) Appearance of the proposed mask in the thermal-IR modality. (Right) Magnification of the highlighted subregion.

[18]. A significant limitation of this method is that it cannot be used for systems involving fewer-than-three cameras.

Many toolkits exist which can be used to perform calibration when provided with a sequence of calibration images by the user. The Caltech Matlab Calibration Toolbox [20] is a popular toolkit for end-to-end calibration. However, this toolkit has several weaknesses which make it ineffective for the problem of efficiently calibrating thermal-IR cameras. For example, the toolkit requires significant manual intervention, such as user guidance in searching for the pattern. Also, it includes the conventional distortion model [11] only and does not take advantage of a high-order distortion model such as the RF model [12], which means that it cannot model severe distortion accurately. Another weakness of this and many other toolkits (such as the example calibration code provided by OpenCV [21]) is that frames must be chosen specifically by the user based on what he thinks would lead to the best possible calibration results. This dependence on user expertise was another motivating factor for this research.

### III. GEOMETRIC MASK FOR CALIBRATION

The proposed calibration pattern consists of a grid of regularly sized squares cut out of a thin material such as a cardboard. It is important that the material is opaque in the modality of interest. The mask can be manufactured by hand using a ruler and a box cutter or, alternatively, using a laser cutter. When used, the pattern is held in front of a backdrop with a different level of thermal radiance. This difference in radiance could be due to difference in temperature or difference in thermal emissivity. Example backdrops include a warm laptop case or a powered computer monitor. When used in this way, the pattern is easily identifiable in the thermal-IR domain, as shown in Fig. 2. If the mask is differently colored to the backdrop, it will also be easily identifiable to visible cameras. Compared with Fig. 1, it can be clearly seen that the corners (which serve as the calibration points) are much clearer than those when using a heated chessboard. Evidence that this improves calibration performance is presented in Section VII.

For the experiments outlined in this paper, the mask pattern was cut out of an A4 sheet of cardboard. Squares were 20 mm wide and spaced with 20-mm separation. A grid of four squares along the shorter axis of the cardboard and six squares along the longer axis was used. These dimensions were selected to ensure that the distribution of corners over the pattern would be similar to that used in conventional chessboard-based methods. The size of the pattern is not critical but should be chosen



according to the camera field of view and the distance to the camera, so that the pattern covers a significant portion of the field of view. For higher resolution cameras (those used in this research were all  $640 \times 480$ ), the square cuts may be made smaller, but it is not clear whether this would actually improve calibration performance. A template of the pattern can be found in [6].

To use the pattern for camera calibration, a video or image sequence must be captured which contains multiple views of the pattern from different angles. This can be achieved by moving the pattern relative to the camera or cameras or vice versa. The best results are achieved with a trajectory that includes a variety of rotations and translations of the pattern, covering most of the camera's field of view. With the current implementation, the entire pattern must be clearly visible to the camera. In the authors' experience, the best results are achieved by holding the pattern as close as possible to the camera while maintaining focus of the entire pattern. For cameras with a shallow depth of field (such as many thermal-IR cameras), this may be difficult. A compromise may need to be made between having a large-enough focal distance to keep the entire pattern in focus simultaneously and holding the pattern close enough to get high-resolution corner features. Estimates of the pattern distances from each camera and the sizes of the calibration workspaces that were used in each experiment are outlined in Section VII.

#### IV. PATTERN SEARCH ALGORITHM

In order to locate the pattern in each image, a new algorithm is proposed which is available in [6]. The present version of the algorithm has not been optimized for speed, and the entire pattern must be present in the image for a successful search to occur. The algorithm consists of the following processes.

- 1) The image histogram is linearly expanded to span from values of 0 to 255, to increase contrast.
- 2) Maximally stable extremal regions (MSERs) [22] are extracted from the image.
- 3) Unfeasible MSERs are filtered out of the putative set.
- 4) A clustering scheme is used to find all and only the MSERs that comprise the actual pattern.
- 5) A local-homography-guided search is used to estimate and then locate the subpixel accurate positions of the corners of each square.

##### A. Histogram Expansion

Histogram expansion is often performed on board for many thermal-IR and visible-spectrum cameras. However, in some cases, this process must be performed after the images are captured. Histogram expansion is needed in order to improve the contrast of the pattern, so that conventional computer vision algorithms used as part of the process of automatically locating the calibration pattern can be effective. Equation (4) was used to improve the image contrast for a standard 8-b grayscale image

$$I'(x, y) = 255 \frac{I(x, y) - \min(I)}{\max(I) - \min(I)}. \quad (4)$$

Here,  $I(x, y)$  represents the original intensity value at coordinate  $(x, y)$  in the image, and  $I'(x, y)$  represents the intensity at the same coordinate in the histogram-expanded image.  $\max(I)$  and  $\min(I)$  are the maximum and minimum intensities present in the original image.

##### B. MSER Extraction

Initially, the OpenCV [21] MSER [22] detector implementation is used to locate candidate features in the image that may represent squares on the calibration pattern. Parameters for the detector can be very permissive as later steps work to remove incorrect features. The specific parameters used in the proposed implementation can be found in [6], with explanations for each parameter found in [21].

##### C. MSER Filtering

A filter cascade is then applied to the candidate features to reduce the list to a smaller subset. The filters applied reject MSERs based on the following properties (listed in order of filter occurrence):

- 1) abnormally shaped;
- 2) too variant in their internal gray levels;
- 3) enclosing other MSERs.

The shape filter measures the height and width of each feature and rejects those which have a ratio that is too large. In the proposed implementation, if the height and width differ by a factor of more than two, then the feature is rejected. This is because such a ratio would represent either a nonsquare feature or a square feature at too great an angle to the camera for accurate corner extraction.

The gray-level variance filter calculates the variance of the pixel intensities in each feature. If a feature has a variance of greater than 256, then it is rejected. This is because a variance this high suggests that the feature is not of uniform-enough intensity to represent a single square in the calibration pattern.

The enclosure filter simply rejects any feature that fully encloses another feature.

Thresholds for these filters were manually set at very conservative levels, in the sense that they are highly unlikely to reject valid MSERs but may allow some invalid MSERs to be retained. Any invalid MSERs will generally be rejected later in the pattern-finding algorithm. The system does not need to be tuned by the user in order for it to be effective.

##### D. MSER Clustering

A clustering algorithm is then implemented which connects all features which are similar in size and average intensity to their neighbors. A maximum size variation of 20% and an intensity variation of 16 are allowed. These thresholds were found to be flexible enough for cameras with large amounts of lens distortion, while still rejecting most invalid features. At this point, only a cluster which is equal to or larger than the number of squares in the actual pattern is retained. If the remaining cluster has more squares than are known to be in the pattern, graph trimming is performed in order to remove any

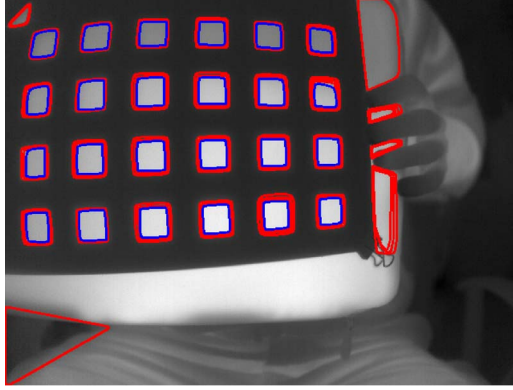


Fig. 3. All putative MSERs extracted by the OpenCV MSER detector implementation. The blue regions represent those that passed the filtering, clustering, and graph-trimming stages.

wrongly attached MSERs. This is a one-by-one removal of the MSERs which have the greatest variation in distances to their nearest three neighbors. The reasoning for this decision is that the MSERs within the pattern will have a small and relatively consistent distance to at least their three nearest neighbors, whereas incorrect MSERs located outside the actual pattern are unlikely to. The effectiveness of the filtering, clustering, and graph-trimming stages of the algorithm is shown in Fig. 3.

At this point, the pattern is assumed to be valid to enable an ordering of the MSERs from top left to bottom right. The validity of this assumption is then tested by ensuring the following.

- 1) Features in the image are surrounded by features that physically surround them in the actual pattern. This works on the assumption that, if the MSER ordering was unable to abide by the basic physical arrangement of the actual pattern, then the extracted pattern is invalid.
- 2) All sets of three adjacent features which represent physically collinear regions are approximately collinear in the image. The threshold chosen limited the variation of the position of the central feature from the line connecting the two outer features to one-third of the length of the connecting line.
- 3) Distances between neighboring MSER centroids vary smoothly. For a single feature, the distance to its nearest neighbor must be at least half of the distance to its furthest neighbor.

#### E. Homography-Guided Search

In order to accurately locate the calibration points under potentially severe lens distortion, a local-homography-based approach is used. Homography mapping can be seen as a linear planar transformation. A homography matrix  $\mathbf{H}$  is used to relate a physical plane to the image coordinates, as shown in

$$\mathbf{x} = \mathbf{H}\mathbf{X}, \quad \begin{bmatrix} u \\ v \\ w \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix}. \quad (5)$$

Here,  $[u \ v \ w]$  represents the coordinates of a point in the image plane, and  $[x \ y \ z]$  represents the coordinates of the point

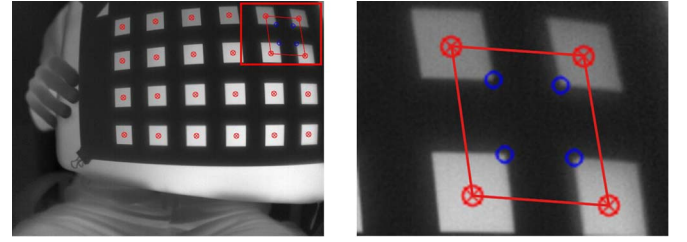


Fig. 4. (Left) All initial square centroids (marked in red). (Right) Magnification of the highlighted subregion. The estimated positions of four internal corners are marked in blue. The red lines connect the four centroids used to guide the search for this subset of internal corner positions.

in the physical plane (in this case, the physical calibration grid). The image-plane coordinate vector may need to be normalized by dividing by  $w$  in order to interpret the coordinates in a meaningful way.

The proposed algorithm initially uses a homography mapping from groups of four noncollinear square centroids to estimate the positions of all of the internal calibration points. The OpenCV function “findHomography()” [21] is used to calculate this homography. This algorithm uses a least squares scheme to compute an initial homography estimate. The Levenberg–Marquardt method [23] is then used to optimize the homography by further reducing the reprojection error. Fig. 4 shows an example of the four centroids selected to estimate the positions of four internal calibration points. This works under the assumption that, in the distorted image, small subregions of the pattern are approximately planar.

Each time an estimate of a calibration point is made, the OpenCV function “cornerSubPix()” [21] is used to refine the position of the calibration point. This function iterates to find the radial saddle point which represents the subpixel accurate location of the corner. The search window for this function is constrained to have a radius equal to half of the estimated distance to the nearest calibration point. This is to prevent the subpixel location from being incorrectly assigned to another nearby corner in the pattern.

Once all internal calibration points have been accurately located, a centrally propagating tactic is employed to locate the remaining calibration points along the edges of the pattern. The propagation involves using the locations of known calibration points to accurately estimate and locate the position of the remaining points from the inside out. Again, a homography mapping is used—this time involving the locations of the physically nearest four known and noncollinear calibration points. Using the internal calibration points as a starting point, the central propagation successively estimates and locates the remaining calibration points in the following order:

- 1) all central edge points—those on the outer rim of the pattern but at least two points in from the corner.
- 2) eight outer edge points—those on the outer rim of the pattern at a distance of one point from the corner.
- 3) four corner points—those at the extreme corners of the pattern.

This approach to locating the calibration points was chosen because it utilizes the lower levels of distortion typically present in the center of the pattern to aid in the search for corners in

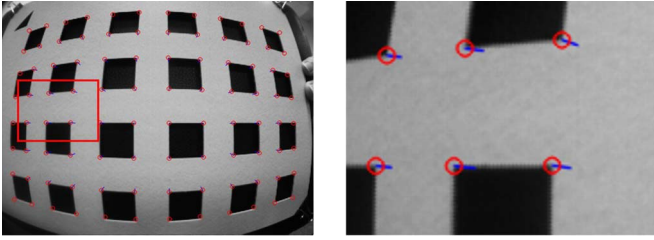


Fig. 5. (Left) Correction of the estimated positions of calibration points using “cornerSubPix()”. (Right) Magnification of the highlighted subregion. The red circle marks the corrected position, while the blue line represents the displacement from the original estimate. These images have been converted to grayscale from a color camera.

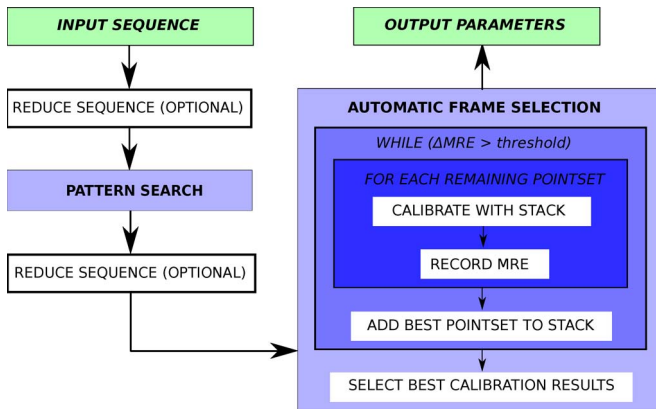


Fig. 6. Structure of the proposed calibration system implementation.

regions with more severe distortion. Using the proposed approach, even the outermost corners which may be present in the most distorted regions of the image can be located accurately. Fig. 5 shows the effectiveness of the approach in correcting the positions of initial local-homography-based corner estimates. These images are taken prior to the final location correction of the calibration points on the edge of the pattern.

## V. SYSTEM IMPLEMENTATION

The proposed implementation was designed for the purpose of thermal and visible-spectrum camera calibration for both the single- and multiple-camera cases.

The proposed end-to-end system accepts one or more image or video sequences and returns near-optimal calibration results without any intervention by the user. The approach utilizes an implementation of the enhanced MCM approach for automatic frame selection [7]. All code associated with the system has been made public in [6] under the General Public License (GPL) [24]. The system is written in the C++ language and makes use of the OpenCV libraries [21]. It is intended that this can form the basis for further development of the tool for the benefit of the research community. It should also be noted that any pattern detection approach can be used within this framework.

The structure of the implementation is shown in Fig. 6.

In the interests of efficiency, the user can optionally specify a maximum number of frames (or sets of frames for extrinsic calibration) from the original sequence to preserve for pattern

searching. The calibration pattern is then searched for in all of these frames, and those in which the pattern and its corners are accurately located are preserved as a point set. This collection of point sets can then be further reduced if required to improve the speed of the Automatic Frame Selection module.

The Automatic Frame Selection module implements the enhanced MCM approach [7]. This approach involves testing multiple combinations of point sets for calibration effectiveness. In the proposed implementation, initially, just a single point set is used to calibrate the camera(s). The accuracy of the calibration results is evaluated in terms of their ability to generalize to the entire sequence of point sets. The point set which, on its own, achieves the lowest MRE is then added to the stack. An optimum second point set is then searched for, and so on.

## VI. EXPERIMENTAL SETUPS

For the experiments, five separate cameras of three different models were used:

- 1) two Thermoteknix Miricle 307K thermal-IR cameras;
- 2) two Videre “Apparen” STH-MDCS2-VAR/-C color cameras;
- 3) one GUIDIR IR210 thermal-IR camera.

The Miricle 307K cameras consist of a long-wave uncooled microbolometer detector sensitive in the 7- to 14- $\mu\text{m}$  range. The cameras have a resolution of  $640 \times 480$  pixels. They are tested to see objects in the temperature range of  $-20\text{ }^{\circ}\text{C}$ – $150\text{ }^{\circ}\text{C}$ . The cameras have a noise-equivalent differential temperature (NEDT) of 85 mK. Fourteen-bit monochromatic images can be streamed from the cameras at 50 Hz using a Universal Serial Bus video class connection. An estimate of the focal length of the camera based on the limited provided data is 598 pixels.

The Videre Apparen color cameras contain a CMOS imager and the options of  $640 \times 480$  or  $1280 \times 960$  pixel resolutions. These cameras use a Bayer color filter array to encode color content—which makes calibration less accurate than a grayscale camera of equal resolution. Images can be streamed from the cameras using an IEEE 1394 (FireWire) interface. A frame rate of 30 Hz can be achieved with the lower resolution mode, which is preferred in this paper. There are insufficient data provided by the manufacturer to estimate the focal length for the camera.

The IR210 camera consists of a Uncooled Focal Plane Array microbolometer detector sensitive in the 8- to 14- $\mu\text{m}$  range. The resolution of the camera is  $384 \times 288$  pixels. The NEDT of this camera is 80 mK at a nominal temperature of  $30\text{ }^{\circ}\text{C}$ . Unfortunately, images must be reduced to an 8-b representation before being retrieved by a PC using an Radio Corporation of America/Bayonet Neill–Concelman interface, at a frame rate of 50 Hz. There are insufficient data provided by the manufacturer to estimate the focal length for this camera.

Two multicamera setups were used in the experiments. The first consisted of a centrally mounted thermal-IR camera (Miricle 307K) surrounded by two color cameras (Videre). All three cameras are mounted in-line in a three-camera variable-baseline stereo configuration. For the experiments, the baseline is fixed at 100-mm separation between each camera. An





Fig. 7. (Left) Multicamera setup 1. The two outer cameras are color cameras, while the central camera is a thermal-IR camera. (Right) Multicamera setup 2. Both cameras operate in thermal IR.

image of the multicamera multimodality setup is shown in Fig. 7. An example application for this multimodality setup would be in tasks where shape and depth were estimated using the two visible cameras and thermal-IR temperature data were mapped onto a generated 3-D model as visual texture.

The second multicamera setup consisted of two thermal-IR cameras (Miracle 307K) mounted in a variable-baseline stereo configuration, also shown in Fig. 7. It is anticipated that this rig will be useful for applications where depth mapping or 3-D reconstruction in thermal IR is needed, such as in problems of nighttime navigation. The separation of the cameras is fixed at 150 mm for the experiments.

In each of the multicamera setups, data were streamed simultaneously from all cameras in Linux operating system. The two open-source programs utilized for data capture were yavta [25] and Dynamic Data eXchange (DDX) [26]. All synchronization was managed by DDX—an open-source software architecture which allows programs to share data at run time through a shared memory mechanism [26]. By using DDX, within a few seconds of all cameras streaming, the images are synchronized within a consistency of approximately 6  $\mu$ s.

All computation was done on a MacBook Pro with a 2.40-GHz processor and 2 GB of RAM.

## VII. EXPERIMENTS

The proposed calibration framework is evaluated and compared to existing approaches for calibrating thermal cameras and multimodality camera setups. Four experiments (Sections VII-A–VII-D) are conducted.

Section VII-A includes a qualitative evaluation of the proposed mask versus the heated chessboard in terms of their convenience for thermal camera calibration. Section VII-B quantitatively compares the accuracy of calibration using the two patterns. Section VII-C evaluates the performance of the proposed pattern-finding algorithm. Section VII-D demonstrates the effectiveness of the implementation and the mask-based approach in fully calibrating a multiple-camera multiple-modality setup.

For Sections VII-B and D, the extended MRE is used as the metric for comparison. In general, MRE is calculated using only the frames included in the calibration process. We chose to calculate MRE using an entire input sequence predominately containing frames that are not actually selected for calibration. This helps to ensure that the metric indicates the generality of

TABLE I  
EVALUATION OF ALTERNATIVE APPROACHES

	Mask	Chessboard
<i>Manufacturing</i>	Less than one hour work. Requires card, cutting.	Printable from a standard printer.
<i>Heating</i>	Mask or the backdrop must be heated or cooled. A variety of convenient methods can be used.	A powerful (500W+) flood lamp and an external power supply is required. Difficult to get even coverage.
<i>Footage</i>	Pattern can be used easily and effectively for several minutes.	The pattern is only effective for a few seconds after heating.
<i>Searching</i>	Algorithms are not very common. However, one is provided in the implementation source code [14].	Generally requires pre-processing (inversion/thresholding). Many conventional algorithms will struggle to find the pattern automatically.
<i>Accuracy</i>	see Experiment VII.B	

the model. Equation (6) shows the equation for extended MRE. The units of MRE are pixels

$$\text{MRE} = \frac{\sum_{m=1}^M \sum_{n=1}^N \|p(m, n) - q(m, n)\|}{MN}. \quad (6)$$

Here,  $M$  is the total number of frames in the extended sequence, and  $N$  is the total number of calibration points per frame.  $p(m, n)$  is the pixel location of a point on the pattern, and  $q(m, n)$  is the reprojected location of that point using an estimate of the pose of the pattern in that particular frame.

Experiments were conducted in a normal office environment, with an ambient temperature of approximately 24 °C. The calibration objects were placed at the minimum focal distance for which the pattern was fully visible, which, for all cameras, was approximately 40 cm from the lens. A workspace of approximately 1 m<sup>3</sup> was required for all experiments.

### A. Evaluation of Alternative Approaches

For this section, the proposed mask-based approach and the traditional heated chessboard are assessed in terms of their key limitations, as shown in Table I.

Effort was made to achieve the best possible performance from the conventional chessboard approach, with much difficulty. A chessboard was printed on an A4 sheet of paper and clipped flat to a rigid folder. A 500-W heat lamp was then used for approximately 5 s to heat the pattern as even as possible. Footage was then captured immediately for approximately 10 s, at which point the image contrast had degenerated significantly. An identical software framework was used for both the proposed method and the conventional method, with the exception of the pattern-finding algorithm. For the chessboard, the OpenCV function “findChessboardCorners()” [21] was used. It should be noted that both pattern-finding algorithms involve the same technique to adjust the corner locations to subpixel accuracy.

Fig. 8 shows the difference in degradation in image quality over time using the two approaches. It is clear from the figure that the window of opportunity using the conventional heated

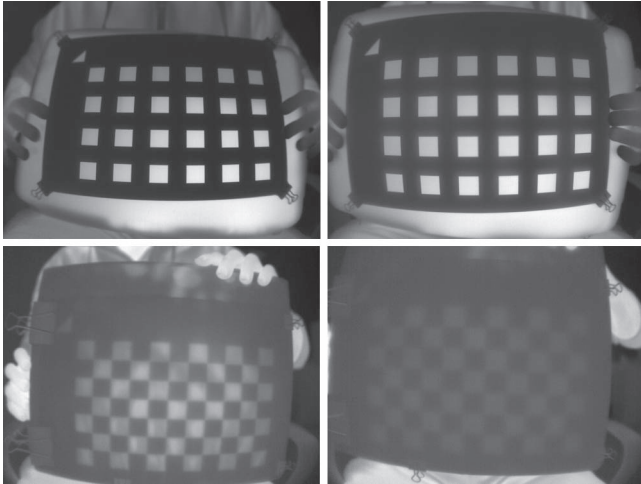


Fig. 8. (Top row) Proposed mask. (Bottom row) Conventional heated chessboard. (Left column) Initial image immediately after heating. (Right column) Image taken 30 s after heating.

TABLE II  
PATTERN EFFECTIVENESS (MRE AND STANDARD DEVIATION)

Configuration	MRE
<i>GUIDE IR210 (mask)</i>	<b>0.324±0.011 px</b>
<i>GUIDE IR210 (board)</i>	0.804±0.015 px
<i>Miricle 307K (mask)</i>	<b>0.284±0.006 px</b>
<i>Miricle 307K (board)</i>	1.274±0.020 px

chessboard is much smaller than that with the mask-based approach. Together with the other limitations of the existing method, this makes the proposed approach appealing for many, if not all, cases.

### B. Comparison of Pattern Effectiveness

The system implementation described in Section IV was used on sets of 200 calibration frames from video sequences of each pattern (both the proposed mask and the conventional heated chessboard) to compare performance. It should be noted that several attempts were required to capture a sufficient number of frames using the heated chessboard, whereas capturing the mask sequence required only one attempt. There was a significant challenge in evenly and effectively heating the chessboard pattern and then quickly capturing footage before the image quality degraded. A limit of ten frames was set for the optimal-frame selection module of the implementation. In the authors' experience, improvements in accuracy are generally negligible using more than ten frames.

Ten trials were performed on each of the two thermal-IR cameras for each pattern, and the average MRE scores were recorded. Each trial used a randomly selected set of 200 frames from a much larger sequence. The results are shown in Table II. From this table, it is clear that calibration points on the mask can be more accurately located than those on the heated chessboard. The improvement is particularly marked for the Miricle 307K camera, which has more extreme lens distortion than the GUIDIR IR210.

Table III shows that the mask-based approach is also more consistent in its estimation of the cameras' focal lengths. While

TABLE III  
PATTERN EFFECTIVENESS (FOCAL LENGTHS AND S.D.)

Configuration	$f_x$	$f_y$
<i>GUIDE IR210 (mask)</i>	<b>638.85±1.35 px</b>	<b>655.24±1.33 px</b>
<i>GUIDE IR210 (board)</i>	644.22±3.00 px	660.01±2.89 px
<i>Miricle 307K (mask)</i>	<b>604.66±1.73 px</b>	<b>606.27±1.65 px</b>
<i>Miricle 307K (board)</i>	615.55±14.71 px	615.73±14.84 px

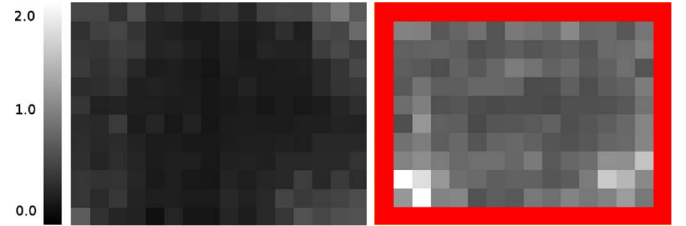


Fig. 9. (Left) Error distribution for the mask. (Right) Error distribution for the chessboard. The scale ranges from MRE of (black) 0.0 to (white) 2.0 or higher. The red shades represent regions which contained no calibration points, and therefore, the error is unknown.

no reliable ground truth was available for any of the cameras, for the Miricle 307K, an estimate was able to be made from provided camera specifications. The estimate of 598.0 pixels compares favorably to the performance of the proposed approach.

To analyze the relationship between reprojection error and image location, an error map was generated from a typical set of results using both patterns. This is shown in Fig. 9. From the images, it can be seen that the central regions tend to have lower errors than the edges. This is likely because of at least two factors.

- 1) The image quality degrades toward the edges because of the curvature of the lens.
- 2) The scaling effect of the distortion multiplying the error.

The mask managed to achieve a more even reprojection error distribution over the camera's field of view than the heated chessboard. Given that, even in the central region, the reprojection error from the mask is significantly lower, this results in a superior error distribution over the conventional approach.

The evaluation shows that the proposed mask is a more effective calibration pattern for thermal-IR cameras than a heated chessboard. However, several failure conditions may cause the mask to be ineffective for the task of calibration. These include an insufficient radiance difference between the mask and the backdrop, the use of a too-thick material (a maximum of 1 mm in thickness is recommended), and the lack of rigidity of the mask.

### C. Performance of Pattern-Finding Algorithm

In order to evaluate the effectiveness of the proposed pattern-finding algorithm, a calibration sequence of 1000 frames from the Miricle 307K thermal camera was captured. For these tests, the null hypothesis was that the pattern was present in the image. A "true positive" (TP) was defined as a correct acceptance of the null hypothesis, while a "true negative" (TN) was defined as a correct rejection. "False positive" (FP) and



TABLE IV  
MASK PATTERN-FINDER SUCCESS RATE

	Positive	Negative
<i>True</i>	0.958	1.000
<i>False</i>	0.000	0.042

TABLE V  
CHESSBOARD PATTERN-FINDER SUCCESS RATE

	Positive	Negative
<i>True</i>	0.225	1.000
<i>False</i>	0.000	0.775

TABLE VI  
PATTERN-FINDER CPU TIME COMPARISON (IN MILLISECONDS)

	Mask		Board	
	True	False	True	False
<i>GUIDE IR210</i>	98	61	44	185
<i>Miricle 307K</i>	186	141	165	473
<i>Videre (visible)</i>	202	181	47	635

“false negative” (FN) correspond to an incorrect acceptance or rejection of the null hypothesis, respectively. Therefore, the conditions in (7) must be satisfied in the tests

$$TP + FN = 1, \quad FP + TN = 1. \quad (7)$$

Table IV shows the results for the proposed algorithm.

From the table, it can be seen that the proposed pattern-finding algorithm is highly effective in locating the mask pattern. It is important that the number of FPs is kept at zero while obtaining the highest possible count of TPs. The failure conditions of the algorithm do not tend to occur in normal controlled calibration footage. One major failure condition is when the pattern is very distant from the camera, such that the corners of the squares on the pattern are no longer clearly visible. In the experiments of this paper, this was a distance of approximately 2.0 m. Using a camera with a larger focal length will result in this maximum distance being larger. Another condition of failure is when the pattern is held at a significant angle (more than 45°) relative to the camera plane. These conditions are generally undesirable for calibration footage regardless of the effectiveness of the detector. However, as mentioned in Section IV, a significant failure condition of greater importance is when part of the board is occluded from view.

In contrast, the results of the OpenCV [21] pattern-finding algorithm in locating the chessboard from a similar calibration sequence are shown in Table V.

The OpenCV function “findChessboardCorners()” [21] can be said to perform very poorly in automatically locating the heated chessboard in a typical sequence of calibration images. However, it should be noted that the algorithm was probably not designed to be used for a task as difficult as thermal-IR camera calibration.

Table VI shows the CPU times in milliseconds of the two pattern-finding algorithms for each of the three types of cameras used in the experiments, when the null hypothesis was both true (the pattern was present) and false (the pattern was not present).

TABLE VII  
THREE-CAMERA SETUP-EXTRINSIC TRANSLATIONS

Method	$T_x$	$T_y$	$T_z$
<i>Mask</i>	98.33±0.34mm	5.60±0.67mm	3.41±0.98mm
<i>Chessboard</i>	95.77±0.36mm	2.89±0.41mm	6.18±0.97mm
<i>Measured</i>	<b>100.00±2.00mm</b>	<b>5.00±4.00mm</b>	<b>5.00±7.00mm</b>

From Table VI, it can be seen that the OpenCV function “findChessboardCorners()” [21] used to find the chessboard is considerably faster than the proposed algorithm when the pattern is present in the image. However, it should be noted that the proposed algorithm has not been optimized, and therefore, significant speed improvements are likely to be achievable. Nevertheless, the difference in processing time is unlikely to have a significant effect on usability for typical calibration tasks. Interestingly, the proposed algorithm is much faster than the OpenCV algorithm when the pattern is not present in the image. This is likely because of the iterative nature of the OpenCV function in continuously reprocessing the image and researching for the pattern until it gives up.

#### D. Multicamera and Multimodality Experiments

For the final part of the performance evaluation, the two multicamera setups outlined in Section VI and shown in Fig. 7 were used. For both the heated chessboard and the proposed mask, sequences of 200 calibration frames were used, and a limit of up to ten frames was set for actual calibration. Ten trials were run for each setup, with each using a randomly selected sequence of 200 frames from a much larger set. The MREs, translations, and rotations were recorded. The same set of camera intrinsic parameters was used for both calibration patterns and setups. The MREs associated with estimating the intrinsic parameters for each camera were 0.349 (Miricle 307K), 0.484 (Videre camera 1), and 0.622 (Videre camera 2). It should be noted that the accuracy of extrinsic calibration is dependent on the accuracy of the intrinsic results.

For the three-camera multimodality setup, the proposed system using the mask approach achieved an average MRE of 1.910 pixels, compared to an MRE of 2.497 pixels using the heated chessboard. For the thermal-stereo setup, the proposed system achieved an MRE of 1.109 pixels versus 2.746 pixels for the conventional approach. It is apparent that the advantage of the proposed mask-based approach is more significant for the case with the two thermal-IR cameras. This difference is because the accuracy improvement of the mask over the heated chessboard does not apply to the visible modality.

Table VII shows the means and standard deviations in the estimated translations between two of the cameras (one thermal and one visible) in setup 1, compared to those estimated using physical measurements. It should be noted that it is very difficult to measure the exact positions of the centers of the cameras, as these are internal and therefore difficult to access. Attempts to measure displacements manually were done with a T-square and a ruler. The  $Z$ -axis is defined in the direction where the leftmost camera is facing, with the  $Y$ -axis being vertical and the  $X$ -axis being horizontal.

Table VIII similarly shows the results for setup 2.

TABLE VIII  
THERMAL-STEREO SETUP-EXTRINSIC TRANSLATIONS

Method	$T_x$	$T_y$	$T_z$
<i>Mask</i>	148.13±0.17mm	3.37±0.24mm	3.46±0.88mm
<i>Chessboard</i>	151.20±2.88mm	2.93±1.61mm	2.98±1.55mm
<i>Measured</i>	150.00±2.00mm	0.00±4.00mm	0.00±4.00mm

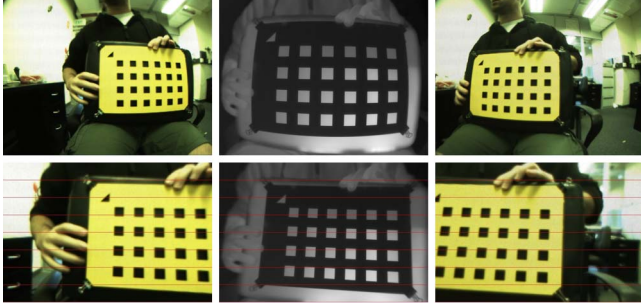


Fig. 10. Demonstration of the intrinsic and extrinsic calibration of a multiple-camera multiple-modality setup, using the proposed approach. (Top row) Original camera views. (Bottom row) Undistorted and rectified camera views. (Left column) Visible camera 1. (Center column) Thermal-IR camera. (Right column) Visible camera 2.

All measurements and estimates of rotations were approximately zero, as expected.

From the tables, it is evident that the translations estimated using the proposed approach are well within the bounds of practical physical measurements. However, the true camera center locations are not known exactly, and so, the results from the automated approach are trusted more by the authors than the physical measurements. However, the results from the heated chessboard approach are also mostly within the bounds of measurements. This suggests that accurate extrinsic calibration may not depend on a high level of intrinsic calibration accuracy. Further research may be needed to more precisely understand the effect of errors in intrinsic calibration on the accuracy of extrinsic calibration.

The proposed implementation is shown to be effective for the task of accurate geometric calibration of a multiple-camera setup involving a thermal-IR camera and two color cameras. Example results from the calibration of the three-camera setup are shown in Fig. 10.

This clearly demonstrates the effectiveness of the mask-based approach for the problem of multiple-modality extrinsic camera calibration. Failure conditions include when there are an insufficient number of frames containing full views of the pattern (less than ten for each camera is not recommended) or when the estimates of the camera intrinsic parameters are very poor.

For the experiments in this paper, intrinsic calibration typically required about 5 min of processing time per camera, with a similar amount of time required for extrinsic calibration. Higher resolution cameras or setups involving larger numbers of cameras would require longer periods of time for calibration.

## VIII. CONCLUSION

A geometric mask has been presented as an alternative pattern for the calibration of thermal-IR cameras. The pattern is easily manufactured and requires none of the extra equipment

existing methods do, such as flood lamps or other light sources. Results showed that it can obtain an improvement in MRE of up to 78% compared to using the conventional approach of a heated chessboard.

An MSER-clustering algorithm has been proposed as a means for locating the calibration mask in images. Calibration points are then located to subpixel accuracy using an inside-to-outside local-homography-based approach. The algorithm was shown to be highly effective in locating the pattern over an extended sequence. Future development for this algorithm could look at using a local line curvature to help find corners in extremely distorted images. Alternatively, a single-parameter lens distortion model could be used to assist the homography computation. Additionally, the case-by-case rejection of individual calibration points based on a local-actance test could help improve accuracy by reducing the negative effect of motion and out-of-focus blur [13].

An end-to-end implementation for intrinsic and extrinsic calibration of cameras was developed, which is shared online in [6] as an open-source project under the GPL [24]. The proposed calibration system does not require any manual intervention beyond providing the initial calibration sequence and pattern dimensions. The system was effectively used to solve the difficult problem of intrinsically and extrinsically calibrating three cameras, including a thermal camera and two color cameras. One potential avenue for future research is the use of region-weighted projection error to ensure that a fitted model is not biased by overfitting to the region of the image most occupied by calibration points and weak fitting to other regions.

Future work planned by the authors includes performing wide-baseline matching on multiple-camera multiple-modality networks, using the approaches in [27] and [28].

## ACKNOWLEDGMENT

Equipment, as well as manufacturing and computing facilities, was provided by both the Commonwealth Scientific and Industrial Research Organisation (CSIRO) and Queensland University of Technology. The authors would like to thank the technical staff at CSIRO for their assistance in the assembly of the capture platform and the affiliated Information Security Institute for the cooperation.

## REFERENCES

- [1] K. Beier and H. Gemperlein, "Simulation of infrared detection range at fog conditions for enhanced vision systems in civil aviation," *Aerosp. Sci. Technol.*, vol. 1, no. 1, pp. 63–71, Jan. 2004.
- [2] P. Bajcsy and R. Kooper, "Integration of data across disparate sensing systems over both time and space to design smart environments," in *Sustainable Radio Frequency Identification Solutions*. Rijeka, Croatia: INTECH, 2010, pp. 281–306.
- [3] Y. M. Ng and R. Du, "Acquisition of 3D surface temperature distribution of a car body," in *Proc. IEEE Int. Conf. Inform. Acquisition*, 2005, pp. 16–20.
- [4] S. Y. Cheng, S. Park, and M. M. Trivedi, "Multiperspective thermal IR and video arrays for 3D body tracking and driver activity analysis," in *Proc. IEEE CVPR*, 2005, pp. 2–3.
- [5] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 11, pp. 1330–1334, Nov. 2000.
- [6] *mm-Calibrator—OpenCV-Based Multi-Modal Camera Calibration System—Google Project Hosting*. [Online]. Available: <http://code.google.com/p/mm-calibrator/>

- [7] S. Rupp, M. Elter, M. Breitung, W. Zink, and C. Küblbeck, "Robust camera calibration using discrete optimization," *Int. J. Appl. Sci., Eng. Technol.*, vol. 1, no. 3, pp. 137–141, 2005.
- [8] S. Rainieri and G. Pagliarini, "Data processing technique applied to the calibration of a high performance FPA infrared camera," *Infrared Phys. Technol.*, vol. 43, no. 6, pp. 345–351, 2002.
- [9] S. Rainieri, F. Bozzoli, and G. Pagliarini, "Characterization of an uncooled infrared thermographic system suitable for the solution of the 2-D inverse heat conduction problem," *Exp. Thermal Fluid Sci.*, vol. 32, no. 8, pp. 1492–1498, Sep. 2008.
- [10] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*. Cambridge, U.K.: Cambridge Univ. Press, 2003.
- [11] D. Brown, "Decentering distortion of lenses," *Photogramm. Eng.*, vol. XXXVII, no. 3, pp. 444–462, 1966.
- [12] D. Claus and A. W. Fitzgibbon, "A rational function lens distortion model for general cameras," in *Proc. IEEE CVPR*, Jun. 2005, vol. 1, pp. 213–219.
- [13] J. N. Ouellet and P. Hebert, "Developing assistant tools for geometric camera calibration: Assessing the quality of input images," in *Proc. 17th ICPR*, Aug. 2004, vol. 4, pp. 80–83.
- [14] D. Liebowitz and A. Zisserman, "Metric rectification for perspective images of planes," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Santa Barbara, CA, 1998, pp. 482–488.
- [15] D. Rzeszutarski and B. Wiecek, "Calibration for 3D reconstruction of thermal images," in *Proc. 9th Int. Conf. QIRT*, 2008, pp. 563–566.
- [16] S. Prakash, P. Y. Lee, T. Caelli, and T. Raupach, "Robust thermal camera calibration and 3D mapping of object surface temperatures," in *Thermosense XXVIII*, J. J. Miles, G. R. Peacock, and K. M. Knettel, Eds. Bellingham, WA: SPIE, 2006, p. 62 050J.
- [17] V. Hilsenstein, "Surface reconstruction of water waves using thermographic stereo imaging," in *Proc. IVCNZ*, 2005, pp. 102–107.
- [18] M. J. Johnson and P. Bajcsy, "Integration of thermal and visible imagery for robust foreground detection in tele-immersive spaces," in *Proc. 11th Int. Conf. Inform. Fusion*, 2008, pp. 1–8.
- [19] T. S. Svoboda, D. Martinec, and T. S. Pajdla, "A convenient multi-camera self-calibration for virtual environments," *Presence: Teleoperators Virtual Environ.*, vol. 14, no. 4, pp. 407–422, Aug. 2005.
- [20] *Camera Calibration Toolbox for Matlab*. [Online]. Available: [http://www.vision.caltech.edu/bouguetj/calib\\_doc/](http://www.vision.caltech.edu/bouguetj/calib_doc/)
- [21] *Welcome—OpenCV Wiki*. [Online]. Available: <http://opencv.willowgarage.com/wiki/>
- [22] J. Matas, J. Matas, O. Chum, M. Urbana, and T. Pajdla, "Robust wide-baseline stereo from maximally stable extremal regions," *Image Vis. Comput.*, vol. 22, no. 10, pp. 761–767, Sep. 2004.
- [23] M. I. Lourakis, "A brief description of the Levenberg–Marquardt algorithm implemented by Levmar," *Matrix*, vol. 3, p. 2, 2005.
- [24] *The GNU General Public License v3.0—GNU Project—Free Software Foundation (FSF)*. [Online]. Available: <http://www.gnu.org/licenses/gpl.html>
- [25] *fastr/yavta—GitHub*. [Online]. Available: <https://github.com/fastr/yavta>
- [26] P. Corke, P. Sikka, J. Roberts, and E. Duff, "DDX: A distributed software architecture for robotic systems," presented at the Australian Conf. Robotics Automation, Canberra, Australia, 2004.
- [27] R. Lakemond, "Multiple camera management using wide baseline matching," Ph.D. dissertation, QUT, Queensland, Australia, Jan. 2010.
- [28] R. Lakemond, C. Fookes, and S. Sridharan, "Dense correspondence extraction in difficult uncalibrated scenarios," in *Proc. Digit. Image Comput.: Techn. Appl.*, Melbourne, Australia, 2009, pp. 53–60.



**Stephen Vidas** (M'07) received the B.Eng. degree (with first-class honors) in electrical engineering as part of the Dean's Scholars Program from Queensland University of Technology, Brisbane, Australia, in 2009, where he is currently working toward the Ph.D. degree in the Speech, Audio, Image and Video Technologies Group, Information Security Institute, working jointly with the Autonomous Systems Laboratory, Information and Communication Technologies Centre, Commonwealth Scientific and Industrial Research Organisation, Queensland Centre for Advanced Technologies, Pullenvale, Australia. His thesis is entitled "Automatic 3D Reconstruction Beyond the Visible Spectrum."

His current research interests include camera calibration, structure from motion, and multiview stereo.

Mr. Vidas was the recipient of the Information Security Institute Best Poster Prize in 2010.



**Ruan Lakemond** (M'06) received the B.Eng. degree in electrical and computer engineering as part of the Dean's Scholars Program and the Ph.D. degree with a thesis on wide-baseline matching in difficult scenarios and the application of these techniques in surveillance from Queensland University of Technology (QUT), Brisbane, Australia, in 2006 and 2010, respectively.

He is currently a Research Associate with the Speech, Audio, Image and Video Technologies Group, Information Security Institute, QUT. His current research focuses on development of wide-baseline matching techniques, camera calibration and management, 3-D reconstruction, and related applications, such as 3-D face recognition.



**Simon Denman** (M'04) received the B.Eng. degree in electrical engineering, the B.I.T. degree, and the Ph.D. degree in the area of object tracking for intelligent surveillance from Queensland University of Technology (QUT), Brisbane, Australia, in 2003 and 2009.

He is currently a Research Fellow with the Speech, Audio, Image and Video Technologies Group, Information Security Institute, QUT. His current research interests include object tracking and soft biometrics, crowd surveillance, event detection, and using intelligent surveillance to extract operational measures.



**Clinton Fookes** (M'06) received the B.Eng. degree in aerospace/avionics and the Ph.D. degree in the field of computer vision from Queensland University of Technology (QUT), Brisbane, Australia, in 1998 and 2004, respectively.

He is currently a Senior Lecturer with the Speech, Audio, Image and Video Technologies Group, Information Security Institute, QUT. He is the Technical Director for the Airports of the Future collaborative research project. His current research interests include biometrics, intelligent surveillance, security,

and complex systems.

Dr. Fookes was the recipient of an Australian Museum Eureka Prize, among other awards.



**Sridha Sridharan** (SM'99) received the B.Sc. degree in electrical engineering and the M.Sc. degree in communication engineering from the University of Manchester Institute of Science and Technology, Manchester, U.K., and the Ph.D. degree in the area of signal processing from The University of New South Wales, Sydney, Australia.

He is currently a Full Professor with the School of Engineering Systems, Queensland University of Technology, Brisbane, Australia, where he is also the Deputy Director of the Information Security Institute

and the Leader of the Research Program in the Speech, Audio, Image and Video Technologies Group.



**Tim Wark** (SM'93) received the Ph.D. degree in multimodal signal processing and pattern recognition from Queensland University of Technology, Brisbane, Australia, in 2000.

He is currently a Principal Research Scientist with the Autonomous Systems Laboratory, Information and Communication Technologies Centre, Commonwealth Scientific and Industrial Research Organisation, Queensland Centre for Advanced Technologies, Pullenvale, Australia. His current research interests lie within in-network processing in wireless sensor

networks and pervasive computing.