Object-based strategy for generating high-resolution four-dimensional thermal surface models of buildings based on integration of visible and thermal unmanned aerial vehicle imagery

Alaleh Fallah [iD],[a] Farhad Samadzadegan,[a] and Farzaneh Dadrass Javan [iD][b],*

[a]University of Tehran, College of Engineering, School of Surveying and Geospatial Engineering, Tehran, Iran

[b]University of Twente, Department of Geo-Information Science and Earth Observation (ITC), Enschede, The Netherlands

ABSTRACT. The mitigation of energy usage in urban areas, especially in buildings, has recently captured the attention of many city managers. Owing to the thermal images' limited resolution, especially at the edges, creating a high-resolution (HR) surface model from them is a challenging process. This research proposes a two-phase strategy to generate an HR four-dimensional thermal surface model of building roofs. In the single-source modification phase, an enhanced thermal orthophoto is produced by retraining the enhanced deep residual super-resolution deep network, and then, using state-of-the-art structures from motion, semi-global matching, and space intersection. The final surface model's resolution is raised by combining thermal data with visible unmanned aerial vehicle images to overcome the limitation of single-source methods in resolution increase. To this end, after visible orthophoto and digital surface model generation, buildings and their boundaries are extracted using the multi-feature semantic segmentation method. Next, in the multi-source modification phase, a fine-registered enhanced thermal orthophoto is generated, and thermal edges are identified around the boundary of the building. The visible and thermal boundaries are then matched, and any smoothness in the temperature edges is eliminated. The results show that the average difference in position between the thermal edges and building boundaries is reduced, and temperature smoothness is completely eliminated at the building edges.

1 Introduction

In the last few years, the mitigation of energy usage in urban areas, particularly in buildings as the main urban objects, has attracted the attention of many city managers.[1,2] Accurate surface temperature data and their spatial distribution aid in identifying heat losses, air and moisture leakages, cracks, insufficient insulation in roofs, and so on.[3,4] Heat leakage from the roof of a building includes about 25% of the total heat loss from a building.[5] Therefore, the amount of energy a building uses is significantly influenced by its roof, and their thermal inspection can help increase life and reduce building maintenance costs.[6]

*Address all correspondence to Farzaneh Dadrass Javan, f.dadrassjavan@utwente.nl

Thermography allows recording, analyzing, and interpreting thermal abnormalities caused by localized building damage or faults.[7,8] The precise location of thermal defects cannot be analyzed using two-dimensional (2D) thermographic images. In this respect, a thermal surface model that displays three-dimensional (3D) information of buildings and thermal information is required to detect, interpret, and measure abnormalities in building roof investigations.[9]

The spatial resolution of such a thermal surface model, especially at the edges, is one of its limitations.[10] In actuality, the spatial resolution of thermal images is typically quite coarse due to the requirement for greater instantaneous field of view in thermal cameras (to ensure that enough energy reaches the detector). Consequently, the thermal surface model created only from original thermal images has a low spatial resolution and few details, which could make it difficult to detect, interpret, and measure thermal anomalies. Therefore, it becomes important to provide a process for creating a high-quality thermal surface model including thermal information and high spatial resolution. In this research, this model is called an high-resolution (HR) four-dimensional (4D) thermal surface model.

1.1 Related Works

Existing methods for producing an HR 4D thermal surface model are classified as single-source and multi-source. In the first group, only thermal images are utilized to increase the quality of the output. Meanwhile, in the second group, fusing thermal information with data from other sensors, such as visible images, light detection and ranging (LiDAR), laser scanners, etc., is used to improve the resolution of thermal surface models.[11–16]

In single-source methods, one solution to modify the quality of the thermal surface model is to enhance the resolution of thermal images using "hardware" methods, which have higher expenses and restrictions.[17] In recent years, researchers have employed super-resolution (SR) methods, which are single-source.[18–21] These methods just use images from a single source to produce HR images and then high-quality surface models. Studies show that increasing the scale factor of SR methods to create spatially enhanced images can result in artificial structures. Therefore, higher scale factors yield lower quality outputs than lower scale factors, which is unacceptable.[22]

In recent years, the development of unmanned aerial vehicles (UAVs) has enabled cost-effective imaging in large numbers, making them an ideal tool for capturing visible images. Therefore, among multi-source methods, the integration of surface models generated from visible images obtained by UAVs with thermal information has attracted considerable research attention.

Nevertheless, the registration of thermal and visible data is a challenge in the production of HR thermal surface models using multi-source methods. To overcome this challenge, some researchers used joint camera systems to capture visible and thermal images concurrently.[23,24] For many projects, joint systems are typically too expensive and uneconomical. Determining the thermal camera's internal orientation arguments requires accurate calibration between two cameras, which is a challenge. Furthermore, separate flights in these systems are impossible with a thermal and visible camera from the same area.

Other studies capture thermal and visible images separately. To develop a model that integrates the geometric correctness and HR of visible images with the thermal data obtained from thermal infrared (TIR) images, Ref. 7 present a method based on the iterative closest point algorithm. Sledz et al. in Ref. 25 projected TIR images onto the digital surface model (DSM) created from visible images to create a much higher geometrically accurate orthophoto. Reference 26 suggested integrating visible and thermal point clouds to produce an HR thermal point cloud from building rooftops. The final point cloud generated using their method has thermal information and a high spatial resolution. Reference 27 proposed a method for combining visual and TIR data obtained from UAVs to create a thermal surface model of an active volcano. In the study by Paziewska,[28] to merge thermal and visible data, point features from the thermal imaging point cloud were interpolated onto the vertices of the visible model.

According to what is stated, single-source methods do not require data from different sources to produce HR 4D thermal surface models. However, they have high costs and construction limitations in the hardware group, and in SR group, the limitation of scale increment exists. On the other hand, the main problem of multi-source methods is accurate registration of data from several sources with varying resolutions.

(a)

(b)

Fig. 1 (a) Visible camera. (b) Thermal camera used in this study.

In addition to the limitations of the methodologies adopted by the cited research, none of the reviewed studies addressed the problem of temperature smoothness at the edges of objects due to the low resolution (LR) of thermal images. Thus, overcoming the thermal images' limitations, which include LR, especially at the edges, is the main challenge in this research for producing an HR 4D thermal surface model. Herein, a two-phase strategy is proposed to create an HR 4D thermal surface model of building roofs using UAV aerial images to overcome the problems of thermal images. In the first phase, an enhanced thermal surface model and an enhanced thermal orthophoto with improved resolution are created based on the single-source method. The visible data are used to integrate with thermal data to increase the resolution of the final surface model in the second phase. To this end, buildings and their boundaries are extracted using a multi-feature semantic segmentation method. Then, using the proposed method, the visible and thermal boundaries are matched, and the smoothness of the thermal edges is eliminated.

The generated HR 4D thermal surface model can be used to visualize the building roof's thermal state and detect thermal anomalies to optimize energy consumption.

2 Materials and Methods

2.1 Sensors

In this research, two datasets including thermal video and visible images are used. Thermal videos are recorded using a Keii HL-640S uncooled focal plane array camera. This camera detects the middle and longwave IR spectrum, which is the TIR region of the IR spectrum. Additionally, visible images are captured using an HR Sony a6000 24 MP camera. The visible and thermal cameras employed in this research, are shown in Fig. 1. Table 1 provides more detailed information about sensors.

2.2 Platforms

Flight is performed using a light weight, multi-rotor UAV with a roll and pitch axis stabilizer. The UAV platform employed in this research is shown in Fig. 2. This UAV has eight motors, its flight altitude is about 400 m, and its maximum flight time is about 35 min.

Table 1 The technical characteristics of the used sensors.

Parameter	Visible camera	Thermal camera
Sensor	CMOS	Uncooled FPA
Focal length (mm)	35	25
Image size (pixels)	6000*4000	640*480
Pixel size (μm)	4.04	17

Fig. 2 UAV platform employed in this research.

(a)   (b)
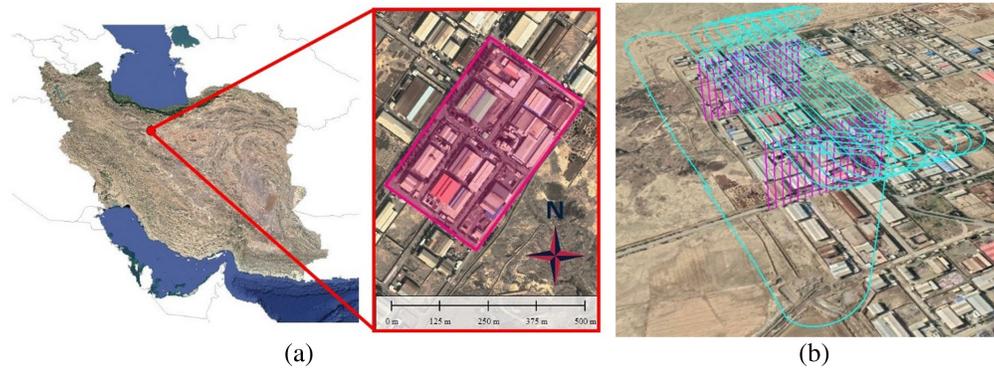
Fig. 3 (a) The study area. (b) The flight plan.

Table 2 Flight plan details.

Parameter	Value
Duration of flight	20 min
Speed of flight	60 km/h
Height of flight	300 m
Overlap	70%
Sidelap	70%
Number of strips	17

2.3 Study Area and Flight Plan

In the southern part of Tehran, Iran, a region serves as the study's location. A UAV ground control station software is used to schedule flights over the approximately 123,000 m² region. The study area and the flight plan are depicted in Fig. 3. Some details about the flight plan are listed in Table 2.

2.4 Ground Control Points

In this research, 33 natural features in the area, such as the corners of buildings, etc., which can be seen in both sets of visible and thermal images, were used as control points. The distribution of these control points in the study area and examples of them in the visible and thermal images are shown in Fig. 4.

Fig. 4 Control points distribution and samples of their zoomed view in visible and thermal images.
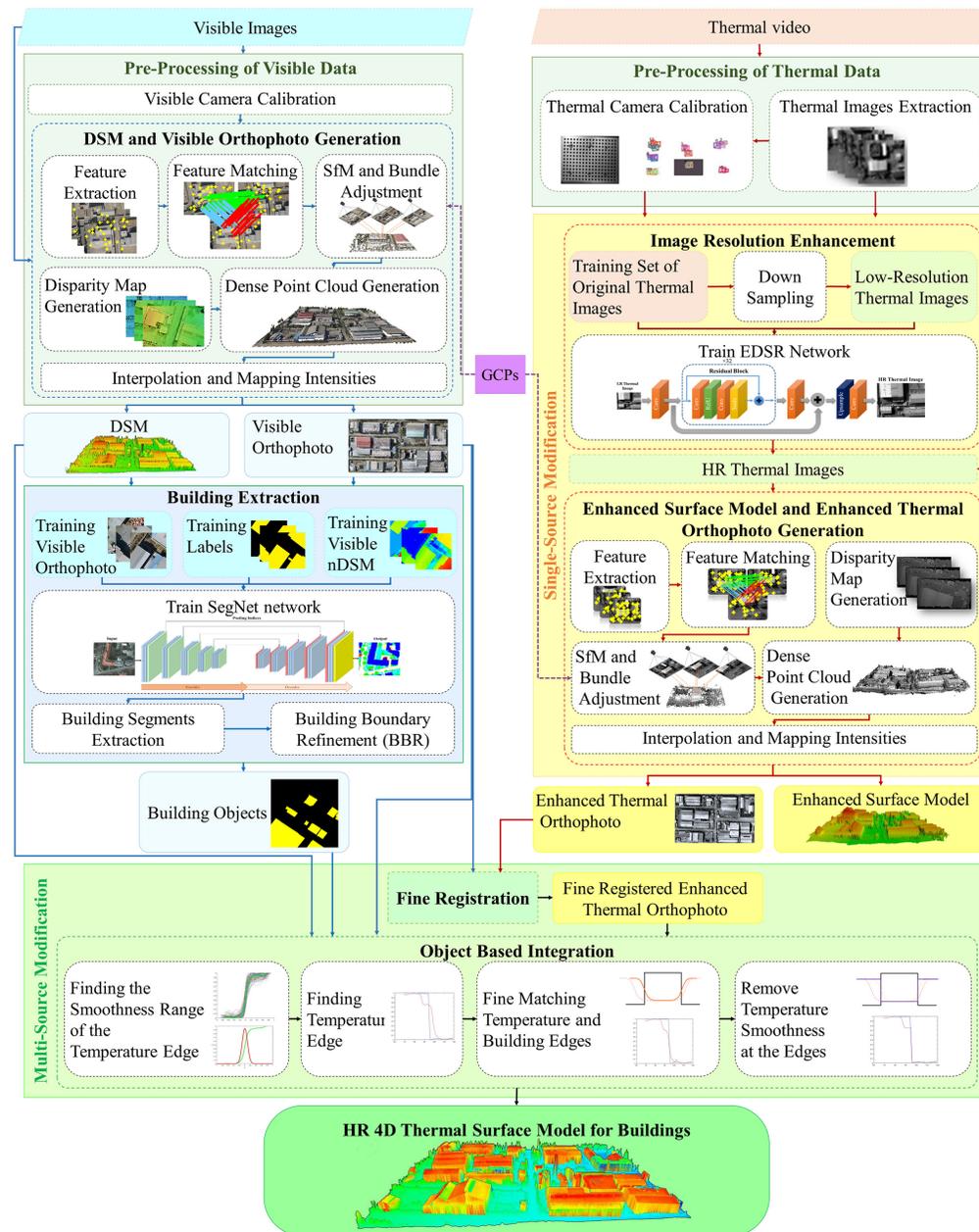
Fig. 5 The proposed method to produce an HR 4D thermal surface model for buildings.

2.5 Methodology

Given the problems of UAV thermal images and the pros and cons of single-source and multi-source methods, a strategy that combines these methods in two phases is proposed to generate an HR 4D thermal surface. Figure 5 displays the proposed method's flowchart. The subsequent sections provide specifics on each step.

2.5.1 *Pre-processing of thermal data*

Both the thermal and visible data need pre-processing to enter the main process of the proposed method. In the case of thermal data, first, the captured video is converted into a string of thermal images. Then, the camera calibration process is performed.

In this study, the traditional camera calibration method is adopted for thermal camera calibration. Before using the photogrammetric methodology, these techniques are utilized to calculate the camera arguments based on image data, such as points or lines with accurate

coordinates.[29] A proper test field must be designed to achieve this. To better detect targets and increase contrast in thermal images, the designed test field for thermal camera calibration should be heated first. Then, images are taken using Zhang's method from various directions and orientations.[30] The camera calibration arguments are then determined using the projection between the features' location in the image and their object coordinates using the following equations (collinearity equations):[31]

$$x = -c\frac{r_{11}(X - X_0) + r_{21}(Y - Y_0) + r_{31}(Z - Z_0)}{r_{13}(X - X_0) + r_{23}(Y - Y_0) + r_{33}(Z - Z_0)}, \tag{1}$$

$$y = -c\frac{r_{12}(X - X_0) + r_{22}(Y - Y_0) + r_{32}(Z - Z_0)}{r_{13}(X - X_0) + r_{23}(Y - Y_0) + r_{33}(Z - Z_0)}. \tag{2}$$

In the equations above, x and y are image coordinates; X_0, Y_0, and Z_0 are coordinates of the projection center; c is focal length; X, Y, and Z are object coordinates; and r represents components of the rotation matrix. Using Brown model equations, lens distortion parameters are determined [Eqs. (3) and (4)][32]

$$x' = x(1 + k_1 r^2 + k_2 r^4 + k_3 r^6 + p_2(r^2 + 2x^2 + 2p_1 xy)), \tag{3}$$

$$y' = y(1 + k_1 r^2 + k_2 r^4 + k_3 r^6 + p_1(r^2 + 2y^2 + 2p_2 xy)), \tag{4}$$

where x' and y' depict the undistorted image coordinates, r is the Euclidean distance between the image coordinates and coordinates of the projection center, p_i is tangential distortion coefficient, and k_i is radial distortion parameter. The accuracy of the calibration technique in geometric calibration can be calculated based on the mean re-projection error.

2.5.2 Single-source modification

Surface models generated from thermal images have lower accuracy because of the lower resolution of these images. On the other hand, a better registration outcome can be obtained by matching and bringing the data resolutions closer to one another.[33] For this purpose, in this phase, an enhanced surface model and an enhanced thermal orthophoto are produced using only thermal data.

There are two main processes in this step. A deep learning (DL)-based single-image SR (SISR) model is trained in the first step to create HR thermal images from LR ones. In the next step, the enhanced surface model and enhanced orthophoto are produced using the outcomes of the earlier steps. The next sections provide details on these two steps.

Image resolution enhancement. A convolutional neural network, named enhanced deep residual super-resolution (EDSR) network, is utilized to apply SISR to improve the thermal image resolution.[34] The EDSR network was selected because of its ease of execution and satisfactory performance, according to recent studies.[35]

The objective in EDSR network training, is to train model f, which is represented in Eq. (5). In this equation, $\widehat{Th}_{HR}$ shows predicted HR thermal image, and Th_{LR} is LR thermal image. In other words, by reducing the distance between $f(Th_{LR})$ and Th_{HR} (HR thermal image), the EDSR network creates a resolution-enhanced image $\widehat{Th}_{HR}$

$$\widehat{Th}_{HR} = f(Th_{LR}). \tag{5}$$

According to Ref. 34 compared to $L2$, the mean absolute error loss function ($L1$ loss) produces better convergence. Therefore, the $L1$ is used to train the EDSR network instead of the $L2$. Equation (6) gives the $L1$ loss function that needs to be minimized

$$e = \frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}\|Th_{HR}(i,j) - f(Th_{LR}(i,j))\|. \tag{6}$$

In Eq. (6), m stands for the number of image rows, n for the number of image columns, i for the index of each row, and j for the index of the column.

After training, a map is created to predict HR images from LR input images. This technique can be applied to super-resolve any LR image to generate an HR image. In the next steps, the enhanced surface model and enhanced orthophoto are created using these enhanced thermal images.

Enhanced surface model and enhanced thermal orthophoto generation. To produce an enhanced surface model from a collection of HR thermal images, first, the exterior orientation arguments of images are computed using the state-of-the-art structures from motion (SfM) method. The SfM algorithm uses the corresponding points found by the scale-invariant feature transform algorithm and also ground control points (GCPs) in a sequential bundle adjustment to establish the input images' exterior orientation parameters. Then a 3D point cloud is generated.[36]

Second, a disparity map is created by applying the semi-global matching (SGM) method to HR thermal images.[37] From a pair of rectified stereo images, SGM calculates a dense disparity map. Numerous researchers have used the SGM because of its satisfactory results in dense stereo-matching applications.[37] A dense point cloud is then produced using space intersection, the disparity maps created from each stereo pair of images, and the exterior orientation arguments. Data gridding is performed to create the enhanced surface model after generating a dense point cloud. After that, intensity values or digital number values, from the corresponding images are assigned to the enhanced surface model,[38] and the enhanced thermal orthophoto is produced.

2.5.3 *Pre-processing of visible data*

Visible images are an input for the second phase; thus, at this stage, this type of data is preprocessed to create appropriate inputs for integration with thermal data. After determining camera calibration parameters, DSM and visible orthophoto are generated from the visible images based on the section "enhanced surface model and enhanced thermal orthophoto generation," except that its inputs are visible images.

2.5.4 *Building extraction*

This research focuses on accurately assigning the temperature data from thermal orthophotos to buildings. Therefore, in this section, buildings are extracted using DSM and visible orthophoto created in previous steps.

DL plays an important role in automatic extraction. It can classify objects accurately and learn complex features. In addition, semantic segmentation methods have been applied in remote sensing tasks like object extraction and detection.[39,40] SegNet is an end-to-end network for semantic pixel-wise segmentation,[41] which is selected in this research due to its better performance compared to other deep architectures, such as FCN, FCN (learn deconv), DeepLab-LargeFOV, DeepLab-LargeFOV-denseCRF, and DeconvNet.[42,43] The structure of SegNet network is depicted in Fig. 6.
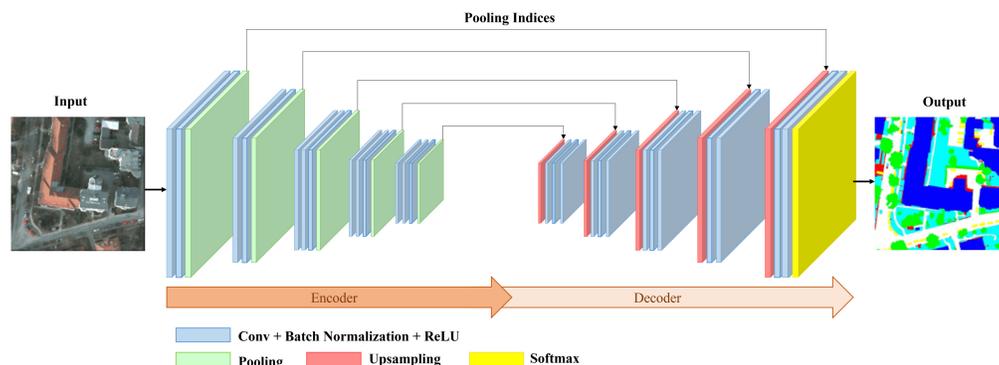

Fig. 6 SegNet architecture for semantic segmentation.

An encoder network, a corresponding decoder network, and a pixel-wise classification layer constitute the fundamental trainable segmentation architecture of the semantic segmentation model SegNet. At the encoder, convolutions using 13 convolutional layers from visual geometry group-16 (VGG-16)[44] and max pooling are performed. The decoder in the reverse step consists of 13 deconvolution layers. Up-sampling and convolutions are carried out at the decoder. The classifier approach is the last layer in the network and uses the soft-max classifier method to forecast the maximum probability of the number of classes.

Good results are obtained when buildings are extracted from a UAV photogrammetric image dataset using a SegNet-based DL semantic segmentation technique in a supervised learning model.[45] Previous research indicates that combining visible images with additional feature bands can increase building segmentation performance.[46–49] Therefore, in this research, the SegNet network and a combination of visible images and normalized DSM (nDSM) feature bands are used to extract the buildings. The nDSM is generated by removing the digital terrain model (DTM) from the DSM using the method proposed by Ref. 50. The data dimension is increased by the nDSM to several views. Consequently, the nDSM feature helps distinguish between the ground and the rooftop and between the shape of buildings and trees.[47]

The visible orthophoto, nDSM, and the image of training labels will be introduced to the SegNet network for training. After the SegNet network has been trained, the generated visible orthophoto and nDSM enter the trained model, and the buildings are extracted.

Although the SegNet network outperformed many other deep architectures, it is important to note that deep architecture networks, including SegNet, generally yield predictions with poor boundaries.[51] Therefore, in this research, the building boundary refinement (BBR) method presented by Ref. 52 is adopted to enhance the building boundaries. In this method, boundary pixels are detected for each building object. Subsequently, in a repeatable procedure, a specific rule is utilized for neighborhood pixels to choose whether each boundary pixel region should grow or shrink. At the end of this step, the accurate building objects are extracted.

2.5.5 *Multi-source modification*

After generating enhanced thermal orthophoto, DSM, and visible orthophoto from thermal and visible sources, the second phase of the proposed method starts. In this phase, first, the enhanced thermal orthophoto is fine registered to the visible orthophoto, and then, an object-based integration operation is performed.

Fine registration.   Since GCPs are employed in the production of surface models, thermal surface model, DSM, and their related orthophotos are registered. This step involves fine registration of enhanced thermal orthophoto and visible orthophoto using the B-Spline registration algorithm.[53] In this method, the input data are transformed under the control of a grid of B-spline control points. An error measurement is employed to determine the degree of misregistration between slave and master images. To achieve the best possible registration between the two sets of data with the fewest possible registration errors, the control points are moved using the quasi-Newton optimizer.

Object based integration.   In this step, the temperature determined from the fine-registered intensified thermal orthophoto should be assigned to extracted buildings. Due to the lower resolution of thermal images, the temperature mapped from fine-registered enhanced thermal orthophoto to DSM will be smooth at the edges. In addition, there is still some deficiency in the accurate registration of temperature edges and the edges of the extracted buildings. Figure 7(a) depicts a hypothetical building in that its temperature is lower than the surroundings. If a cross-section is considered on this building, the temperature profile assigned to it will be like Fig. 7(b). In this figure, the black profile shows a cross-section of a building in the DSM, and the red profile depicts the temperature assigned to the cross-section. The deficiency in the registration and edge smoothness is clear in the temperature profile. The following steps are taken to accurately register the edges and eliminate temperature smoothness at the edges of the building.
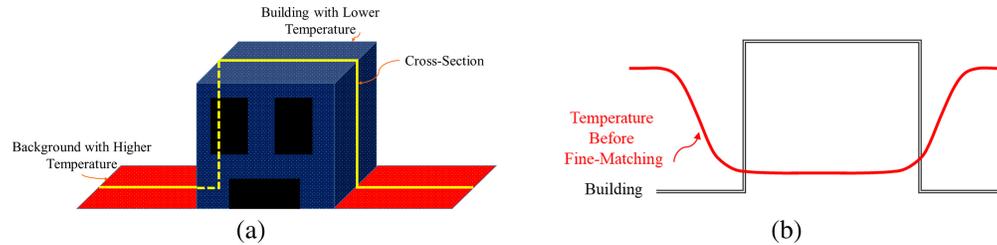
Fig. 7 (a) A hypothetic building. (b) The deficiency in the registration and edge smoothness for a cross-section of a hypothetic building.

- Finding the smoothness range of the temperature edge

Images are regarded in the imaging process to be the outcome of an imaging function $F(.)$ applied to objects, which can be presented as the following equation:

$$\text{Image}(x, y) = F(\text{Object}(x, y)). \tag{7}$$

The imaging function $F(.)$ is determined by taking into account the linear system assumption as a set of 2D convolutions of objects with the point spread functions (PSFs). The PSFs are made up of elements of the imaging system like $\text{PSF}_{\text{Atmosphere}}$, PSF_{Lens}, and $\text{PSF}_{\text{Sensor}}$[54]

$$\text{Image}(x, y) = \text{PSF}_{\text{Atmosphere}} * \text{PSF}_{\text{Lens}} * \text{PSF}_{\text{Sensor}} * \text{Object}(x, y) = \text{PSF}_{\text{System}} * \text{Object}(x, y). \tag{8}$$

In optical analysis, a system's line spread function (LSF), which depicts the picture of an ideal line, is typically favored over the PSF because it is simpler to measure. The PSF and LSF can be converted into one another. High-contrast edges are good objectives for assessing the spatial response. Therefore, the object is often chosen so that the reflection is strong on one side of the edge and weak on the other.

Edge profiles are produced at each edge point once edge locations are extracted. After smoothing and consistency checks, the edge spread function (ESF) is incorporated into the LSF calculation. In this research, the mean value of the ESFs computed for different edges is calculated to obtain an ESF for the entire image. Then, differentiation is used on the mean ESF profile to determine LSF. Next, a Gaussian function smooths and removes the noise from the resulting LSF curve. The LSF calculation equation is expressed in the following equation:

$$\text{LSF}(x) = \frac{d}{dx}[\text{ESF}(x)]. \tag{9}$$

Figure 8 displays the visual relation between LSF and ESF. As shown in the figure, the smoothness area of the edge is located within the range of blue dashed lines.

The difference between the two independent variable values at which the dependent variable is equal to half of its maximum value is called the full width at half maximum (FWHM).[55] If Eq. (10) is the density of a normal distribution, where σ is the standard deviation and μ is the expected value the maximum value of the function of Eq. (10) is obtained from Eq. (11)
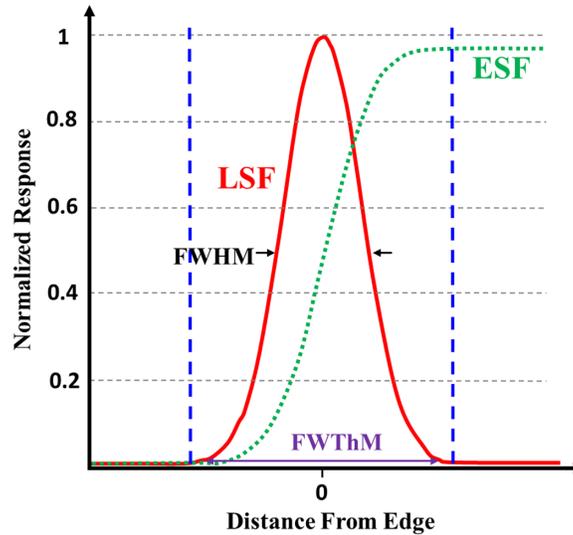
$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \cdot \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right], \tag{10}$$

$$f_{\max} = f_X(\text{mode}(X)) = f_x(\mu) = \frac{1}{\sqrt{2\pi}\sigma}. \tag{11}$$

And so, on considering the definition of FWHM, its bounds satisfy the following equation:

$$f_X(x_{\text{FWHM}}) = \frac{1}{2}f_{\max} = \frac{1}{2\sqrt{2\pi}\sigma}. \tag{12}$$

Using Eq. (12), Eq. (13) can be developed and finally Eq. (15) is obtained to calculate full width at one-thousandth maximum (FWThM) or the smoothness range[56] by solving Eq. (14)

Fig. 8 The relationship between LSF and ESF.

$$f_X(x_{\text{FWThM}}) = \frac{1}{1000} f_{\max} = \frac{1}{1000\sqrt{2\pi}\sigma}, \tag{13}$$

$$\frac{1}{1000\sqrt{2\pi}\sigma} = \frac{1}{\sqrt{2\pi}\sigma} \cdot \exp\left[-\frac{1}{2}\left(\frac{x_{\text{FWThM}} - \mu}{\sigma}\right)^2\right], \tag{14}$$

$$\text{FWThM} = 2\sqrt{2 \ln 1000}\sigma. \tag{15}$$

- Finding temperature edge

At the edge of buildings, there is usually a strong temperature difference between the building and non-building areas. The following steps should be performed to find the edge pixel in fine-registered thermal orthophoto: first, a line should be fitted to each edge's pixels for each extracted building using the least square method.

If the coordinates of the building edge pixels are $(x_1, y_1), \ldots, (x_n, y_n)$, then Eq. (16) is used to determine the best line fitted to them

$$y = mx + b. \tag{16}$$

In this equation, m is the slope of best fitted line and b is the y-intercept.

In the second step, for each pixel of the edge by coordinate (x_p, y_p), the line perpendicular to the best-fitted line should be extracted by the following equation:

$$y = -\frac{1}{m}x + y_p - mx_p. \tag{17}$$

Then, for each pixel in the desired edge, the temperature profile is extracted from fine-registered enhanced thermal orthophoto in the direction perpendicular to the edge in a certain range. This certain range can be determined depending on the accuracy of the registration in the previous steps, as well as the resolution of the thermal data, by trial and error. The edge's location is the point with the greatest slope in the temperature profile.

- Fine matching temperature and building edges

After finding the temperature edge pixel, the temperature assigned to the building is modified by shifting the temperature profile in a direction perpendicular to the edge and matching its edge pixel with the building's edge. By performing these steps, the temperature assignment error due to the mismatch of edges in thermal data and visible data is minimized. In Fig. 9(a), the orange profile shows the cross-section depicted in Fig. 7(b) after fine-matching the edges. The dashed line in this figure depicts the temperature profile before fine-matching the edges.
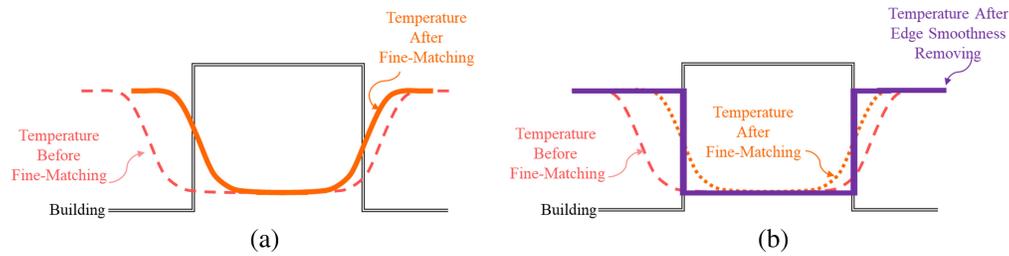
Fig. 9 (a) Remove the deficiency in the registration. (b) Remove temperature edge smoothness.

- Remove temperature smoothness at the edges

The temperature within the smoothness range is determined by taking into account the temperature of the last pixel on either side of the smoothness range. This results in the loss of temperature smoothness in the building edges. Figure 9(b) displays the temperature assigned to the cross-section of the building after removing temperature smoothness at the edges.

At the end of the implementation of the proposed method, an HR 4D thermal surface model of buildings is generated in which the building's temperature is accurate and sharp at the edges.

3 Results

In the following sections, the results of the proposed approach are assessed for the test area in the south of Tehran.

3.1 Pre-Processing Results of Thermal Data

The thermal video is recorded by the thermal camera and then converted into images with a size of 640×480 pixels. Then, thermal camera calibration is done using a rectangular calibration board that has 13×17 hollow circles. To identify the locations of targets in various images, the calibration board also includes six coded targets [Fig. 10(a)]. Imaging from multiple views is performed after heating the calibration board [Fig. 10(b)].

Adaptive thresholding is utilized to produce binary images.[57] The centers of the circles are determined in the next stage[58] [Fig. 10(c)] and are regarded as the image's coordinates. Then, the camera is calibrated after forming the object's coordinates. A total of 221 calibration points
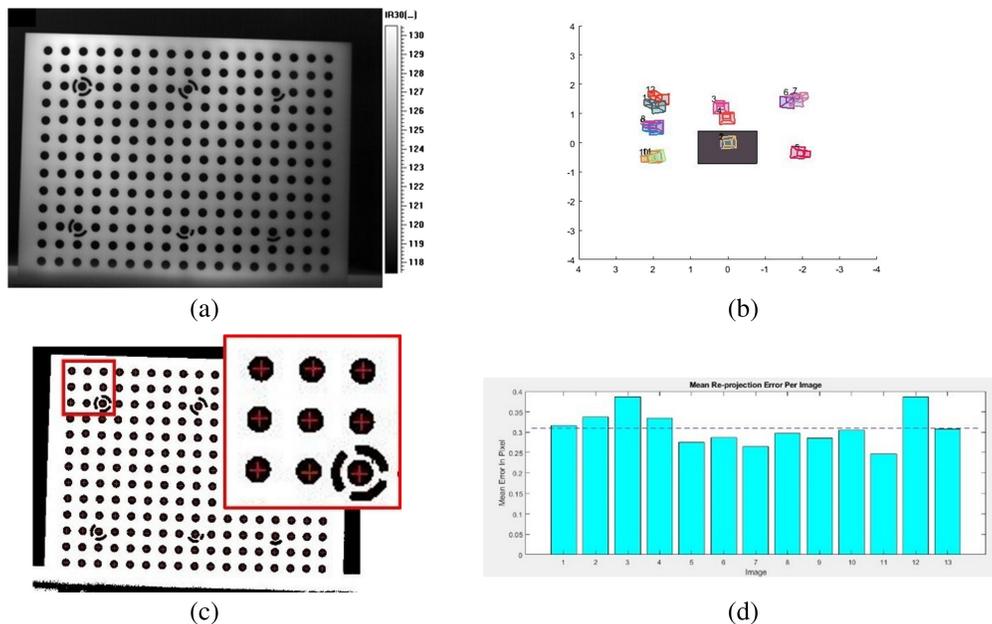
Fig. 10 (a) Calibration board, (b) camera positions, (c) finding the circles' centers, and (d) mean re-projection error chart.

extracted from 13 images are used, and the average value of mean re-projection error is estimated to be 0.315 pixels [Fig. 10(d)].

3.2 Results of Single-Source Modification

In this phase, first, the radiometric and spatial resolution of the thermal images is increased by utilizing the EDSR network. In the training step, 434 images (the thermal video's captured frames) are utilized for training the EDSR network, and 186 images are used to assess the accuracy of the calculated trained model. The structural similarity index[18] and peak signal-to-noise ratio[18] values are determined as 0.9401 and 36.72, respectively.

The trained model is employed to produce HR thermal images from the other 669 original thermal images. The scale factor for generating HR thermal images in this research is two. The size of the original images and enhanced images are 640×480 and 1280×960, respectively. In addition, the pixel size of the original images and enhanced images are 17 and 8.5 μ, respectively. Following the image resolution enhancement stage, the enhanced images are utilized to produce an enhanced surface model with an 11 cm resolution. In addition, the resolution of the original surface model is 22 cm.

Comparing the original surface model and the enhanced surface model with a DSM (created from visible images) reveals that the enhanced surface model contains greater details than the original surface model (Fig. 11). Furthermore, Fig. 11 illustrates that the edges of the objects are clearly crisper in the enhanced surface model than in the original surface model.
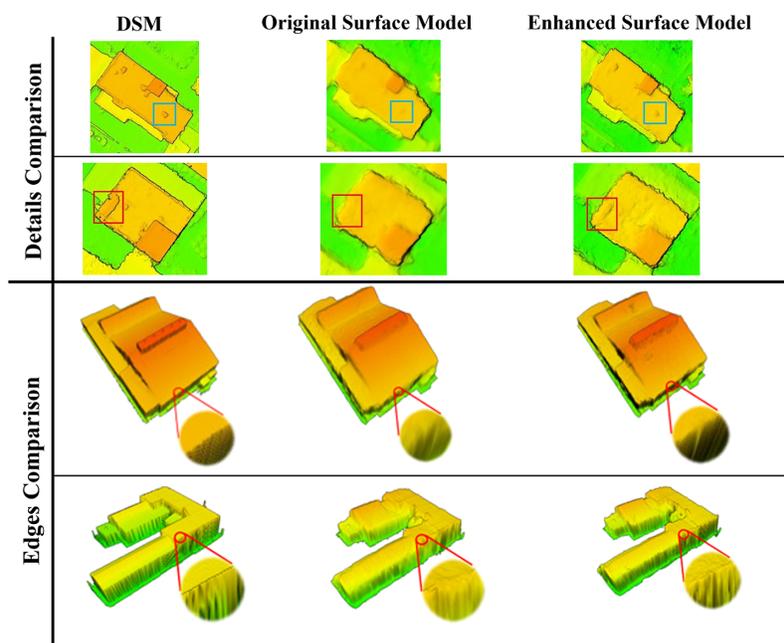
The enhanced thermal orthophoto generated from super-resolved images and enhanced surface model is depicted in Fig. 12.

## 3.3 Pre-Processing Results of Visible Data

In this step, DSM and visible orthophoto are generated from visible images to combine with thermal data. DSM is generated from visible images with a size of 6000×4000 pixels. The resolution of the DSM is 4 cm. The DSM and the visible orthophoto generated from visible images are shown in Fig. 13.

3.4 Results of Building Extraction

Visible orthophoto and DSM are generated from visible images taken in a different zone of Tehran to detect and extract buildings in the investigated area. Then, the nDSM is generated

Fig. 11 Visual comparison of the details and edges of the original surface model and the enhanced surface model to the reference DSM.
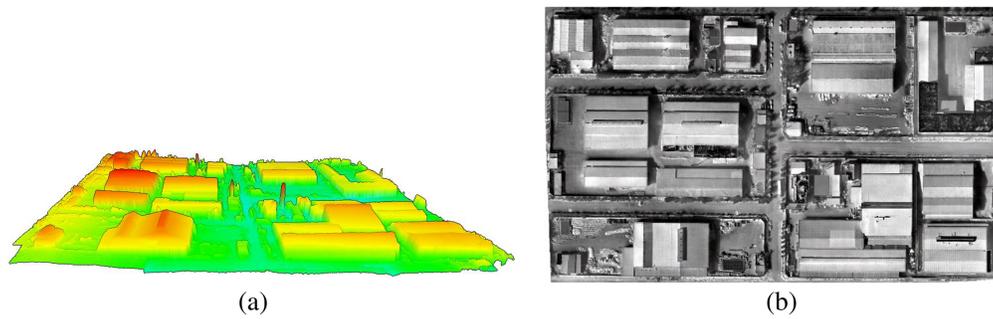
**Fig. 12** (a) Enhanced surface model. (b) Enhanced thermal orthophoto.
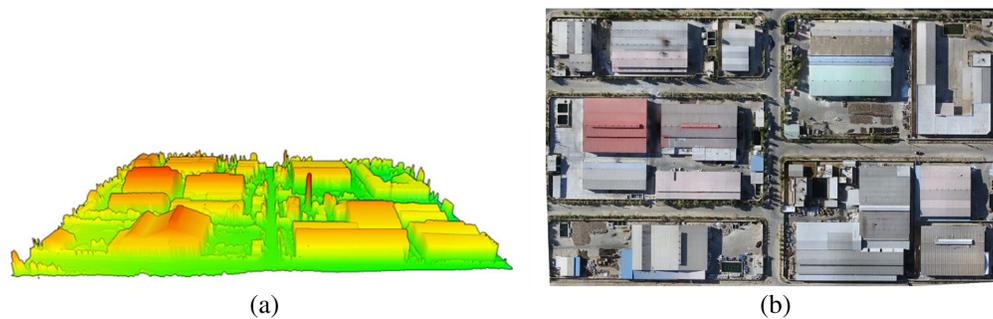
Fig. 13 (a) 3D view of DSM. (b) Visible orthophoto generated from visible UAV images.

from the DSM. The orthophoto and the nDSM are cropped into small parts with a size of 480×480 pixels. The cropped orthophotos are utilized as base maps to label the ground objects as building and non-building. Thus, the final dataset for training the SegNet network consists of 862 labeled images, cropped visible images, and cropped nDSMs, which are all connected by the same ID.

The SegNet network parameters used in the semantic segmentation learning process are the same as those used by Ref. 41. The visible orthophoto, nDSM, and labeled image samples of the training dataset are shown in Fig. 14.
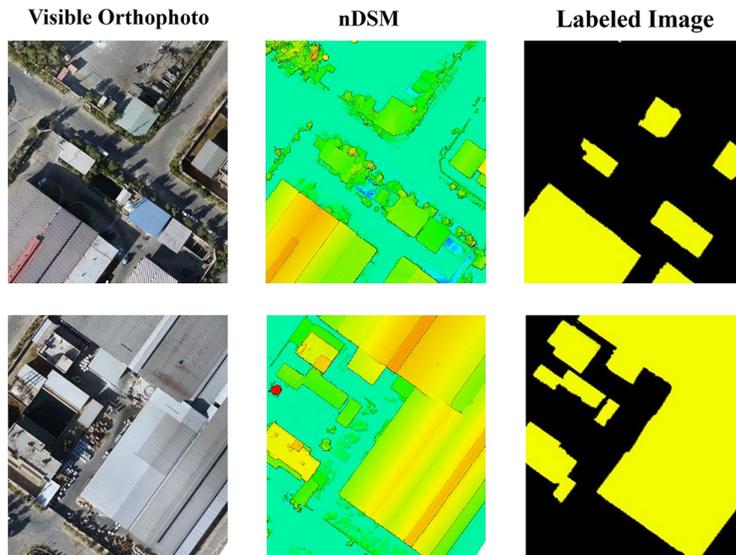
After the network's training, the visible orthophoto and the nDSM of the studied area are fed to the network, and its output is a labeled image that determines the buildings in the area.
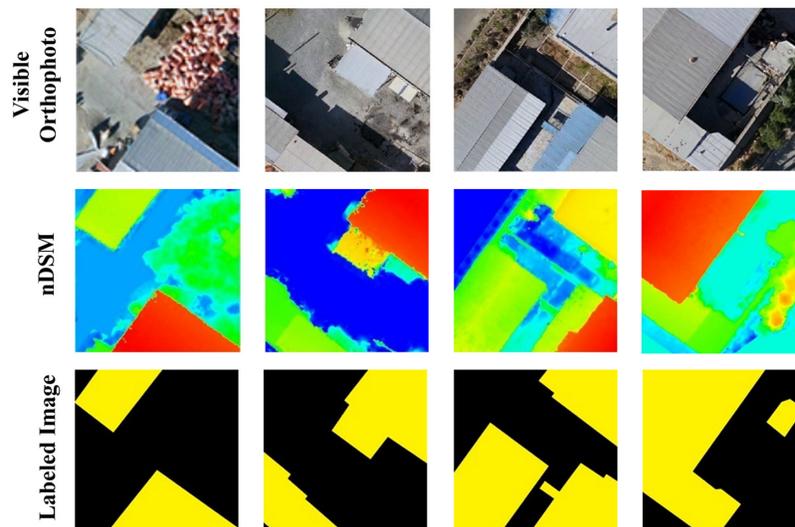
Fig. 14 Samples of the training dataset for building extraction.

Fig. 15 Overview of all the buildings extracted from the study area.

Fig. 16 Samples of original inputs and segmentation to building and no-building class results.

Figure 15 depicts an overview of all the buildings extracted from the study area. Figure 16 shows the original inputs and the segmentation into building and non-building classes from a close-up view of two sample areas.

Subsequently, BBR is used to remove the remaining shortcomings in the extracted building boundaries. Figure 17 illustrates the outcome of using the BBR to correct overshoot and

**Fig. 17** Result of applying BBR algorithm to the building objects.

undershoot errors in several sample test areas. Evidently, the majority of remaining building boundary errors are eliminated, and results are more acceptable.

## 3.5 Results of Multi-Source Modification

After detecting the building boundaries and generating fine-registered enhanced thermal ortho-photo, temperature profiles are determined around each boundary pixel at the edges of buildings. Temperature profiles for some edges with proper quality and distribution in the thermal image are calculated to determine the smoothness of the edges. After that, the mean temperature profile is determined for these temperature profiles. Figure 18(a) depicts these temperature profiles (color-ful dashed lines) and their mean (thick green line). Considering this mean profile as the ESF curve, the LSF curve is calculated and its FWThM is determined as equal to 13 pixels according to Eq. (15). That is, on each side of the edge pixel, an average of 6 pixels are affected by the edge smoothness. Figure 18(b) depicts the relationship between the calculated mean profile (green curve) and its LSF (red curve).

The point with the highest slope in the temperature profile is searched in the range of twice the smoothness of the edges to find the temperature edge around the boundary of the buildings. After fine-registering the edges, the temperature smoothness at the edges is removed.

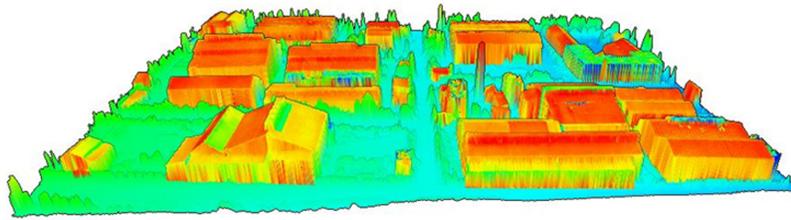Figures 19 and 20 display a 2D view and a 3D view of HR 4D thermal surface model representation of all the buildings in the studied area, respectively.

(a) (b)
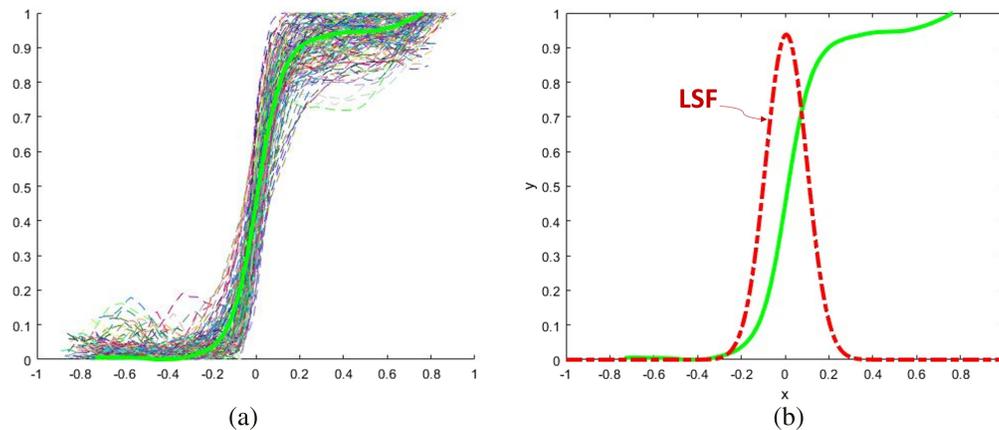
Fig. 18 (a) Temperature profiles and their mean profile. (b) Relationship between the calculated mean profile and its LSF.
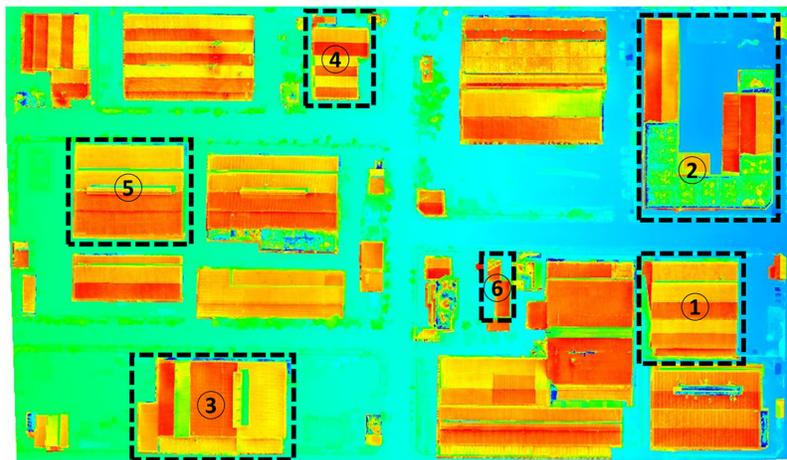
Fig. 19 2D view of HR 4D thermal surface model representation of all the buildings in the studied area.

Fig. 20 3D view of HR 4D thermal surface model representation of all the buildings in the studied area.

4 Discussion

In this research, a method is proposed to generate an HR 4D thermal surface model of buildings based on the integration of visible and thermal UAV imagery.

For data collection, visible and thermal imaging has been performed on two flights and at two separate times because thermal imaging sensors can display hot areas with greater contrast at night. The reason for this is that the ambient temperature, and more importantly, the core temperatures of unheated objects and surroundings, are often substantially lower at night than during the day. In addition, using separate thermal and visible cameras is more cost-effective than using multiple camera systems. Furthermore, in multiple camera systems, it is challenging to identify the internal orientation characteristics of the thermal camera using precise calibration between cameras.[23]

Considering the limited resolution of thermal images, even the enhanced surface model cannot achieve the desired quality required in many fields due to the scaling constraint on increasing the resolution.[21,22] Therefore, this study attempts to generate an HR 4D thermal surface model of buildings in two phases. In the first phase, it generates an enhanced surface model and an enhanced thermal orthophoto from thermal images, the resolution of which is closer to that of DSM and visible orthophoto, respectively and facilitates registration. In the second phase, an HR 4D thermal surface model of the buildings is created by integrating the DSM of the buildings and the temperature extracted from the enhanced thermal orthophoto.

In the first phase, because the implementation of the EDSR network is easy, it is used for the SR of thermal images. The use of other DL networks is suggested to reduce registration challenges.

In the building extraction step, using nDSM instead of DSM in the SegNet-based DL semantic segmentation strategy removes the effects of ground slope in extracting buildings. However, it seems possible to increase the overall accuracy of building extraction by adding other features, such as the vegetation mask, and enlarging the training data for segmentation. Besides, the BBR method overcame some of the shortcomings of determining the exact boundaries of buildings.

Note that, in the second phase, the search range to find the temperature edge around the boundary of the buildings should be proportional to the accuracy of registration. Determining too large or too small a size of this search range can lead to errors in the process of assigning temperature to the edges of buildings.

It is crucial to note that in the studies conducted so far, after creating a surface model (or point cloud) from visible images, the temperature is assigned to the HR surface data using various techniques, such as interpolation. In those studies, it is assumed that registration is carried out with the maximum possible accuracy, and the possibility of shifting between the boundaries extracted from thermal and visible data is not taken into account. However, in this research, the average distance between the positions of thermal and visible building boundaries in the original model (generated from original thermal orthophoto) is calculated to be about 6 pixels. However, this distance was reduced to about 4 pixels in the enhanced model (generated from enhanced thermal orthophoto) and less than 1 pixel in the HR model (generated from proposed method). This means the average difference in the position of the temperature edges and building boundaries was reduced by about 83.3%. Examples of how the temperature profiles are positioned in relation to the height profile of the building's edge are illustrated in Fig. 21.
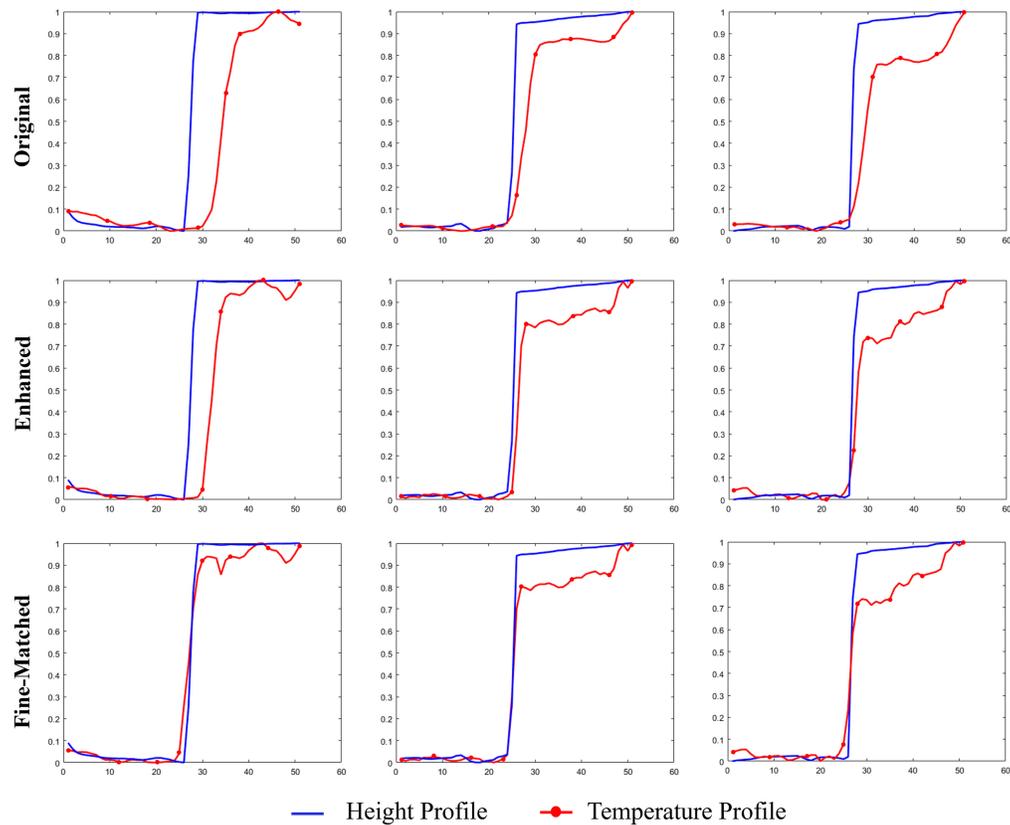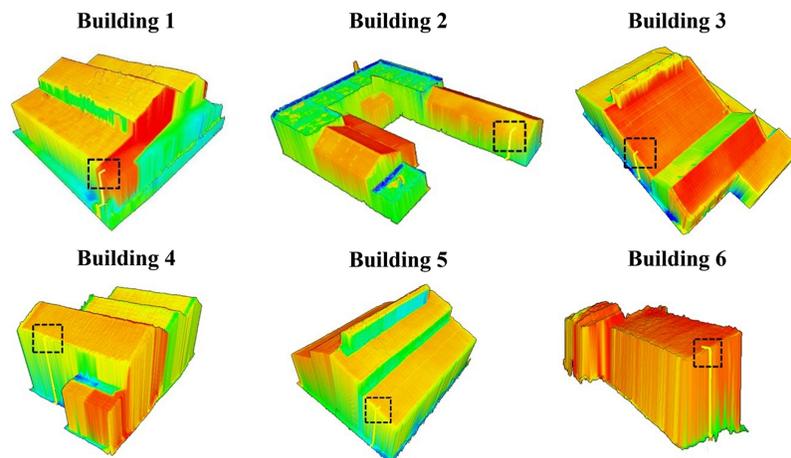
Fig. 21 Comparison of the temperature profile (red) and height profile (blue).

This figure displays the temperature profiles in comparison to the height profile in three modes: original, enhanced, and fine-matched. As it is clear in Fig. 21, the shift between the temperature profiles compared to the height profile in the second row is less than in the first row. Moreover, in the third row, the displacement value is reduced to less than one pixel.

In the proposed method, the smoothness of temperature at the edges is removed. A profile perpendicular to the edge of each highlighted building in Fig. 19 is determined and analyzed to investigate this issue. Figure 22 shows the 3D view of the highlighted buildings and the position of the specified profiles.

Figure 23 depicts the profile marked on each building in Fig. 21 in more detail.

Fig. 22 3D view of selected buildings and edge profile position on each building.
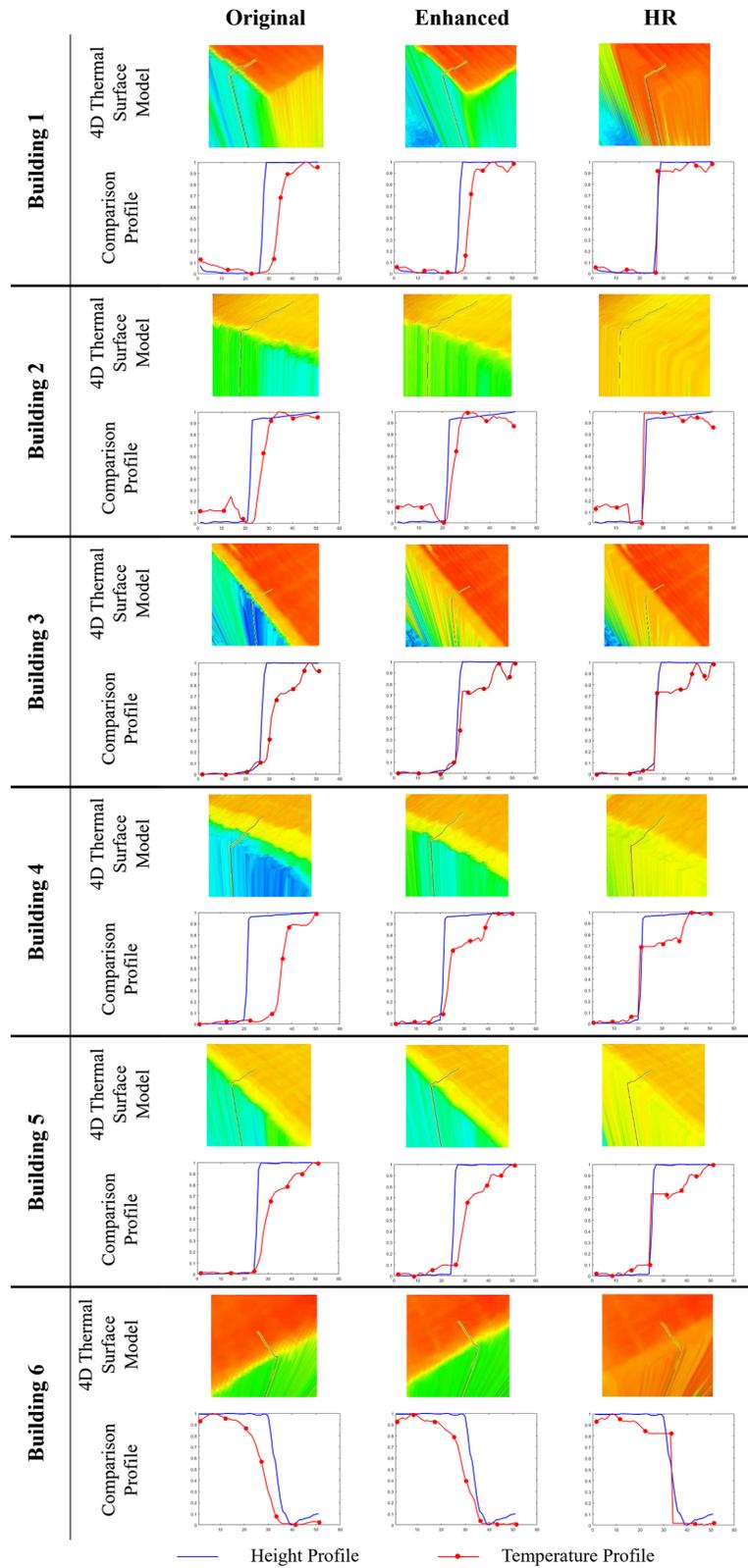
Fig. 23 Comparison of 4D thermal surface models and temperature profile data.

In the comparison profiles in Fig. 23, the blue curve shows the height profile, and the red curve shows the temperature profile. As shown in Fig. 23, the HR 4D thermal surface model has a better match between the visible and thermal edges of the buildings in addition to the high spatial resolution obtained from visible data. Furthermore, the smoothness of the edges is not present in this model.

Note that the original profile has a wider range of temperature smoothness compared to the enhanced profile. That is, if the single-source modification stage is skipped, the area where the temperature changes would be larger. However, due to the proper performance of SR methods in enhancing the spatial resolution, performing the single-source modification step increases the accuracy of the assigned temperature.

Therefore, the main challenge in this research was the accurate matching of the thermal boundary and the object boundary, as well as removing the temperature smoothness at the edges of the buildings, which was successfully overcome. At the end of the implementation of the proposed method, an HR 4D thermal surface model was generated that can be utilized to optimize energy usage.

5 Conclusions

This research proposed a two-phase strategy to generate an HR 4D thermal surface model. In the single-source modification phase, an enhanced surface model and an enhanced thermal orthophoto were produced using DL-based SISR methods. Due to the limitation of scale increment in SR methods, visible orthophoto and DSM were first created from UAV visible images. Then, in the building extraction step, after retraining a SegNet network, the nDSM and visible orthophoto were fed to the trained model, and the buildings were extracted. After that, data integration was performed in the multi-source modification phase. In this way, a fine-registered enhanced thermal orthophoto was first generated. Following the temperature edges' smoothness range determination, the building's boundaries and the temperature edges first coincided, and then, the temperature smoothness at the edges was eliminated.

The results demonstrated great accuracy in both thermal and spatial information of the produced HR 4D thermal surface model. The matching of temperature edges and visible boundaries was increased by about 83.3%. The temperature smoothness at the edges caused by the LR of thermal images was completely eliminated.

The quality of the generated model is directly affected by the accurate determination of the building's boundaries; thus, even though the evaluation confirmed the effectiveness of the proposed method, future studies should be focused on more accurate extraction of buildings and their boundaries. For more accurate investigations of thermal anomalies, the mismatch of temperature edges and boundaries of objects on the roofs should be found and removed. Furthermore, the temperature edges' smoothness in the inner area of the roof of each building should be removed. Additionally, to determine the absolute temperature, the thermal sensor radiometric calibration is advised with the intention of employing the HR 4D thermal surface model for interpretation.

Disclosures

Code and Data Availability

References

1. Z. Chen et al., "Interpretable machine learning for building energy management: a state-of-the-art review," *Adv. Appl. Energy* **9**, 100123 (2023).

2. S. Brandi, M. Fiorentini, and A. Capozzoli, "Comparison of online and offline deep reinforcement learning with model predictive control for thermal energy management," *Autom. Constr.* **135**, 104128 (2022).

3. E. Mandanici et al., "A multi-image super-resolution algorithm applied to thermal imagery," *Appl. Geomat.* **11**(3), 215–228 (2019).

4. J. Pan, "Analysis of human factors on urban heat island and simulation of urban thermal environment in Lanzhou city, China," *J. Appl. Remote Sens.* **9**(1), 095999 (2015).

5. H. Goudarzi and A. Mostafaeipour, "Energy saving evaluation of passive systems for residential buildings in hot and dry regions," *Renew. Sustain. Energy Rev.* **68**, 432–446 (2017).

6. J. Zhang et al., "Thermal infrared inspection of roof insulation using unmanned aerial vehicles," *Int. Arch. Photogramm Remote Sens. Spatial Inf. Sci* **40**(1), 381 (2015).

7. E. Maset et al., "Photogrammetric 3D building reconstruction from thermal images," *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci.* **IV-2/W3**, 25 (2017).

8. S. Farrag, S. Yehia, and N. Qaddoumi, "Investigation of mix-variation effect on defect-detection ability using infrared thermography as a nondestructive evaluation technique," *J. Bridge Eng.* **21**(3), 04015055 (2016).

9. D. Borrmann, J. Elseberg, and A. Nüchter, "Thermal 3D mapping of building façades," in *Intelligent Autonomous Systems*, S. Lee et al., Eds., Vol. **12**, pp. 173–182, Springer, Berlin, Heidelberg (2013).

10. Y. Ham and M. Golparvar-Fard, "Rapid 3D energy performance modeling of existing buildings using thermal and digital imagery," in *Constr. Res. Congr. 2012: Constr. Challenges in a Flat World* (2012).

11. D. Borrmann et al., "A mobile robot based system for fully automated thermal 3D mapping," *Adv. Eng. Inf.* **28**(4), 425–440 (2014).

12. M.-D. Yang, T.-C. Su, and H.-Y. Lin, "Fusion of infrared thermal image and visible image for 3D thermal model reconstruction using smartphone sensors," *Sensors* **18**(7), 2003 (2018).

13. I. Campione et al., "3D thermal imaging system with decoupled acquisition for industrial and cultural heritage applications," *Appl. Sci.* **10**(3), 828 (2020).

14. J. Zhu et al., "Generation of thermal point clouds from uncalibrated thermal infrared image sequences and mobile laser scans," *IEEE Trans. Instrum. Meas.* **72**, 1–16 (2023).

15. M. Xu et al., "SMFD: an end-to-end infrared and visible image fusion model based on shared-individual multi-scale feature decomposition," *J. Appl. Remote Sens.* **18**(2), 022203 (2024).

16. M. Previtali et al., "Rigorous procedure for mapping thermal infrared images on three-dimensional models of building façades," *J. Appl. Remote Sens.* **7**(1), 073503 (2013).

17. L. Yue et al., "Image super-resolution: the techniques, applications, and future," *Signal Process.* **128**, 389–408 (2016).

18. Y. Zhang et al., "A CNN-based subpixel level DSM generation approach via single image super-resolution," *Photogramm. Eng. Remote Sens.* **85**(10), 765–775 (2019).

19. P. Burdziakowski, "Increasing the geometrical and interpretation quality of unmanned aerial vehicle photogrammetry products using super-resolution algorithms," *Remote Sens.* **12**(5), 810 (2020).

20. M. Pashaei et al., "Deep learning-based single image super-resolution: an investigation for dense scene reconstruction with UAS photogrammetry," *Remote Sens.* **12**(11), 1757 (2020).

21. A. Fallah et al., "Intensifying the spatial resolution of 3D thermal models from aerial imagery using deep learning-based image super-resolution," *Geocarto Int.* **37**, 116 (accepted) (2022).

22. J. Shermeyer and A. Van Etten, "The effects of super-resolution on object detection performance in satellite imagery," in *Proc. IEEE Conf. Comput. Vision and Pattern Recognit. Workshops* (2019).

23. F. Javadnejad et al., "A photogrammetric approach to fusing natural colour and thermal infrared UAS imagery in 3D point cloud generation," *Int. J. Remote Sens.* **41**(1), 211–237 (2020).

24. C. Daffara et al., "A cost-effective system for aerial 3D thermography of buildings," *J. Imaging* **6**(8), 76 (2020).

25. A. Sledz, J. Unger, and C. Heipke, "Thermal IR imaging: image quality and orthophoto generation," *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.-ISPRS Arch.* **XLII-1**, 413–420 (2018).

26. M. Dahaghin et al., "Precise 3D extraction of building roofs by fusion of UAV-based thermal and visible images," *Int. J. Remote Sens.* **42**(18), 7002–7030 (2021).

27. Z. E. Wakeford et al., "Combining thermal imaging with photogrammetry of an active volcano using UAV: an example from Stromboli, Italy," *Photogramm. Rec.* **34**(168), 445–466 (2019).

28. J. Paziewska and A. Rzonca, "Integration of thermal and RGB data obtained by means of a drone for interdisciplinary inventory," *Energies* **15**(14), 4971 (2022).

29. K. Yan et al., "A decoupled calibration method for camera intrinsic parameters and distortion coefficients," *Math. Prob. Eng.* **2016**, 1–12 (2016).

30. Z. Zhang, "A flexible new technique for camera calibration," *IEEE Trans. Pattern Anal. Mach. Intell.* **22**(11), 1330–1334 (2000).

31. P. Grussenmeyer and O. A. Khalil, "Solutions for exterior orientation in photogrammetry: a review," *Photogramm. Rec.* **17**(100), 615–634 (2002).

32. D. C. Brown, "Close-range camera calibration," *Photogramm. Eng.* **37**(8), 855–866 (1971).

33. C. Zhao et al., "Effects of spatial resolution on image registration," *Proc. SPIE* **9784**, 97840Y (2016).

34. B. Lim et al., "Enhanced deep residual networks for single image super-resolution," in *Proc. IEEE Conf. Comput. Vision and Pattern Recognit. Workshops* (2017).

35. W. Yang et al., "Deep learning for single image super-resolution: a brief review," *IEEE Trans. Multimedia* **21**(12), 3106–3121 (2019).

36. T. Phuc Truong et al., "Registration of RGB and thermal point clouds generated by structure from motion," in *Proc. IEEE Int. Conf. Comput. Vision Workshops* (2017).

37. H. Hirschmüller, "Semi-global matching-motivation, developments and applications," *Photogramm. Week* **11**, 173–184 (2011).

38. M. Jauregui et al., "Digital orthophoto generation," *Int. Arch. Photogramm. Remote Sens.* **33**(B4/1; Part 4), 400–407 (2000).

39. E. Maggiori et al., "Can semantic labeling methods generalize to any city? the inria aerial image labeling benchmark," in *IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)* (2017).

40. S. Tian et al., "A novel deep embedding network for building shape recognition," *IEEE Geosci. Remote Sens. Lett.* **14**(11), 2127–2131 (2017).

41. V. Badrinarayanan, A. Kendall, and R. Cipolla, "Segnet: a deep convolutional encoder-decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.* **39**(12), 2481–2495 (2017).

42. V. Badrinarayanan, A. Handa, and R. Cipolla, "SegNet: a deep convolutional encoder-decoder architecture for robust semantic pixel-wise labelling," arXiv:1505.07293 (2015).

43. A. Kendall, V. Badrinarayanan, and R. Cipolla, "Bayesian SegNet: model uncertainty in deep convolutional encoder-decoder architectures for scene understanding," arXiv:1511.02680 (2015).

44. K. Simonyan, "Very deep convolutional networks for large-scale image recognition," arXiv:1409.1556 (2014).

45. W. Boonpook et al., "A deep learning approach on building detection from unmanned aerial vehicle-based images in riverbank monitoring," *Sensors* **18**(11), 3921 (2018).

46. Y. Xu et al., "Building extraction in very high resolution remote sensing imagery using deep learning and guided filters," *Remote Sens.* **10**(1), 144 (2018).

47. W. Boonpook, Y. Tan, and B. Xu, "Deep learning-based multi-feature semantic segmentation in building extraction from images of UAV photogrammetry," *Int. J. Remote Sens.* **42**(1), 1–19 (2021).

48. W. Sun and R. Wang, "Fully convolutional networks for semantic segmentation of very high resolution remotely sensed images combined with DSM," *IEEE Geosci. Remote Sens. Lett.* **15**(3), 474–478 (2018).

49. D. Marmanis et al., "Classification with an edge: improving semantic image segmentation with boundary detection," *ISPRS J. Photogramm. Remote Sens.* **135**, 158–172 (2018).

50. T. Krauß, "A new simplified DSM-to-DTM algorithm–DSM-to-DTM-step," https://www.preprints.org/manuscript/201807.0017/v1 (2018).

51. B. Bischke et al., "Multi-task learning for segmentation of building footprints with deep neural networks," in *IEEE Int. Conf. Image Process. (ICIP)* (2019).

52. H. Mohammadi, F. Samadzadegan, and P. Reinartz, "2D/3D information fusion for building extraction from high-resolution satellite stereo images using kernel graph cuts," *Int. J. Remote Sens.* **40**(15), 5835–5860 (2019).

53. Z. Wu et al., "Medical image registration using B-spline transform," *Int. J. Simul. Syst. Sci. Technol* **17**(48), 1.1–1.6 (2016).

54. F. D. Javan, F. Samadzadegan, and P. Reinartz, "Spatial quality assessment of pan-sharpened high resolution satellite imagery based on an automatically estimated edge based metric," *Remote Sens.* **5**(12), 6539–6559 (2013).

55. M. Zhou et al., "Land cover classification from full-waveform LIDAR data based on support vector machines," *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.* **41**, 447–452 (2016).

56. Y. Mei-Woo, "Determination performance of gamma spectrometry co-axial hpGE detector in Radio-chemistry and Environment Group, Nuclear Malaysia," in *Research and Development Seminar*, Malaysia (2014).

57. C. D. Prakash and L. J. Karam, "Camera calibration using adaptive segmentation and ellipse fitting for localizing control points," in *19th IEEE Int. Conf. Image Process.* (2012).

58. J.-N. Ouellet and P. Hébert, "Precise ellipse estimation without contour point extraction," *Mach. Vision Appl.* **21**(1), 59 (2009).

Alaleh Fallah is a PhD candidate in photogrammetry and remote sensing at the University of Tehran, Iran. She received her BSc degree in surveying engineering and MSc degree in photogrammetry from the University of Tehran, Iran, in 2011 and 2014, respectively. Her research interests include high-resolution 3D model generation, thermal and visible data fusion, and deep learning.

Farhad Samadzadegan received his PhD in photogrammetry engineering from the University of Tehran, Tehran, Iran, in 2001. Currently, he is working as a full professor in the Faculty of Surveying and Geospatial Engineering at the University of Tehran, Tehran, Iran. He has more than 20 years of experience in designing and developing digital photogrammetric and remote sensing software and systems.

Farzaneh Dadrass Javan is an assistant professor at ITC faculty, University of Twente, the Netherlands. She received her MSc and PhD degrees from the University of Tehran, Tehran, Iran, in 2010 and 2014, respectively, both in photogrammetry. Currently, she is mainly focused on UAV-based applications in geo-science and the fusion of UAVs with other satellite- and terrestrial-based sources of data.