



Intensifying the spatial resolution of 3D thermal models from aerial imagery using deep learning-based image super-resolution

Alaleh Fallah, Farhad Samadzadegan & Farzaneh Dadrass Javan

To cite this article: Alaleh Fallah, Farhad Samadzadegan & Farzaneh Dadrass Javan (2022): Intensifying the spatial resolution of 3D thermal models from aerial imagery using deep learning-based image super-resolution, Geocarto International, DOI: [10.1080/10106049.2022.2082544](https://doi.org/10.1080/10106049.2022.2082544)

To link to this article: <https://doi.org/10.1080/10106049.2022.2082544>



© 2022 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



Published online: 22 Jun 2022.



Submit your article to this journal [↗](#)



Article views: 293



View related articles [↗](#)



View Crossmark data [↗](#)

Intensifying the spatial resolution of 3D thermal models from aerial imagery using deep learning-based image super-resolution

Alaleh Fallah^a , Farhad Samadzadegan^a  and Farzaneh Dadrass Javan^{a,b} 

^aSchool of Surveying and Geospatial Engineering, College of Engineering, University of Tehran, Tehran, Iran; ^bDepartment of Geo-Information Science and Earth Observation (ITC), University of Twente, Enschede, The Netherlands

ABSTRACT

Nowadays, 3D thermal models can play an important role in buildings' energy management while acquiring multisource data to generate a high-resolution 3D thermal model. Consequently, in this article, a method for intensifying 3D thermal model using deep learning-based image super-resolution is presented. In the proposed method, first, the enhanced deep residual super-resolution (EDSR) deep network is re-trained based on thermal aerial images. Second, the resolution of low-resolution thermal images is enhanced using the newly trained network. Finally, the state-of-the-art structures from motion (SfM), semi global matching (SGM) and space intersection are utilized to generate intensified 3D thermal model from the resolution enhanced thermal images. Spatial evaluations indicate a 5% increase in edge-based image fusion metric (EFM) for the intensified 3D model. Besides, the evaluations show that the modulation transfer function (MTF) curves of the intensified 3D thermal model are closer to a reference model against the original 3D thermal model.

ARTICLE HISTORY


Received 11 August 2021
Accepted 22 May 2022

KEYWORDS

Thermal image enhancement; 3D thermal model intensification; deep learning; single image super-resolution (SISR)

HIGHLIGHTS

- A 3D thermal model intensification solution using EDSR is proposed which is independent of hardware techniques and multi-source data.
- Considering the importance of edge sharpness in the intensified 3D thermal model, the quality of edges is assessed using MTF curves and the EFM metric.
- In comparison to the original 3D thermal model, the MTF curves of the intensified 3D thermal model are closer to the MTF curve of the high-resolution 3D model.
- The EFM metric shows higher values for MTF curves of the intensified 3D thermal model against MTF curves of the original 3D thermal model.

CONTACT Farzaneh Dadrass Javan  f.dadrassjavan@utwente.nl

© 2022 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

1. Introduction

Nowadays, the use of a 3D thermal model generated from aerial thermal images, in which the digital number (DN) values from thermal images are mapped, are more commonly applied. These models are mainly used in building inspections which helps in detecting of heat losses, incomplete insulation in roofs, cracks, air and moisture leakages, etc. Also, detecting areas with energy leaks to reduce energy consumption plays an important role in buildings energy management (Rakha et al. 2018). Thus, it seems necessary to have accurate and precise information about the surface temperature and its spatial pattern (Mandanici et al. 2019). In this regard, 3D thermal models that present thermal information and provide 3D building roofs information can be used to detect, interpret and measure thermal anomalies in building and roof inspections (Borrmann et al. 2013).

One of the challenges of having such a 3D thermal model is its spatial resolution. In fact, because of the need for larger IFOVs in thermal cameras (to ensure that enough energy reaches the detector), the spatial resolution of thermal images is usually fairly coarse. Consequently, the 3D thermal model generated only based on thermal images provides a low spatial resolution and few details that may challenge the process of detecting, interpreting and measuring thermal anomalies (Khodaei et al. 2015).

The proposed methods to generate a high-quality 3D thermal model can be divided into two main groups as multi-source and single-source methods. In multi-source methods, researchers use thermal images and information from other resources such as laser scanners data and RGB images or panchromatic images simultaneously (Oreifej et al. 2014; Borrmann et al. 2013; Antón and Amaro-Mellado 2021). That is while in single source methods, only thermal images are used and attempt to improve the quality of the final product.

The multi-source methods are classified into two categories of methods that focus on image space and those works that on object space. Research that increases thermal image resolution by fusing it with RGB or panchromatic images falls into the first category (Ma et al. 2019). In the methods focus on object space, researchers have used fusion-based techniques for enhancing the accuracy of 3D thermal models. As an instance, in some studies the 3D model is generated using RGB images or laser scanner data and then, the thermal information is mapped onto the 3D model (Ham and Golparvar-Fard 2013; Yang et al. 2018; Javadnejad et al. 2020; Alba et al. 2011; Borrmann et al. 2013). In other studies, 3D models from thermal and RGB images have been generated separately and registered by different methods to enhance the resolution of the 3D thermal model (Maset et al. 2017; Javan and Savadkouhi 2019; Dahaghin et al. 2021).

Obviously, the multi-source methods are not applicable in the cases where only thermal images are available. Acquiring multisource data and registering data from various data sources are the challenging process because different sensors are concentrated on different operating ranges and environmental conditions (Zhang et al. 2019). In addition, regarding the simultaneous use of visible and thermal sensors, it seems necessary to mention these points that first dual sensors that capture thermal and RGB images at the same time are normally too expensive and not cost-effective for lots of projects. Second, thermal images should be taken at night and when there is no light reflection, while RGB images should be taken during the day and when enough light exists. Simultaneous capturing of these two images is used mainly in cases related to interpretation application. In the 3D information extraction projects, it is mostly recommended not to take RGB and thermal images at the same time.

In the single source methods, an attempt is made to increase the 3D thermal model resolution only by increasing the thermal image resolution. One way to achieve this is the

resolution enhancement of thermal images through ‘hardware’ techniques, which entails high costs and many limitations (Yue et al. 2016). Therefore, resolution enhancement of thermal images independent of hardware techniques can be more practical for generating high-quality 3D thermal models.

As another solution in the single source methods, super-resolution (SR) techniques that use images of just one sensor have become a potential way for obtaining high resolution images (Yue et al. 2016) and, then, high-quality 3D model. Indeed, SR is a technique that reconstructs a higher-resolution image or sequence from the given low resolution images (Yue et al. 2016; Dong et al. 2016).

Against various methods of producing super-resolution images, single-image super-resolution (SISR) techniques that use limited low-resolution information from a single image to estimate the mapping from low-resolution to high-resolution space have been used in many studies (Kim and Kwon 2010). The SISR methods are classified into three groups including interpolation-based methods that suffer from accuracy deficiency (for example, bi-cubic interpolation), reconstruction-based methods that are usually time-consuming and learning-based methods (Yang et al. 2019). Among the learning-based methods, deep learning (DL) solutions, particularly convolutional neural networks (CNNs), are superior to other methods because they are actually able to enhance the data in an information-theoretical sense (Kansal and Nathan 2020; Liebel and Körner 2016).

Nowadays, researchers have focused on providing networks that can produce better high-resolution images against simple up-sampling methods such as bi-cubic interpolation. Some of these networks are CNN-based SR (SR-CNN)(Dong et al. 2016), very deep super-resolution (VDSR) (Kim et al. 2016), enhanced deep residual super-resolution (EDSR) (Lim et al. 2017), super-resolution network for multiple degradations (SRMD) (Zhang et al. 2018), very deep residual channel attention networks (RCAN) (Dai et al. 2019), etc.

Researchers have also used SR techniques in the 3D information extraction from RGB images. For example, Zhang et al. (2019) generated high-quality DSMs by improving the quality of images using super-resolution methods. They applied various super-resolution methods to the RGB images and compared the quality of the resulting DSMs. In another work, Burdziakowski (2020) utilized SR algorithms to increase the geometric and interpretative quality of the final photogrammetric products. They concluded the photogrammetric products generated from high-altitude images processed by the SR algorithm showed a similar quality to the reference products generated from low-altitude images and, in some cases, even improved their quality. In Pashaei et al. (2020), the ability of a DCNN-based SISR model, named enhanced super-resolution generative adversarial network (ESRGAN), to predict the spatial information degraded or lost in a hyper-spatial resolution unmanned aerial system (UAS) RGB image set was studied. Results showed the accurate extraction of interior and exterior imaging geometry from a super-resolved image set.

Although the super-resolution of RGB and panchromatic images has been extensively studied, the DL-based enhancement of thermal images is a newer field of research. In some research for thermal images enhancement, network training is done based on images from the visible spectrum at different colour space representations (Choi et al. 2016; Almasri and Debeir 2018). In contrast to these methods, a number of studies in recent years have focused on the production of high-resolution thermal images through network training using low-resolution thermal images (Rivadeneira et al. 2019; Kansal and Nathan 2020).

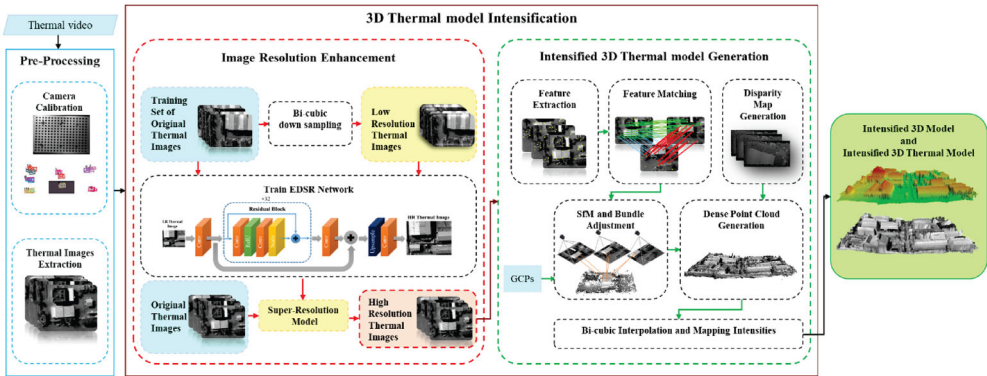


Figure 1. Proposed method for generating high-resolution 3D thermal model from low-resolution aerial thermal images.

This research proposes to generate a high-resolution 3D thermal model from lower resolution aerial thermal imagery captured by a low-cost camera mounted a low weight drone. For this purpose, super-resolution thermal images are generated based on training a deep network while quality of aerial thermal images and the information content is lower in thermal imagery than RGB images. This research aims also to investigate how much this enhancement can improve the spatial resolution of the final 3D thermal model generated based on low quality aerial thermal images. For the purpose, a DL-based super-resolution network is trained and applied to generate intensified 3D thermal model. To evaluate the capability of the proposed methodology, the effect of SISR algorithm on 3D thermal model intensification is investigated by measuring spatial resolution evaluation criteria and answer to the question whether this low-cost solution can help to improve the quality of 3D models generated by low quality and cost images.

The remainder of this article is organized as follows: [Section 2](#) describes the methodology, the quality assessment metrics for evaluating the intensified 3D thermal model and dataset that is used in this article. In [Section 3](#), the results of generating a 3D thermal model based on the proposed solution are presented and discussed in detail. Lastly, [Section 4](#) provides a conclusion and future perspective.

2. Materials and methods

2.1. Methodology

Given the importance of producing high-quality 3D thermal the problems of conventional solutions, increasing the spatial resolution of aerial thermal images using DL-based SISR seems to be a practical solution for producing intensified 3D thermal models. In [Figure 1](#) the proposed method to generate an intensified 3D thermal model is presented. Accordingly, after training the selected deep network, the spatial resolution of each thermal image extracted from the thermal video will be increased, then, intensified 3D model is generated from generated high-resolution aerial thermal images. Afterwards, the DN values of high-resolution aerial thermal images are mapped onto the intensified 3D model and the intensified 3D thermal model is generated. It is noteworthy mentioning that converting thermal image DN values into absolute temperature values is not in the scope of this article. If needed, radiometric calibration and conversion of thermal DN value into temperature could be implemented.

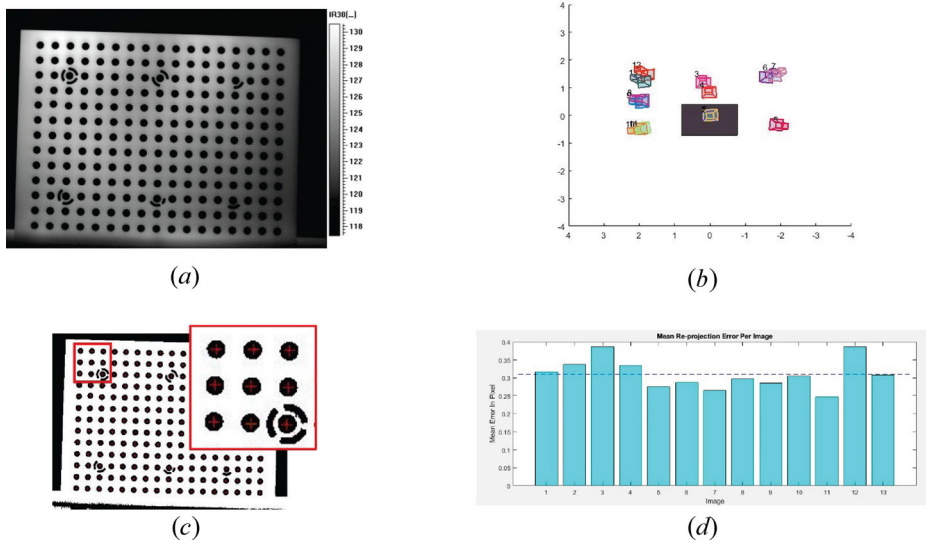


Figure 2. Thermal camera calibration. Calibration target (a), Camera positions (b), Detection of centres of circles based on Ouellet method (c), and mean re-projection error chart (d).

2.1.1. Pre-processing

This step includes two basic phases of thermal image extraction from captured string video frames and the thermal camera calibration. The thermal camera captures the video which is converted into images. The size of extracted image frames is 640×480 pixels.

One of the main photogrammetric parts here is camera calibration, which intends to calculate the parameters of camera model from 2D images (Peng and Li 2010). Currently, the techniques for camera calibration are classified into two main groups: the traditional camera pre-calibration and the self-calibration strategies. The traditional camera calibration methods are performed before using the photogrammetric procedure and solve the camera parameters based on accurate scene information, such as points or lines with precise coordinates. Thus, these methods normally result in more accurate camera calibration (Yan et al. 2016). To this end, an appropriate test field must be designed. Many effective factors such as simple structure, material, shape, and target dimensions should be considered in test field designing. The features of the test field, as well as their coordinates, should be easily identified as well. The projection between image location of features and their object coordinates are then employed to conclude the camera calibration parameters.

In a study conducted by Usamentiaga et al. (2017), the camera calibration process was done using both chessboards and circular patterns. Comparing the obtained results revealed better accuracy for the circular pattern. Consequently, for the proposed method, a planar circular test field is designed and images are captured from several directions and orientations based on Zhang's method (Zhang 2000). The test field is a rectangular calibration board (with 13×17 hollow circles). The diameter of circles is 12 mm and the distances between their centres are 24 mm. Also, six coded targets are embedded on the calibration board to identify the position of targets in each of images (Figure 2a). Due to the limitation of the wavelength detectable by thermal cameras, the calibration board is heated for better detection and higher contrast, followed by imaging from multiple views (Figure 2b). After imaging, adaptive thresholding is used to generate binary images (Prakash and Karam 2012). In the next step, the geometric centres of the circles are identified (Ouellet and Hébert 2009) (Figure 2c).

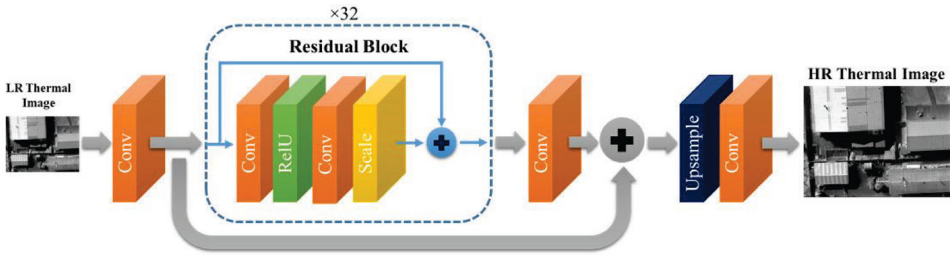


Figure 3. Network structure of EDSR.

After determining the geometric centre of targets in all images, each point is re-projected from object space to image space by collinearity equations (Equations (1) and (2)) (Javan and Savadkouhi 2019):

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

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

In the above equations, x and y are image coordinates, c is focal length, X_0 , Y_0 , and Z_0 are coordinates of projection center, X , Y , and Z are object coordinates, and r denotes elements of the rotation matrix. To determine lens distortion parameters, Brown model equations are used (Equation (3) and (4)) (Brown 1971):

$$x' = x(1 + k_1r^2 + k_2r^4 + k_3r^6 + p_2(r^2 + 2x^2 + 2p_1xy)) \quad (3)$$

$$y' = y(1 + k_1r^2 + k_2r^4 + k_3r^6 + p_1(r^2 + 2y^2 + 2p_2xy)) \quad (4)$$

where x' and y' show image coordinates that have no distortion, k_i represents the lens radial distortion parameter, p_i is lens decentring distortion factor, and r is the radial distance from the perspective point projected on the image plane.

In geometric calibration, the mathematical model converting from the target coordinate system and the corresponding coordinates in the image space is also computed. Therefore, the accuracy of the calibration algorithm can be estimated based on the mean re-projection error. Figure 2d presents the mean re-projection error per image. As it is obvious from the figure, the average value of 0.315 pixels is estimated over 13 images and 221 calibration points.

2.1.2. 3D thermal model intensification

As shown in Figure 1, this phase consists of two main steps. In the first step, the process of producing a high-resolution thermal image from a low-resolution one is done by training a DL-based SISR model. In the next step, the process of producing the intensified 3D thermal model is performed based on the outputs of the previous steps. Details of these two steps are given in the followings.

2.1.2.1. Image resolution enhancement. To enhance the thermal image resolution, the EDSR network, which is a convolutional neural network (CNN), is used to apply SISR (Lim et al. 2017). Since in this study showing the efficiency of SISR methods in 3D thermal model intensification is important, the EDSR network is chosen due to its simplicity of implementation and acceptable reported performance in recent research (Yang et al. 2019). The network structure of EDSR is presented in Figure 3. The EDSR network has

not yet been trained for thermal aerial imagery and for the first time, its capability in producing an aerial thermal super-resolved image is investigated here.

The main difference between low-resolution and high-resolution images is in high-frequency details such that the EDSR network can learn the mapping between low- and high-resolution images. Due to the necessity of reference high-resolution images for deep network training, in this step, original thermal images are down-sampled by the bi-cubic method to generate low-resolution thermal images as the network input (Lim et al. 2017). This makes it possible to compare the output of the EDSR network with the reference high-resolution image (actually original thermal images).

Consider Th_{LR} as a low-resolution thermal image, Th_{HR} as a high-resolution thermal image and \widehat{Th}_{HR} as the estimated high-resolution thermal image, the goal is to train model f , which holds in Equation (5).

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

In other words, the EDSR network generates a resolution enhanced image \widehat{Th}_{HR} by minimizing the distance between $f(Th_{LR})$ and Th_{HR} .

The EDSR network is trained using the mean absolute error loss function (L1 loss) instead of L2. Lim et al. (2017) found that L1 loss results in better convergence than L2. The L1 loss function that should be minimized is presented in Equation (6). In this equation, m is the number of the rows of the images, i represents the index of each row, n is the number of columns of the image and j represents the index of the column.

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

In selecting the images for training, an attempt is made to include a variety of features (such as buildings, roads, trees, land) in the training dataset. The extracted frames are first converted into low-resolution images by the bi-cubic interpolation method with the scale factor of 2; then 70% of these images are used for the training process and the remaining 30% images are used for validating the accuracy of the training process. It should be noted that the use of down-sampled images and original images for deep network training is a common a technique in this field and research by (Shermeyer and Van Etten 2019; Lim et al. 2017) show that the network learns to produce the super-resolved image independent of the spatial resolution of the input image.

The training set of low-resolution images and their corresponding original thermal images are entered into the EDSR network for training. It is important to note that in order to prepare a thermal image similar to the RGB image and to allow entry into the EDSR network for training; the thermal image DN values are repeated in three channels. The standard network parameters as set in the original paper (Lim et al. 2017) are used. The patch size of 48×48 pixels is chosen and augmented by random horizontal flips and 90 rotations per patch. The use of image patches and augmentation of the data during the training process effectively increase the amount of available training data and ensure the production of a robust model. Moreover, a variety of features in the images selected for training help to more robustness of the trained model. Since the EDSR network is not been used before for high resolution aerial thermal images, for this study it is developed in C ++.

After the training process, the map for predicting high-resolution image from the input low-resolution image is determined. In this way, any low-resolution image can be super-resolved and the high-resolution image can be produced. Thus, each of another original

thermal images are extracted from the video of the study area is separately converted into high-resolution thermal images using the trained model and scale factor 2. These enhanced images are used to generate the intensified 3D thermal model in the next steps.

2.1.2.2. Intensified 3D thermal model generation. For generating an intensified 3D thermal model from a set of high-resolution thermal images, the first state-of-the-art SfM algorithm is used to compute exterior orientation parameters of images. Thus, the key points are detected and matched using Scale Invariant Feature Transform (SIFT) method. These matched points are used in a sequential bundle adjustment to determine the exterior orientation of input images and then to generate a sparse 3D point cloud (Truong et al. 2017).

Second, the semi-global matching (SGM) algorithm is applied to high-resolution thermal images for generating a disparity map (Hirschmüller 2011). SGM estimates a dense disparity map from a rectified stereo image pairs. In this method, the stereo matching problem is formulated to determine the disparity image D for energy $E(D)$ minimization using Equation (7).

$$E(D) = \sum_p (C(p, D_p)) + \sum_{q \in N_p} P_1 T[|D_p - D_q| = 1] + \sum_{q \in N_p} P_2 T[|D_p - D_q| > 1] \quad (7)$$

where the first term is sum of the matching costs of all pixels for the disparity values in D . In the second term, the P_1 parameter is a constant penalty that is added to all pixels q in the neighbourhood N_p of p , for which the disparity varies a little (i.e. 1 pixel). The third term in all greater disparity differences adds a larger constant penalty P_2 . Satisfactory outcomes of the SGM have made it appropriate for dense stereo matching applications and thus encouraged many researchers to use it (Hirschmüller 2011). Then, consistency check and peak removal are used to eliminate errors and blunders caused by conditions like spectral discontinuities and hidden areas from the generated disparity maps (Mohammadi et al. 2019). Nevertheless, some blunders still stay in purified disparity maps due to the incompleteness of these two algorithms (Mohammadi et al. 2019).

Afterwards, the generated disparity maps from all the stereo pairs of images and computed exterior orientation parameters are utilized to generate a dense point cloud of the scene by applying space intersection. After generating a dense point cloud, the intensified 3D model is produced by data gridding.

Finally, thermal image DN values from the high-resolution thermal images are mapped onto the intensified 3D model for producing an intensified 3D thermal model. It should be noted that although the DNs do not refer to absolute temperature in the thermal image, their differences indicate relative temperature differences. Darker radiometric information infers less heat reflection in the area than in areas with lighter radiometric information.

2.2. Quality assessment

In this study, by considering the fact that the produced intensified 3D thermal model consists of two elements of intensified 3D model and the texture (orthophoto) which is mapped onto the intensified 3D model; both the 3D model and the orthophoto are evaluated separately.

First, the intensified 3D model is evaluated using two statistical criteria, i.e. the Root Mean Square Error (RMSE) and Mean Relative Error (MRE), to compare the low-resolution 3D model (3D model generated from original thermal images) and the intensified

3D model against the reference 3D model. Here, the reference 3D model is considered as R and the generated 3D model from thermal images as G . Equations (8) and (9) represent RMSE and MRE, respectively.

$$RMSE(G, R) = \sqrt{\frac{1}{N} \sum_{i=1, j=1}^{i=n, j=n} (G_{ij} - R_{ij})^2} \quad (8)$$

$$MRE(G, R) = \frac{1}{N} \sum_{i=1, j=1}^{i=n, j=n} \frac{|G_{ij} - R_{ij}|}{R_{ij}} \quad (9)$$

where N is the total number of pixels in 3D model, and (i, j) represents the position at i^{th} row and j^{th} column. The lower values for RMSE and MRE indicate more similarity of the produced 3D model to the reference 3D model (Zhang et al. 2019). It should be noted that due to the use of control points in the 3D model production process, the generated and reference 3D models are registered planimetrically and vertically; only the Z accuracy is checked using the RMSE and MRE criteria.

The criteria such as RMSE and MRE cannot examine the geometric quality of produced 3D thermal models. Considering the importance of edge information in 3D thermal model quality, especially in building boundaries, the Modulation Transfer Function (MTF) as an edge-based quality metric is used for evaluating the geometric quality of generated 3D thermal model. The main idea of this method is to extract the appropriate edges in the reference and produce models and compare their MTF curves based on the Line Spread Function (LSF).

To calculate MTF, first, the high-contrast edges should be extracted to evaluate the spatial resolution of the produced models. Therefore, step edges, which are defined using Equation (10), are considered as appropriate targets for evaluation.

$$step(x) = \begin{cases} 1 & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases} \quad (10)$$

After extracting edge locations, edge profiles are extracted. For each point on the edge, straight lines are created perpendicular to the edge and crossing the edge. The resulted Edge Spread Function (ESF) is introduced to the LSF calculation after smoothing and consistency checks. This is done to avoid noise in final MTF and prevent the algorithm from selecting weak or unsteady edges (Javan et al. 2013). To calculate LSF, differentiation is applied to the ESF profile. Next, the generated LSF curve is smoothed and its noises are removed by a Gaussian function. Equation (11) shows the LSF calculation equation:

$$LSF(x) = \frac{d}{dx}[ESF(x)] \quad (11)$$

Discrete Fourier transform from the generated LSF function leads to MTF (Equation (12)). The normalized MTF is computed by dividing the absolute transformed function values by the first absolute value.

$$MTF = \left| \int LSF(x)e^{2\pi ix} dx \right| \quad (12)$$

After calculating MTF for the extracted edges in the reference and produced models, the model quality assessment is performed by comparing the calculated MTF curves. This assessment is based on the idea that any edge in the reference model should appear in the produced models with similar MTF curves. The lower degradation of the MTF curve



Figure 4. Thermal camera (a) RGB camera (b) and UAV platform (c) used in study.

shows better spatial quality of the produced models. For the numerical comparison of MTF curves, the statistical variance index is used. In this way, it is possible to measure the distance between MTF curves using the edge-based image fusion metric (EFM) introduced in (Javan et al. 2013) (Equation(13)):

$$V_i = MTF_i^{generated} - MTF_i^{reference}$$

$$Var = \frac{1}{N} \sum_{i=1}^N (V_i - \bar{V})^2 \quad \text{Where} \quad \bar{V} = \sum_{i=1}^N V_i / N \quad (13)$$

$$EFM = 1 - Var$$

where MTF_i denotes the MTF value at spatial frequency i . \bar{V} is the mean of the variable V_i . Also, N is the total number of spatial sample frequencies. To be compatible with other measures, the EFM will be defined in such a way that a higher value indicates less distinction between the produced and the reference models, suggesting a higher spatial quality and a higher similarity.

2.3. Dataset

The study is carried out in Charmshahr industrial area in the southern part of Tehran, Iran.

2.3.1. Sensor and platform

In this study, thermal videos are captured using an uncooled focal plane array camera Keii HL-640S mounted on a lightweight multi-rotor Unmanned Aerial Vehicle (UAV) that has the roll and pitch axis stabilizer. This camera detects the Thermal InfraRed (TIR) region of InfraRed (IR), which covers middle and longwave part of IR spectrum. Moreover, for generating the reference 3D model a high-resolution RGB Sony a6000 24 MP camera equipped with a 35 mm lens is used to collect RGB images. These RGB images are only used to generate a reference 3D model and evaluate the enhanced spatial resolution of the intensified 3D model. Figure 4 shows the thermal and RGB cameras as well as UAV platform used in this study.

More details about the sensors and platform are provided in Table 1.

2.3.2. Study area and flight plan

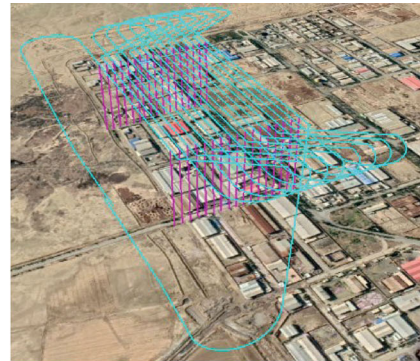
Flight planning for the Charmshahr region with the area of approximately 123,000 m² is done using Mission planner as a UAV ground control station software. Figure 5 shows the studied area and the flight plan. Table 2 provides some information over the flight plan.

Table 1. The technical specifications of applied UAV Platform and sensors.

Thermal camera		RGB camera		UAV platform	
Sensor	Uncooled FPA	Sensor	CMOS	Maximum flight altitude (m)	400
Pixel size (μm)	17	Pixel size(μm)	4.04	Number of motors	8
Video frame size (pixels)	640*480	Image size (pixels)	6000*4000	Maximum flight time without camera (minutes)	35
Focal length (mm)	25	Focal length (mm)	35		
Thermal sensitivity ($^{\circ}\text{C}$)	0.05				
Temperature range ($^{\circ}\text{C}$)	-20 to +150				



(a)



(b)

Figure 5. Study area (a) and the flight plan (b).**Table 2.** Flight plan information.

Parameter	Value
Flight height (m)	300
Flight speed(km/h)	60
Flight duration (minutes)	20
Number of strips	17
Overlap (%)	70
Sidelap (%)	70
Total extractable frames	12000

3. Results and discussion

3.1. Results

As mentioned in the methodology, for generating an intensified 3D thermal model, using the proposed algorithm, the spatial and radiometric resolution of the images is enhanced. In the training phase, 620 frames are extracted from the thermal video are used. 70% of these images (434 images) are used for the training process. The accuracy of the determined trained model is validated by other 186 images and the peak signal-to-noise ratio (PSNR)(Zhang et al. 2019) and structural similarity index (SSIM) (Zhang et al. 2019) values are obtained as 36.72 and 0.9401, respectively.

After the training model generation, each all of another the 669 original thermal images extracted from the video are separately converted into high-resolution thermal



Figure 6. Overall view of some samples for visual comparison between original images extracted from the thermal video and their corresponding high-resolution images.

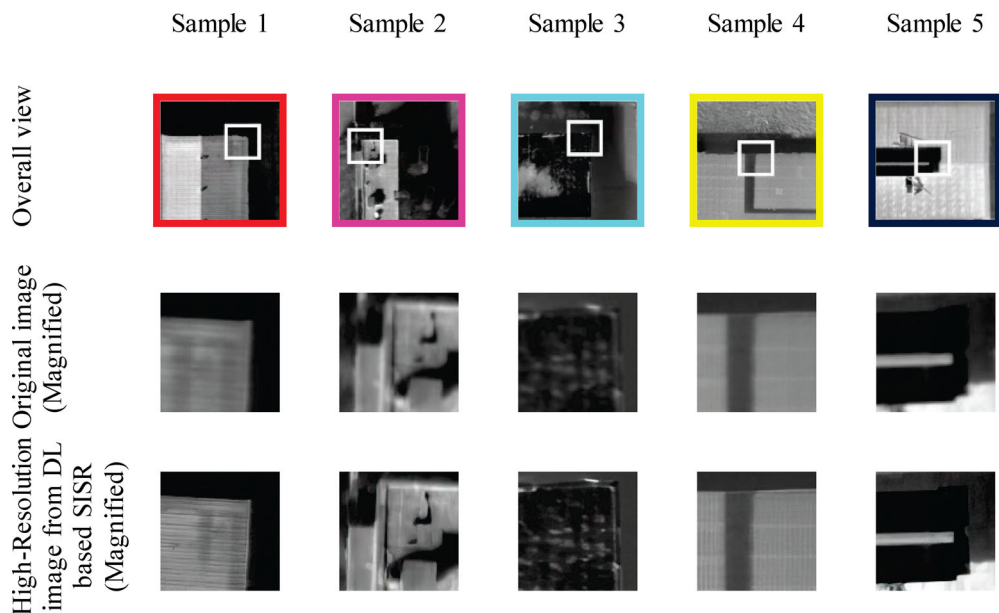


Figure 7. visual comparison between original images and their corresponding generated high-resolution images.

images using the trained model and scale factor 2. The size of the original images is 640×480 pixels and their pixel size is $17 \mu\text{m}$. The size of the enhanced images is 1280×960 and their pixel size is $8.5 \mu\text{m}$. These enhanced images are used to generate the intensified 3D thermal model in the next steps.

Figure 6 illustrates the overall view of some samples for visual comparison between original images extracted from the thermal video and their corresponding high-resolution images resulted from DL-based SISR.

In Figure 7, for each sample of Figure 6, a magnified view of the original image and its corresponding high-resolution image is compared.

After image resolution enhancement step, the resolution-enhanced images are used to generate intensified 3D thermal model of about 11 cm resolution. Also, the low-resolution 3D thermal model (generated from original thermal images) has a resolution of 22 cm.

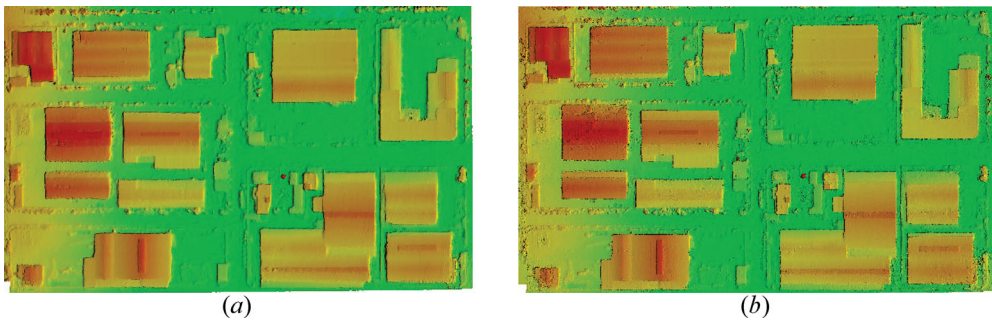


Figure 8. Generated 3D models. Low-resolution 3D model (a) and intensified 3D model (b).

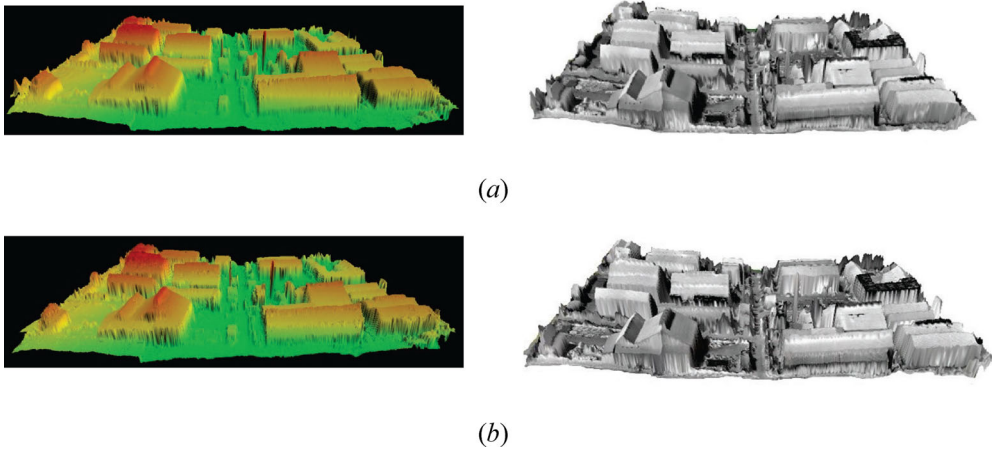


Figure 9. 3D view of generated 3D models. Low-resolution 3D model and its 3D thermal model(a) and intensified 3D model and its 3D thermal model(b).

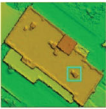
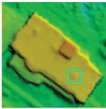
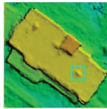


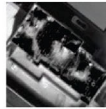
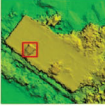
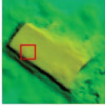
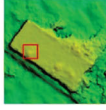



	3D Model			Orthophoto		
	Reference	Low-resolution	Intensified	Reference	Low-resolution	Intensified
Object 1						
Object 2						

Figure 10. Visual comparison of objects details in low-resolution and intensified 3D models and in low-resolution and intensified orthophoto.

Figure 8 presents the low-resolution 3D model and intensified low-resolution 3D model. Also, 3D views of these two 3D models and their corresponding 3D thermal models are represented in Figure 9.

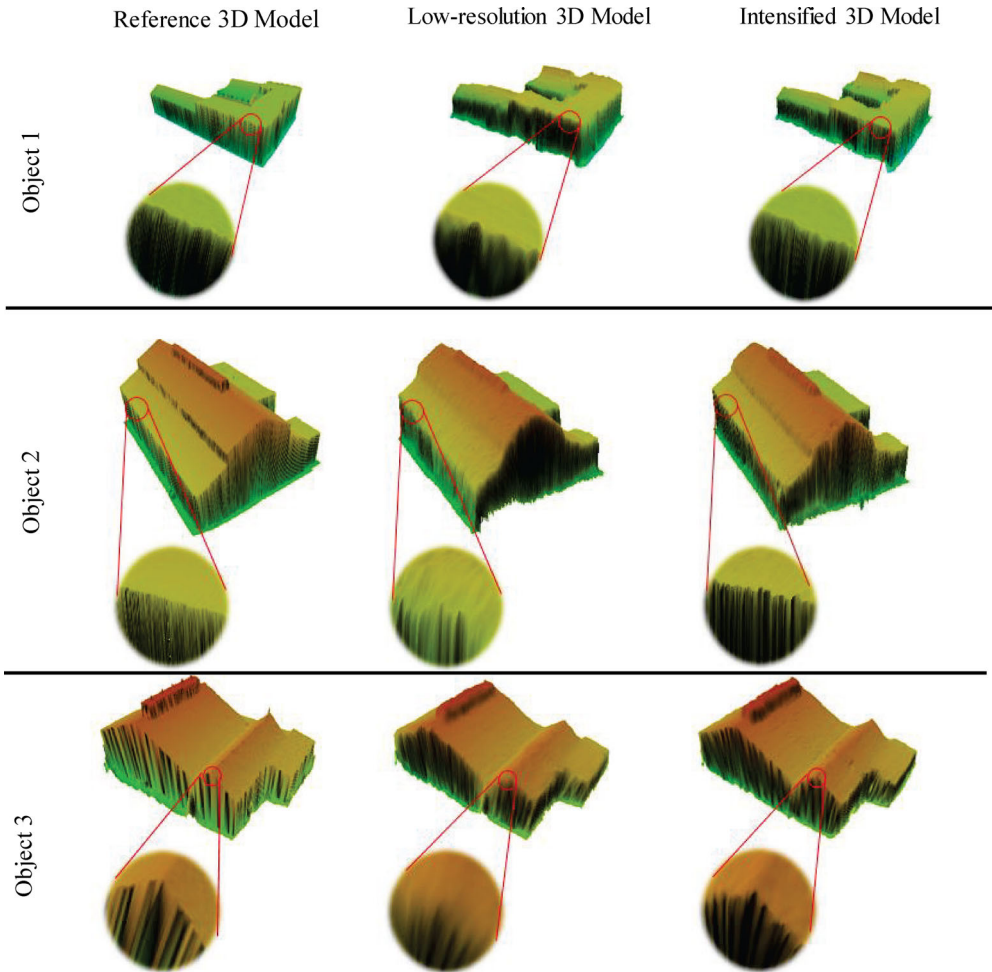


Figure 11. Visual comparison of 3D view of objects edges.

3.2. Discussion

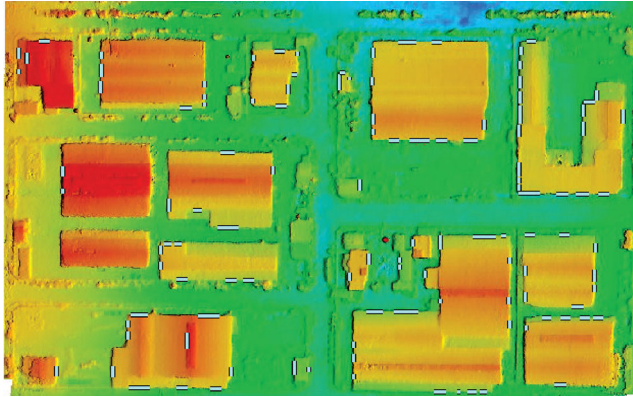
In this section, the quality of the low-resolution and intensified 3D thermal models is compared against the reference 3D model by resolution of four cm which is generated from RGB images captured using a high-resolution RGB camera.

To compare low-resolution 3D thermal model and intensified 3D thermal model against the reference 3D model, the produced 3D thermal models are interpolated to the same resolution as the reference 3D model by the bi-cubic interpolation method. [Figure 10](#) shows a visual comparison of low-resolution 3D model and intensified 3D model against the reference 3D model and their corresponding orthophotos. [Figure 11](#) shows a visual comparison of 3D view of object edges of low-resolution 3D model and intensified 3D model. The visual comparison illustrates that objects in intensified 3D model provide more details than the low-resolution 3D model. Moreover, the edges on the intensified 3D model are sharper than those of the low-resolution 3D model.

In [Figure 10](#), two objects in 3D models and orthophotos are visually compared. As can be seen, the intensified 3D model has more details. In the highlighted regions, small objects in the intensified 3D model are more distinctive than the low-resolution 3D

Table 3. RMSE and MRE metrics calculated for low-resolution 3D model and intensified 3D model.

Metric	Low-resolution 3D model	Intensified 3D model
RMSE(m)	0.55	0.29
MRE (%)	0.092	0.062

**Figure 12.** Suitable edges to MTF curves and EFM metric calculation extracted from 3D models.

model, which is due to the higher quality of spatially enhanced images using SISR. On the other hand, it is clear that the intensified 3D model, compared to the reference 3D model, still has shortcomings in the details and quality of the edges. Although the spatial resolution of the reference 3D model is three times better, Shermeyer and Van Etten (2019) showed that even if the spatial resolution of the super-resolved images is improved by a higher scale factor, SISR methods cannot achieve the quality of original images at the same spatial resolution. In addition, these studies demonstrate that increasing the scale factor to produce spatially enhanced images can reduce the quality of the results against lower scale factors and can cause artificial structures in the outputs, which is not acceptable.

In [Figure 11](#), three objects are shown in the 3D view to compare the sharpness of the edges visually. For each object, a magnified view is represented. The edges of the objects are obviously sharper in the intensified 3D model than low-resolution 3D model. However, as before, there are still shortcomings compared to the reference 3D model.

In addition to the visual evaluation, quantitative assessment of produced 3D models is of particular importance. As mentioned before, the 3D thermal model is a 3D model, onto which the texture (orthophoto) is mapped, so in this part, the spatial accuracy of the intensified 3D model and the orthophoto is evaluated separately. To measure the effect of SISR on the quality of the intensified 3D model, the RMSE and MRE metrics once are calculated between low-resolution 3D model and reference 3D model and again between the intensified 3D model and reference 3D model ([Table 3](#)).

Besides, considering the importance of sharpness of the edges in 3D models, the quality of edges is evaluated based on MTF curves and EFM metric. For this purpose, proper edges are extracted from 3D models using Canny edge extraction operators. Then, the line segments are detected from the edges using the direct Hough transform (Mukhopadhyay and Chaudhuri 2015). Finally, those that are too long, too short, or slanted are eliminated by calculating the length and slope of the extracted lines. [Figure 12](#) shows the suitable edges to MTF curves and EFM metric calculation extracted from 3D model. The mean values of the EFM metric for all the extracted edges are 0.928 and

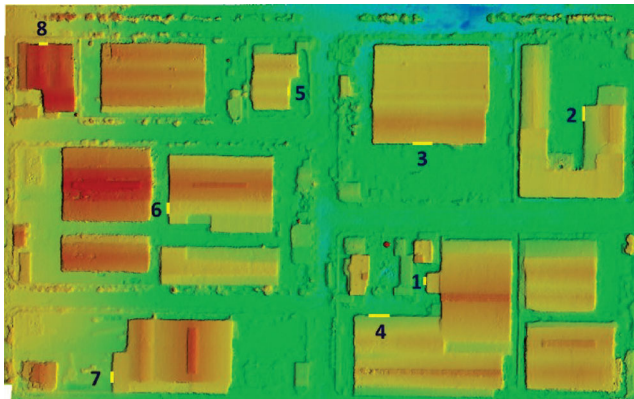


Figure 13. Overall view of 8 selected samples in study area for MTF and EFM analysis.

0.979 for low-resolution and intensified 3D models, respectively. The higher mean value of EFM metric for the intensified 3D model than that of low-resolution 3D model shows the closeness of the intensified 3D model to the reference 3D model.

For more analysis, eight edges with suitable distribution in the study area are selected for evaluation. The overall view of these samples is provided in Figure 13.

The edges in Figure 13 are magnified and compared in Figure 14. Each edge in the reference 3D model, Low-resolution 3D model, and intensified 3D model is shown and the MTF curves are presented.

As can be seen in Figure 14, the MTF curves of the intensified 3D model are closer to the MTF curves of the reference 3D model compared to the low-resolution 3D model and, suggesting that the proposed method improves the geometric quality of the 3D model. Moreover, it is obvious that although the edges are spatially enhanced in the intensified 3D model, there are still some distortions on the edges.

Finally, the EFM metric is computed to compare MTF curves with respect to the reference 3D model (Table 4).

As can be seen in Table 4, the results for the intensified 3D model that are generated based on the proposed method have higher closeness values, indicating the higher spatial quality of the intensified 3D model than the low-resolution 3D model.

Also, for evaluating the spatial accuracy of the orthophotos generated from high-resolution thermal images, the MTF curve is plotted for each selected edge from 3D model in low-resolution, enhanced and reference orthophotos. Then, the average MTF curves are calculated. As can be seen in Figure 15, MTF curve of the enhanced orthophoto has lower degradation; therefore, the spatial resolution of the enhanced orthophoto is better than the low-resolution orthophoto.

It is noteworthy that if the interpolation-based SISR methods (such as bi-cubic) are used instead of the EDSR to enhance the resolution of the images, although a dense 3D model is produced, the accuracy of the produced 3D model would be lower than the intensified 3D model generated using the proposed method. This is because although this method is simple, the high-frequency details of the image are not restored (Ooi and Ibrahim 2021).

To illustrate this point, a bi-cubic dense 3D model is produced and compared with the intensified 3D model. Therefore, the MTF curve is plotted for 20 edges of the proposed intensified 3D model and bi-cubic dense 3D model. Finally, the average MTF curve is calculated. As shown in Figure 16, the average MTF curve of the intensified 3D model has

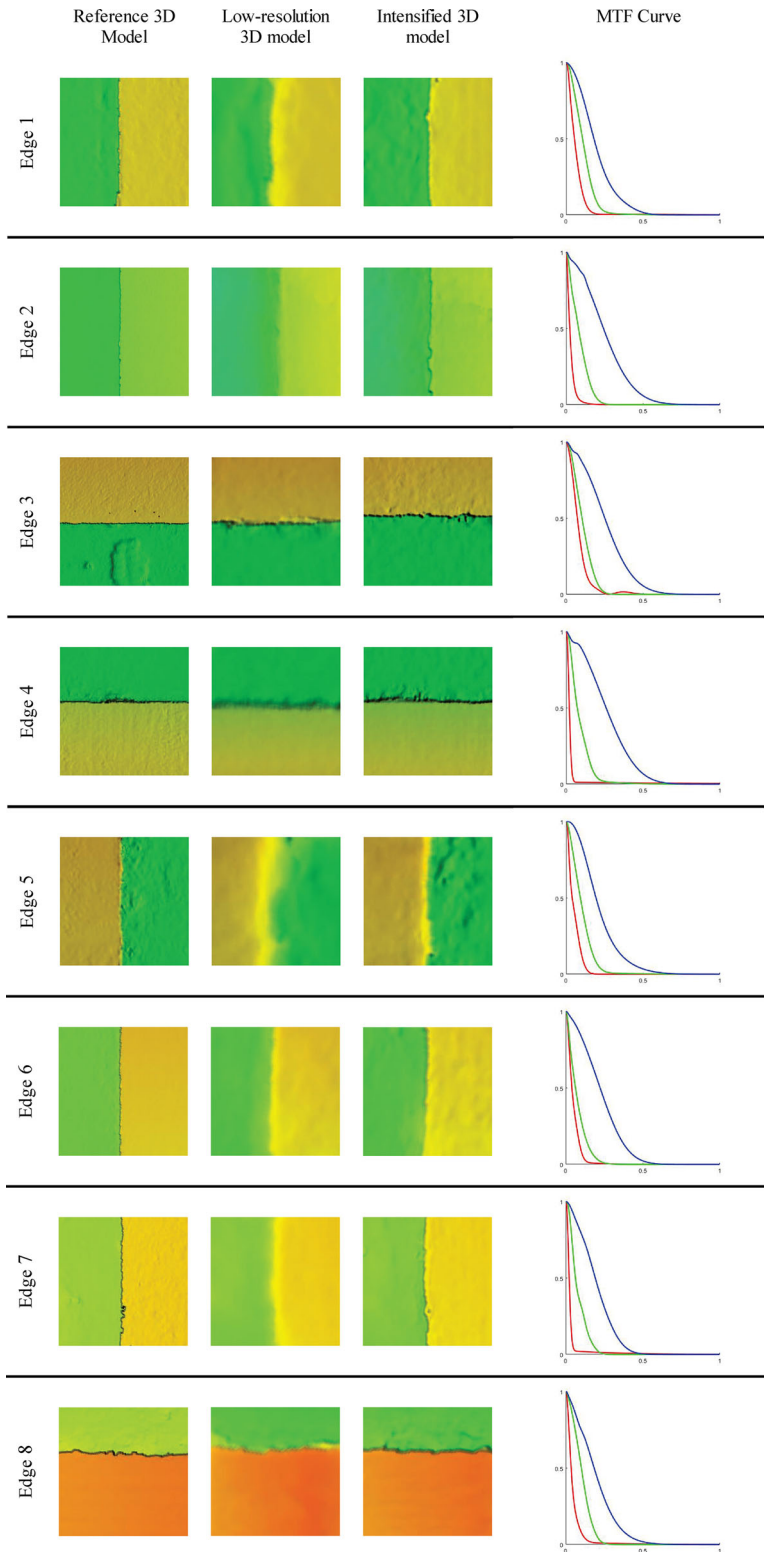


Figure 14. Magnified representation of the selected edges and their MTF curves. In MTF curve diagram, the blue curve represents the MTF of the reference 3D model, the red curve represents the MTF of the Low-resolution 3D model and, the green curve represents the MTF of the intensified 3D model.

Table 4. EFM values

3D model	Edge 1	Edge 2	Edge 3	Edge 4	Edge 5	Edge 6	Edge 7	Edge 8
Low-resolution 3D model	0.959	0.913	0.948	0.901	0.946	0.936	0.926	0.950
Intensified 3D model	0.988	0.963	0.969	0.953	0.981	0.960	0.972	0.987

Closeness of low-resolution 3D model and intensified 3D model to Reference 3D model.

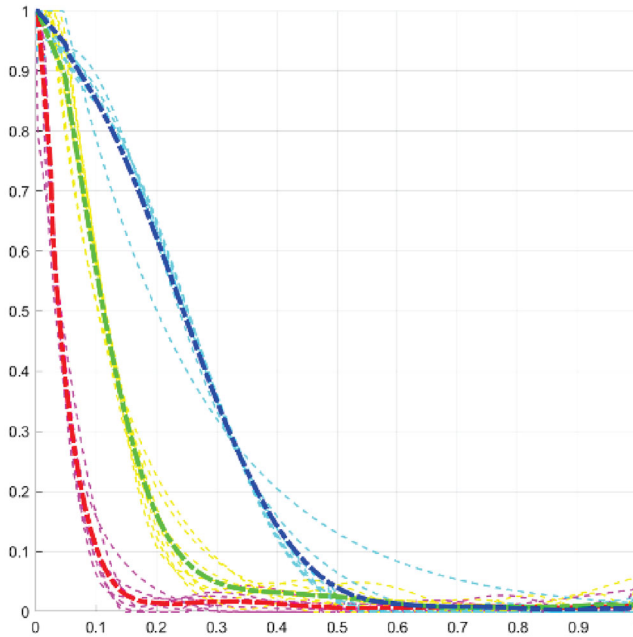


Figure 15. The average MTF curve for selected edges in low-resolution orthophoto (red curve), intensified orthophoto (green curve) and reference orthophoto (blue curve).

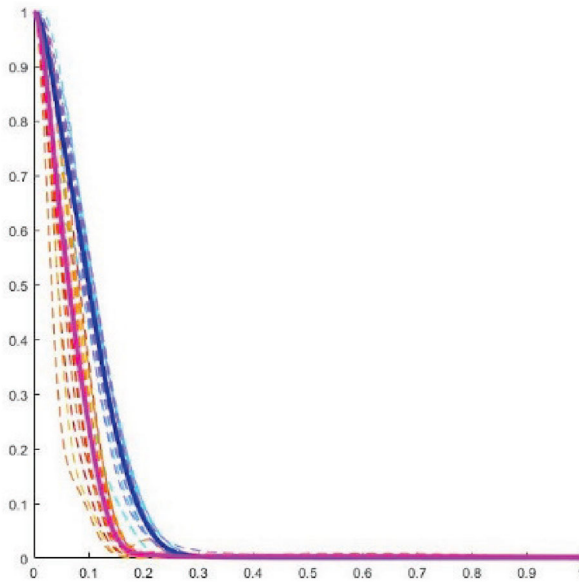


Figure 16. Comparison between the average MTF curves for some edges in intensified 3D model (blue curve) and bi-cubic dense 3D model (pink curve).

lower degradation and the spatial resolution of the intensified 3D model is better than the bi-cubic dense 3D model.

4. Conclusions

Nowadays, 3D thermal model that presents thermal information and provides 3D building roofs information can play an important role in energy management of urban area. By considering the challenges of acquiring multisource data and registering data from various data sources and using hardware techniques to generate a high-resolution 3D thermal model, in this article, a method is presented for intensifying 3D thermal model using SISR algorithm. The effect of SISR algorithm on 3D thermal model intensification is investigated by measuring spatial resolution evaluation criteria.

The efficiency of the proposed method was evaluated from two criteria, i.e. RMSE and MRE, to compare the low-resolution 3D model and intensified 3D model against the reference 3D model. Evaluations indicate 47% improvement in RMSE and 0.03% improvement in MRE for the intensified 3D model that indicates an increase in the quality of the result. In addition, for spatial evaluation, considering the importance of edge information in 3D model quality, especially in building boundaries, the sharpness of the edge information in the intensified 3D model is computed using MTF curves and EFM metric. The results show that the MTF curves of the intensified 3D model are closer to the MTF curve of the reference 3D model. Moreover, the EFM metric shows higher values for MTF curves of the intensified 3D model against MTF curves of the low-resolution 3D model. In addition, investigating MTF curves of orthophotos shows MTF curve of enhanced orthophoto has lower degradation and the spatial resolution of the enhanced orthophoto is better than the low-resolution orthophoto. In addition, although the evaluation shows the efficiency of the proposed method to generate 3D thermal models with higher spatial resolution, further studies should be concentrated on producing better results in the boundaries of the objects. Also, it is necessary to investigate the possible incensement in spatial resolution and the use of different methods of DL-based SISR and their effect on the quality of 3D thermal models. In addition, with the aim of using the intensified 3D thermal model for the interpretation purposes, the radiometric calibration of the thermal sensor is suggested in order to determine the absolute temperature.

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

Alaleh Fallah  <http://orcid.org/0000-0002-1211-0165>

Farhad Samadzadegan  <http://orcid.org/0000-0003-3439-5648>

Farzaneh Dadrass Javan  <http://orcid.org/0000-0002-0910-8160>

References

- Alba MI, Barazzetti L, Scaioni M, Rosina E, Previtali M. 2011. Mapping infrared data on terrestrial laser scanning 3D models of buildings. *Remote Sens.* 3(9):1847–1870.
- Almasri F, Debeir O. 2018. Multimodal sensor fusion in single thermal image super-resolution. Paper presented at: the Asian Conference on Computer Vision.

- Antón D, Amaro-Mellado J-L. 2021. Engineering graphics for thermal assessment: 3D thermal data visualisation based on infrared thermography, GIS and 3D point cloud processing software. *Symmetry*. 13(2):335.
- Borrmann D, Elseberg J, Nüchter A. 2013. Thermal 3D mapping of building façades. In *Intelligent autonomous systems*. Berlin, Heidelberg: Springer; vol. 12; p. 173–82.
- Brown DC. 1971. Close-range camera calibration, photogrammetric engineering. *Eng Remote Sens*. 37(8): 855–866.
- Burdziakowski P. 2020. Increasing the geometrical and interpretation quality of unmanned aerial vehicle photogrammetry products using super-resolution algorithms. *Remote Sens*. 12(5):810.
- Choi Y, Kim N, Hwang S, Kweon IS. 2016. Thermal image enhancement using convolutional neural network. Paper presented at: the 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS).
- Javan FD, Savadkouhi M. 2019. Thermal 3D models enhancement based on integration with visible imagery. *Int Arch Photogramm Remote Sens Spat Inf Sci*. 263–269.
- Dahaghin M, Samadzadegan F, Javan FD. 2021. Precise 3D extraction of building roofs by fusion of UAV-based thermal and visible images. *Int J Remote Sens*. 42(18):7002–7030.
- Dai T, Cai J, Zhang Y, Xia S-T, Zhang L. 2019. Second-order attention network for single image super-resolution. Paper presented at: the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.
- Dong C, Loy CC, He K, Tang X. 2016. Image super-resolution using deep convolutional networks. *IEEE Trans Pattern Anal Mach Intell*. 38(2):295–307.
- Ham Y, Golparvar-Fard M. 2013. An automated vision-based method for rapid 3D energy performance modeling of existing buildings using thermal and digital imagery. *Adv Eng Inform*. 27(3):395–409.
- Hirschmüller H. 2011. Semi-global matching-motivation, developments and applications. *Photogramm Week*. 11:173–184.
- Javadnejad F, Gillins DT, Parrish CE, Slocum RK. 2020. 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.
- Javan FD, Samadzadegan F, Reinartz P. 2013. Spatial quality assessment of pan-sharpened high resolution satellite imagery based on an automatically estimated edge based metric. *Remote Sens*. 5(12): 6539–6559.
- Kansal P, Nathan S. 2020. A multi-level supervision model: a novel approach for thermal image super resolution. Paper presented at: the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops.
- Khodaei B, Samadzadegan F, Javan FD, Hasani H. 2015. 3D surface generation from aerial thermal imagery. *Int Arch Photogramm Remote Sens Spatial Inf Sci*. XL-1/W5(1):401–405.
- Kim J, Lee JK, Lee KM. 2016. Accurate image super-resolution using very deep convolutional networks. Paper presented at: the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.
- Kim KI, Kwon Y. 2010. Single-image super-resolution using sparse regression and natural image prior. *IEEE Trans Pattern Anal Mach Intell*. 32(6):1127–1133.
- Liebel L, Körner M. 2016. Single-image super resolution for multispectral remote sensing data using convolutional neural networks. *Int Arch Photogramm Remote Sens Spatial Inf Sci*. XLI-B3:883–890.
- Lim B, Son S, Kim H, Nah S, Lee KM. 2017. Enhanced deep residual networks for single image super-resolution. Paper presented at: the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops.
- Ma J, Ma Y, Li C. 2019. Infrared and visible image fusion methods and applications: a survey. *Inf Fusion*. 45:153–178.
- Mandanici E, Tavasci L, Corsini F, Gandolfi S. 2019. A multi-image super-resolution algorithm applied to thermal imagery. *Appl Geomat*. 11(3):215–228.
- Maset E, Fusiello A, Crosilla F, Toldo R, Zorzetto D. 2017. Photogrammetric 3D building reconstruction from thermal images. *ISPRS Ann Photogramm Remote Sens.Spat Inf Sci*. 4:25.
- Mohammadi H, Samadzadegan F, Reinartz P. 2019. 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.
- Mukhopadhyay P, Chaudhuri BB. 2015. A survey of Hough transform. *Pattern Recognit*. 48(3):993–1010.
- Ooi YK, Ibrahim H. 2021. Deep learning algorithms for single image super-resolution: a systematic review. *Electronics*. 10(7):867.

- Oreifej O, Cramer J, Zakhor A. 2014. Automatic generation of 3d thermal maps of building interiors. *ASHRAE Trans.* 120(2)
- Ouellet J-N, Hébert P. 2009. Precise ellipse estimation without contour point extraction. *Mach Vis Appl.* 21(1):59–67.
- Pashaie M, Starek MJ, Kamangir H, Berryhill J. 2020. Deep learning-based single image super-resolution: an investigation for dense scene reconstruction with UAS photogrammetry. *Remote Sens.* 12(11):1757.
- Peng E, Li L. 2010. Camera calibration using one-dimensional information and its applications in both controlled and uncontrolled environments. *Pattern Recognit.* 43(3):1188–1198.
- Truong TP, Yamaguchi M, Mori S, Nozick V, Saito H. 2017. Registration of RGB and thermal point clouds generated by structure from motion. Paper presented at: the Proceedings of the IEEE International Conference on Computer Vision Workshops.
- Prakash CD, Karam LJ. 2012. Camera calibration using adaptive segmentation and ellipse fitting for localizing control points. Paper presented at: the 2012 19th IEEE International Conference on Image Processing.
- Rakha T, Liberty A, Gorodetsky A, Kakillioglu B, Velipasalar S. 2018. Heat mapping drones: an autonomous computer-vision-based procedure for building envelope inspection using unmanned aerial systems (UAS). *Technology|Architecture + Design.* 2(1):30–44.
- Rivadeneira RE, Suárez PL, Sappa AD, Vintimilla BX. 2019. Thermal image superresolution through deep convolutional neural network. Paper presented at: the International Conference on Image Analysis and Recognition.
- Shermeyer J, Van Etten A. 2019. The effects of super-resolution on object detection performance in satellite imagery. Paper presented at: the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops.
- Usamentiaga R, Garcia DF, Ibarra-Castanedo C, Maldague X. 2017. Highly accurate geometric calibration for infrared cameras using inexpensive calibration targets. *Measurement.* 112:105–116.
- Yan K, Tian H, Liu E, Zhao R, Hong Y, Zuo D. 2016. A decoupled calibration method for camera intrinsic parameters and distortion coefficients. *Math Probl Eng.* 2016:1–12.
- Yang M-D, Su T-C, Lin H-Y. 2018. Fusion of infrared thermal image and visible image for 3D thermal model reconstruction using smartphone sensors. *Sensors.* 18(7):2003.
- Yang W, Zhang X, Tian Y, Wang W, Xue J-H, Liao Q. 2019. Deep learning for single image super-resolution: a brief review. *IEEE Trans Multimedia.* 21(12):3106–3121.
- Yue L, Shen H, Li J, Yuan Q, Zhang H, Zhang L. 2016. Image super-resolution: the techniques, applications, and future. *Signal Process.* 128:389–408.
- Zhang K, Zuo W, Zhang L. 2018. Learning a single convolutional super-resolution network for multiple degradations. Paper presented at: the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.
- Zhang Y, Zheng Z, Luo Y, Zhang Y, Wu J, Peng Z. 2019. A CNN-based subpixel level DSM generation approach via single image super-resolution. *Photogramm Eng Remote Sens.* 85(10):765–775.
- Zhang Z. 2000. A flexible new technique for camera calibration. *IEEE Trans Pattern Anal Machine Intell.* 22(11):1330–1334.