Image-based construction of building energy models using computer vision

Citation for published version:
Dino, IG, Sari, AE, Iseri, OK, Akin, S, Kalfaoglu, E, Erdogan, B, Kalkan, S & Alatan, A 2020, 'Image-based construction of building energy models using computer vision', *Automation in Construction*, vol. 116, 103231. https://doi.org/10.1016/j.autcon.2020.103231

Digital Object Identifier (DOI):
10.1016/j.autcon.2020.103231

Published In:
Automation in Construction

Image-based construction of building energy models using computer vision

Ipek Gursel Dino[1,2] (ipekg@metu.edu.tr), Tel: + 90-312-210-2262, Corresponding Author

Alp Eren Sarı[1] (sari.eren@metu.edu.tr)

Orcun Koral Iseri[1,2] (koral.iseri@metu.edu.tr)

Sahin Akin[1,2] (sahin.akin@metu.edu.tr)

Esat Kalfaoglu[1] (esat.kalfaoglu@metu.edu.tr)

Bilge Erdogan[3] (b.erdogan@hw.ac.uk)

Sinan Kalkan[1] (skalkan@metu.edu.tr)

A. Aydın Alatan[1] (alatan@metu.edu.tr)

[1] Center for Image Analysis (OGAM), Middle East Technical University, Dumlupınar Bulvarı No:1, 06800 Ankara, Turkey.

[2] Department of Architecture, Middle East Technical University, Dumlupınar Bulvarı No:1, 06800 Ankara, Turkey

[3] Heriot-Watt University, School of Energy, Geoscience, Infrastructure and Society, Institute for Sustainable Building Design, Edinburgh, Scotland EH14 4AS

Abstract (100-150 words)

Improving existing buildings' energy performance requires energy models that accurately represent the building. Computer vision methods, particularly image-based 3D reconstruction, can effectively support the creation of 3D building models. In this paper, we present an image-based 3D reconstruction pipeline that supports the semi-automated modeling of existing buildings. We developed two methods for the robust estimation of the building planes from a 3D point cloud that (i) independently estimate each plane and (ii) impose a perpendicularity constraint to plane estimation. We also estimate external walls' thermal transmittance values using an infrared thermography-based method, with the surface temperatures measured by a thermal camera. We validate our approach (i) by testing the pipeline's ability in constructing accurate surface models subject to different image sets with varying sizes and levels of image quality, and (ii) through a comparative analysis between the calculated energy performance metrics of a ground truth and calculated energy simulation model.

Keywords: 3D modeling; building energy modeling; computer vision; 3D reconstruction; infrared imaging

1 Introduction

Existing buildings are increasingly being placed into focus in the Architecture, Engineering, Construction, and Facility Management (AEC/FM) industry, due to their great potential for performance improvement and meaningful environmental impact. AEC/FM applications on existing buildings typically require 3D models that represent the precise, as-is conditions. Such models can support activities regarding a wide range of areas including safety / health assessment, space planning, procurement, cost estimation, life cycle assessment, sustainability assessment, performance monitoring, operations and maintenance, scheduling and retrofit / refurbishment / renovation planning [1]. In the literature, the benefits of rapid assessment of as-built building conditions are reported to enhance the efficiency of building and maintenance operations [2]. The effective and timely execution of AEC/FM tasks, specifically those that target increased energy performance, call for approaches that precisely model the existing buildings and quantify building performance through simulation tools. Energy simulation has the potential to reduce buildings' environmental impact, improve occupant comfort and indoor environmental quality and facilitate innovation in AEC [3]. For existing buildings, energy simulations can also complement real monitored building data for operational optimization or retrofit. A simulation-based virtual model has the capacity to analyze the building's past behavior to calibrate the program for improved predictive potency, or predict the building's response to alternative scenarios [4]. Dynamic energy simulation tools adopt a forward-modeling approach that begins with a description of the building and components, providing a physical description of the building (design geometry, thermal characteristics of the building envelope, internal heat gains, infiltration and occupancy profiles), its systems (system types and sizes, control schedules, outdoor air requirements) and components (HVAC components) [5]. Amongst these, the former is the most fundamental category that the other categories are based upon. Therefore, a correct description of the building form and envelope thermal properties is critical for the reliability of simulation-based performance assessment.

Despite the key role of simulations in performance assessment, the difficulties in the construction of energy models has been a major obstacle against their widespread use. Manual modeling based on building documentation (i.e. drawings, specifications, schedules) and walk-through audits have been proven to be labor-intensive and difficult, mainly due to missing or outdated building documentation (i.e. drawings or models), and the process's proneness to imprecision and error [6]. This is also due to buildings undergoing undocumented major changes and the degradation of materials and building systems over time. As a result, the rapid acquisition of spatial information and the automation of modeling appear as key factors in the effective and timely execution of AEC/FM tasks.

ICT techniques can help automate building modelling. Remote sensing technologies such as Light Detection and Ranging (LiDAR) are widely used to obtain 3D point clouds of target geometries based on distance measurements. LiDAR can be used for planning retrofit, spatial planning, resource and construction progress tracking [7–13]. Despite the volume of existing research on automated as-built generation, these steps still largely remain as semi-automated and labor-intensive processes that involve human labor to various degrees [14]. Moreover, LiDAR-based approaches' high cost, high level of operational expertise and complexity prevent their widespread use.

Computer vision methods, especially 3D reconstruction techniques, offer potentials in the data acquisition and modeling of building or component geometry in unstructured physical environments. 3D reconstruction from multiple views mainly relies on finding the "projections" (or occurrences) of 3D scene points in 2D multiple views [15]. Given its 2D projections in different views, the source 3D scene point can be estimated by simple geometric calculations. In buildings, image-based reconstruction techniques are reported to be advantageous over LiDAR for their low cost of technology implementation and data collection, but at the same time are expensive in data processing for 3D point cloud generation [16]. Another difference between the two methods is the sparsity of the generated point cloud models in image-based techniques. However, image-based reconstruction techniques have proven to be a robust alternative to LiDAR, and recently resulted in models that are comparable to LiDAR in their accuracy

[17].

The AEC/FM applications that use computer vision methods primarily focus on the data capturing and modeling of existing built forms, monitoring and progress tracking. Dimitrov and Goldparvar-Fard [18] developed a vision-based approach based on structure-from-motion, multi-view stereo, together with a voxel coloring algorithm to generate a volumetric reconstruction of a construction site and detect progress using 4D as-planned BIM models. Kim et al. [19] apply a data-driven scene-parsing method that recognizes construction objects classes in images. Park et al. [20] develop a content-based image retrieval approach for the automated registration of photos to 4D BIM and identification of BIM objects for construction project management. Tang et al. [21] propose an automatic method for reconstructing semantically rich indoor 3D building models including components of indoor environments such as space, wall, floor, ceilings, windows, and doors from low-quality RGB-D sequences. Brillakis et al. [22] propose a videogrammetric framework for acquiring spatial data of infrastructure using low-cost high resolution video cameras that are traversed around a scene to produce a dense 3D point cloud.

Infrared thermal imaging is a viable, non-destructive technique for fast and accurate building diagnostics and material characterization. Typical building thermal performance assessment practices make use of IR testing to detect problems of heat losses, thermal bridges, air leakage and moisture sources and missing / damaged thermal insulation [23]. Since the identification of problems requires the manual and simultaneous interpretation of infrared (IR) images and RGB images, this process is limited in applicability due to its dependence on human expertise in combining IR and RGB images [24]. Image fusion, therefore, is used to combine multiple input images of the same object into a composite image that contains critical thermal information. The fusion of thermal and visual images for building 3D modelling has been addressed in the previous literature in AEC/FM. Yang et al. [25] propose a method for thermal model reconstruction from thermal and RGB images, which builds a 3D mesh model with surface temperature values. González-Aguilera et al. [26] develop an approach for the automatic registration of infrared (IR) images and 3D-laser scanner models and the combination of thermographic and geometric data in a thermographic 3D-model to locate thermal defects and quantify heat losses. Ribaric et al. [24] propose a knowledge-based system that uses the fusion of information extracted from low-resolution IR images and high-resolution visual RGB images for non-destructive testing and façade diagnostics. Merchán et al. [27] improve the existing modelling techniques that generate colored 3D models of outdoor scenes by decoupling the color integration and geometry reconstruction stages. The development of non-destructive, non-contact techniques for architectural heritage conservation is addressed by Costanzo et al. through the fusion of terrestrial laser scanning and the infrared thermal images [28]. Lagüela et al. [29] combine geometric information with thermal data using a new procedure that builds thermographic 3D models by the fusion of infrared mosaics and visible images. Adan et al. [30] improve the accuracy and soundness of existing approaches in a system that fuses information from 3D laser scanners, RGB cameras and thermal cameras to generate dense thermal 3D point clouds. Schramm et al. [31] reduce the impact of correspondence problems when scanning objects with few geometric features by developing an imaging system consisting of an IR camera and a near-infrared (NIR) depth sensor. Ham and Golparvar-Fard [32] develop a vision-based approach for 3D spatio-thermal modeling by automatically generating and superimposing the 3D building and thermal point clouds to build 3D energy performance models.

The automated recognition of relevant objects in a 3D scene and extracting useful semantic information using machine learning methods is also an important research direction. The image-based detection and classification of building materials has been addressed previously using various methods including Support Vector Machine classifiers [12], Multilayer Perceptron (MLP), Radial Basis Function (RBF), and Support Vector Machines (SVM) [33] and Neuro-Fuzzy systems (NFS) [34]. Similarly, the automated detection of building components can greatly ease the generation of as-is models of existing buildings. Xiong et al. develop a method to automatically convert raw 3D point data, which can also identify and model the main structural components of an indoor environment (i.e. walls, floors, windows) by point cloud voxelization, planar patch extraction and the labeling of patches as building elements [8]. Wang et al. follow a similar approach to plane segmentation, an edge and boundary detection algorithm, followed by a rule-based building envelope component classification system [35]. The resulting data is converted to gbXML to process further in energy simulation applications. Valero et al. propose a Terrestrial Laser Scanning data-processing pipeline that builds semantic 3D models of building interiors using Boundary Representation (B-Rep) models, which recognizes openings by detecting the moldings around empty areas in 3D points [36]. Liu et al. propose a method for the remote monitoring of external cladding using classical operators and fuzzy logic algorithms [37]. The inspection of steel frame manufacturing is addressed by Martinez et al. using a vision-based framework and a knowledge-based decision-making system [38]. Automating the acquisition, updating and management of knowledge in historic / heritage buildings is also an active research area that is in need of novel methods to capture historic building elements in a high level of detail [39]. Automated methods for construction project monitoring and the detection of construction equipment / workers is addressed using various methods including deep learning and virtual reality [40], vector-quantized histograms for material classification [33], vision-based algorithms that detect partition components [41], 2D Continuous Wavelet Transform for the automatic segmentation of stones in walls [42], convolutional neural networks to detect workers and heavy equipment [43], vision-based algorithms using as-built video data to recognize activity states of construction activities [44], and convolutional networks and transfer learning to detect construction equipment [45]. Adan et al. focus on the recognition of non-structural components such as MEP components using a consensus strategy for depth-based and color-based recognitions [46].

Critical factors for the successful implementation of image-based 3D reconstruction include usability, reliability, and ease-of-use [47]. While 3D point cloud data extraction from an image collection has been previously addressed widely, obtaining a useful model from point cloud data using fit-to-purpose plane segmentation approaches is still an active research area in AEC/FM. The extraction of a 3D model (surface or wireframe model, BIM or energy model) demands either cumbersome manual work or computational approaches to streamline the modeling process that require different degrees of complexity. Despite the success of the existing work in the automation of modeling, executing tasks on huge, unstructured point clouds require a high level of expertise and experience on software tools to be able to operate on raw geometries, which can hinder the effectiveness of the task. During the generation and processing of large 3D point clouds, approaches tailored for specific purposes are needed to efficiently integrate, update, manage, analyze, and visualize 3D points [48]. Moreover, imaged based methods demonstrate reduced robustness when the scenes do not have sufficiently textured areas, contain excessively repetitive textures, and when lighting varies dramatically across the views [49]. The lack of distinguishable features in indoors (especially on walls or ceilings) needed for image registration is also an obstacle against the reliable convergence of feature-based methods. Another problem encountered in interior spaces is occlusion due to furniture, curtains or other indoor objects. This makes it difficult to clearly identify and extract objects of interest, and reduces the reliability of model construction [50]. Therefore, high levels of tolerance to missing data is required for indoor modeling, as compared to outdoor modeling.

In this paper, we present an image-based 3D reconstruction pipeline that supports the semi-automated energy modeling of existing buildings. The main motivation behind the proposed approach is the pressing necessity for the fast, easy and low-cost method for reliable and precise building energy modeling. As an alternative to high-cost LiDAR data, our approach makes use of unstructured visible (optical) and thermal images of a room captured using readily available cameras. The developed pipeline can generate different models during its successive stages, including point clouds, planar surface models and building energy models. For the latter two models, we developed two methods for the robust estimation of the building planes from the initially generated 3D point cloud. The first method independently estimates each plane using RANdom SAmple Consensus (RANSAC) and singular value decomposition (SVD). The second method builds upon the same method by imposing a plane perpendicularity (or any angular, in general) constraint to plane estimation step to improve geometric precision. Surface planarity, minimum surface complexity and the reduction of unnecessarily intricate geometries are desirable qualities the surface models.

The thermal characterization of opaque external building surfaces in also of key importance for energy simulations in the calculation of conductive heat transfer between the indoor and outdoor environments. As discussed previously, IR imaging can be used for the quantitative characterization of building constructions using the surface temperatures measured by a thermal camera. In our pipeline, we calculate external walls' thermal transmittance using an existing IR thermography–based procedure [51]. In the construction 3D thermal point clouds, we follow a similar approach to Ham and Golparvar-Fard [32], who calculated actual thermal resistance of building assemblies at the level of 3D points and converted it into a single value for each building element. Our pipeline differs in the following: (i) Ham and Golparvar-Fard perform separate 3D reconstruction from the thermal images and visible band images, which, however, is inefficient and unnecessary given that the visible band and thermal cameras are calibrated. In our approach, we perform 3D reconstruction only from the visible band images, and the thermal information is transferred using the transformation between the cameras. (ii) We propose a novel mechanism for integrating structural priors into the pipeline to get more accurate reconstructions even with small number of images.

The main practical advantage of our approach is that users' involvement required for the extraction of architectural elements of an indoor space (e.g., walls, windows, floors and ceilings) is based on simple interaction tasks on 2D images. Specifically, after the construction of a point cloud, architecturally significant elements are identified by the users by partially or fully marking these elements on 2D images. This eliminates the need for complex software tools and human expertise for point cloud editing, and makes our approach a viable, rapid, low-cost and easy-to-use alternative to the existing approaches in the literature.

The rest of this paper is structured as follows. Section 2 presents the development of the proposed pipeline, and the phases of calibration, 3D point cloud generation, 3D room modeling, fusion with thermal data and thermal transmittance value calculation. Section 3 presents and discusses the results of the two experiments conducted for the validation of the developed pipeline. Finally, Section 5 concludes this work and points to some future research directions.

2 Material and methods

In this paper, we propose a methodology for the semi-automated modeling of existing buildings using 3D reconstruction. Specifically, we develop a pipeline that merges digital 2D visible (optical) images and thermal images of a room into a single 3D building model with thermal transmittance values assigned to the external walls. The developed pipeline generates different models including point clouds, planar surface models and energy models. The pipeline uses visible band images registered with the corresponding thermal images as input data. Visible band images are utilized for 3D model generation whereas thermal images are used for obtaining thermal transmittance values of external walls.

Our pipeline is composed of the following main steps (Figure 1):

- The calibration of the imaging system: This step estimates the parameters of the cameras (e.g. focal length, scaling factor or distortion) and the rotation & translation between the two cameras.
- 3D point cloud generation: This step uses multiple views of a room to estimate a set of 3D points of the room. For this, a structure-from-motion technique and a multi-view stereo method are employed.
- 3D model generation: From the sparse set of 3D points generated in the previous step, the walls of the room are calculated in 3D.
- Fusion with thermal data, and thermal resistance calculation: The 3D model of the room is populated with the temperature values from the thermal camera, and with this, the external walls'

thermal resistance values are calculated.

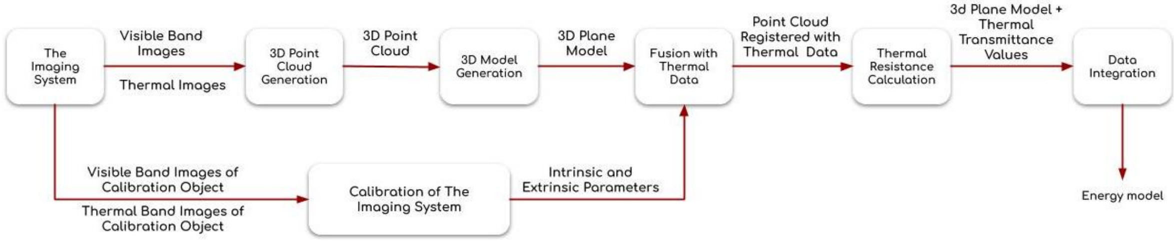


Figure 1 An overview of the proposed pipeline for constructing an energy model of a room from visible band and thermal images.

2.1 The Calibration of the Imaging System

Physical properties of a camera are defined by its *intrinsic* parameters, which mainly describe the focal length and optical center. The *extrinsic* parameters, on the other hand, describe the physical location (rotation and translation) of a camera with respect to a reference coordinate system. Using two cameras together, such that information can be shared between them, requires knowing both the intrinsic and the extrinsic parameters of the cameras. The process of computing the intrinsic and extrinsic (rotation and translation) parameters of a camera at the same time is known as *camera calibration* in computer vision. We calibrated our visible band and thermal cameras using MATLAB's camera calibration toolbox as follows:

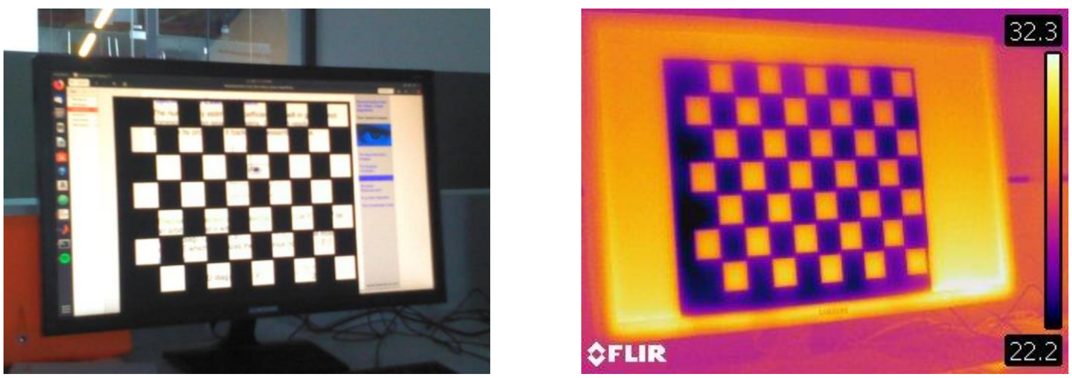
Visible Band Calibration: We manufactured a checkerboard calibration object using a piece of cardboard, whose images were taken by the visible band camera (

a. Figure 2-a). We calculated the intrinsic and extrinsic parameters of the visible band camera with 20 pictures of the calibration object captured by the camera.

Thermal Calibration: The same checkerboard pattern is used for the calibration of the thermal camera since the pattern is observable in the thermal image (

b. Figure 2-b). We similarly calculate the intrinsic and extrinsic parameters of the thermal camera using 20 images of the calibration object captured by the thermal camera.

c. Calculation of rotation and translation between the two cameras: The visible band camera and the thermal camera are mounted on the same camera body (Figure 3), which makes calculating the transformation (translation and rotation) between them easy and practical. The extrinsic parameters were obtained with respect to the same calibration object, which is visible for both visible band and thermal cameras. We exploit this fact to calculate the relative translation and rotation between the two cameras.



(a) (b)

Figure 2 The checkerboards used for calibrating the visible band and thermal cameras.

Figure 3 A view of FLIR E60.

2.2 3D Point Cloud Reconstruction

For generating a high-quality 3D point cloud of the room, we mainly rely on Structure from Motion (SfM), a widely-used technique in computer vision (Figure 4). In SfM, a set of images of an environment is used to obtain 3D information about the environment. These images are assumed to be captured at different positions (and possibly with different cameras) and to contain overlapping views of the environment. From these overlapping views, or more technically, the visual information that corresponds to the same 3D entities in the environment, the positions of the cameras and the 3D coordinates of the pixels can be identified.

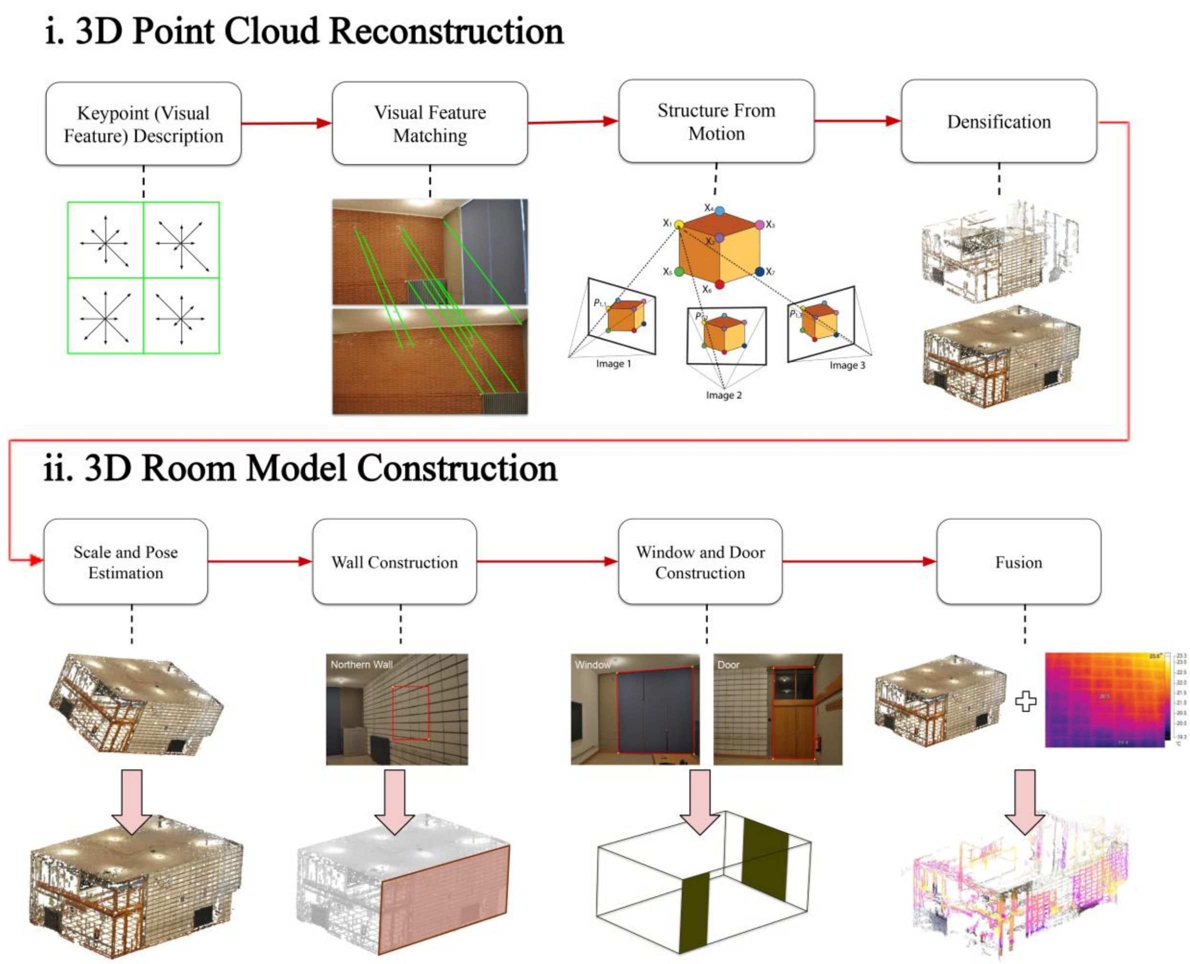


Figure 4 An overview of the main steps of 3D point cloud generation and 3D model construction. The dense 3D point cloud generated by step (i) is provided as input to the 3D room model construction step (ii)

To perform 3D reconstruction to obtain a point cloud, we developed a software tool by adapting the OpenMVG library [52]. Our tool follows the steps outlined in Figure 4 and is described in detail in the following sections.

2.2.1 Keypoint (Visual Feature) Description

SfM relies on finding matching pixels across the different views. However, not every pixel in an image carries visually meaningful information, and therefore, trying to find which pixels carry meaningful information and how we can represent this information are very critical. To find those "useful" pixels (called keypoints) as well as to represent information at and around those keypoints, we use the Scale-Invariant Feature Transform (SIFT) method [53]. SIFT finds keypoints by looking at intensity changes in an image at multiple scales. If there is a consistent change at a pixel at different scales, then that pixel is assumed to carry useful information. For representing such a keypoint, SIFT calculates a summary of how the intensity changes around the keypoint in the form of a 128-dimensional vector.

2.2.2 Visual Feature Matching

The previous step has identified in each image useful keypoints and represented them as (feature) vectors. Before estimating the 3D coordinates, the keypoints that correspond to the same 3D points should be identified. This is accomplished by comparing the feature vectors across the different views, and the closest feature vectors are identified as matching. For matching we use a cascaded method [54], which results in a set of potential matches between features in different views. Then, a post-processing step is employed to remove matches that are geometrically incorrect using AC-RANSAC (acronym for A Contrario RANdom SAmple Consensus) [55].

2.2.3 Structure from Motion (SfM)

SfM essentially formulates and solves jointly (1) the 3D reconstruction of matching points and (2) the estimation of the relative 3D distance and pose between the images. For SfM, we use the method proposed by Moulon et al. [55] due to its robustness and adaptive capacity. The method constructs an initial 3D model using the best matching two images and continues reconstruction by adding the remaining images iteratively.

2.2.4 Densification using Multi-view Stereo (MVS)

Since a dense 3D point cloud is needed for 3D reconstruction, it is essential to densify the sparse 3D point cloud computed by SfM. For densification, we employed an existing algorithm for multi-view stereopsis [56], which densifies a given 3D point cloud by interpolation.

2.2.5 Enhancement through Interaction

The 3D point cloud generated by the previous step requires adjustments on the scale, and selection of the planar regions on the walls, and the doors and the windows by mouse clicks.

2.2.5.1 Adjusting the Scale and Pose of the 3D Model

The 3D point cloud obtained in the previous step needs to be corrected for its scale and pose. We first select three points from the captured visual band images (Figure 5). We also identify these points' actual coordinates by manually measuring the distances to a selected origin point in the room. With this information at hand, we can calculate the transformation between the current model and the target model with correct scale and orientation. For this, we compute a *similarity transformation:*

$$\mathbf{x}' = \mathbf{A}\mathbf{x} + \mathbf{t}, \tag{1}$$

where $\mathbf{x} \in \mathfrak{R}^3$ is a 3D point in the original, scale-free, arbitrarily oriented 3D model; $\mathbf{A}$ is $\mathfrak{R}^{3\times3}$ orthogonal matrix with rotation and scaling elements; $\mathbf{t} \in \mathfrak{R}^3$ is translation; and $\mathbf{x}' \in \mathfrak{R}^3$is the scale-corrected, orientation-corrected 3D point. The solution is obtained using a non-linear least square optimization method [57] that provides the lowest mean squares error (i.e. $\sum_i (x_i' - x_i)\ ^2/n$). This similarity transformation is applied onto the reconstructed 3D point cloud to correct the scale, orientation and translation of the 3D point cloud. However, this process can introduce some degree of discrepancy, since it is inherently challenging to select the pixel that corresponds to a known 3D point. This imprecision amplified further with the use of low resolution images, where finding the correct match between the model and the image pixel is challenged further. Moreover, the rooms themselves are not perfect constructions (e.g. rounded corners, tilted walls), which makes the process noisier.

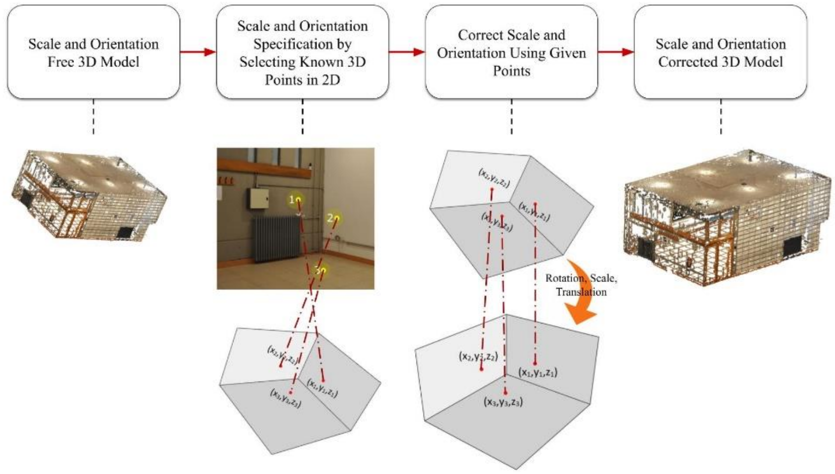


Figure 5 Scale and pose correction

2.3 3D Room Model Construction

The dense 3D model has a set of 3D points in 3D space. For correctly constructing a 3D model, the surfaces need to be identified and estimated as 3D planes. The equations of a plane for each boundary surface (walls, floor and ceiling) are computed for the points that lie within a rectangular shape that is manually marked by the user on the images using our tool. In this process, it is important to mark only the surface portions that are free of any possible obstructions, in order to eliminate the objects that can misinform plane fitting. As such, the algorithm can effectively function in physical environments that are heavily obstructed or cluttered.

We developed two methods for the robust estimation of the corresponding planes from the 3D point cloud. The first method, namely the "Baseline Plane Estimation from 3D Point Cloud" (BPE) method, is based on RANSAC and SVD, which estimates one plane corresponding to each wall, ceiling or floor independently. The second method, namely the "Robust 3D structure estimation with geometric constraints" (RSEC) method, is an improvement over the first method by increasing the precision of the final model. RSEC builds upon BPE and exploits the assumption of the rectangularity of the room.

The baseline method in Section 2.3.1 is very similar to the method used for 3D model construction from visible band images in Ham and Golparvar-Fard [32]. However, the second method, RSEC, to be presented in Section **Error! Reference source not found.** is our contribution.

2.3.1 Baseline Plane Estimation from 3D Point Cloud (BPE) Method

The BPE method is based on least-squares plane fitting algorithms with RANSAC. As described above, the user selects a four-corner polygon on each wall such that the points in the polygon are coplanar. The 3D points that are bounded by this region are first found in the 3D point cloud. Afterwards, our algorithm tries to fit a plane to each wall to construct the geometric model of the room in 3D as follows:

<u>Using RANSAC for 3D plane estimation</u>

Several points are randomly selected from each surface to estimate a plane. The performance of the plane fit is measured by computing the inlier ratio of our model estimation; in other words, we evaluate the success of the estimated model based on the ratio of the points in the estimated model. A distance thresholding method is utilized for inlier decision. If the distance between a point and its corresponding estimated plane is below a certain threshold (which is basically a hyperparameter), the point is an inlier point. After this process is iterated a number of times, the geometric model with the highest inlier ratio is kept as the best model.

<u>Plane Fitting</u>

We adapted a least square error plane fitting algorithm for model estimation in the proposed pipeline. First, we have N 3D points, $\mathbf{x}_1, \dots, \mathbf{x}_N$ with $\mathbf{x}_i \in \mathbb{R}^3$, sampled by RANSAC that can be stacked in a $N \times 3$ matrix as $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N]^T$ where T denotes transpose. As typically performed as a preprocessing step, the center point of the set is calculated and subtracted from all points to shift the center of the points to the origin. As a result, the new points and the new $N \times 3$ point matrix become:

$$\bar{\mathbf{x}} = \frac{1}{N}\sum_{i=1}^{N} \mathbf{x}_i \,, \tag{2}$$

$$\mathbf{x}_i' = \mathbf{x}_i - \bar{\mathbf{x}}, \tag{3}$$

$$\mathbf{X}' = [\mathbf{x}_1', \mathbf{x}_2', \dots, \mathbf{x}_N']^T. \tag{4}$$

The goal of plane fitting is to find a normal vector $\mathbf{n} \in \mathbb{R}^3$ that minimizes the mean square error (distance) of the 3D points that are expected to be on the wall:

$$\mathbf{n}^* \leftarrow \arg\min_{\mathbf{n}} \sum_{i=1}^{N} |\mathbf{n}^T \mathbf{x}_i|^2 = \arg\min_{\mathbf{n}} \mathbf{n}^T \mathbf{X}'^T \mathbf{X}' \mathbf{n}, \qquad s.t.\ \left|\left|\mathbf{n}\right|\right|_2^2 = 1. \tag{5}$$

The Lagrange multipliers method is an appropriate choice for minimizing this cost since we need to perform least squares minimization with the constraints. The cost function can be expressed in Lagrangian multipliers as:

$$J(\mathbf{X}'; \mathbf{n}, \lambda) = \mathbf{n}^T \mathbf{X}'^T \mathbf{X}' \mathbf{n} + \lambda(1 - \mathbf{n}^T \mathbf{n}). \tag{6}$$

To find a solution that minimizes the function J, we need to take the derivative of J and equate to 0:

$$\frac{\partial J}{\partial \mathbf{n}} = 2\mathbf{X}'^T \mathbf{X}' n - 2\lambda \mathbf{n} = 0, \tag{7}$$

$$\mathbf{X}'^T \mathbf{X}' \mathbf{n} = \lambda \mathbf{n}. \tag{8}$$

The vector $\mathbf{n} \in \mathbb{R}^3$ satisfying Eq. 8 is an eigenvector of $\mathbf{X}'^{\mathrm{T}} \mathbf{X}'$. We need to find the cost $\mathbf{n}^T \mathbf{X}'^T \mathbf{X}' \mathbf{n}$ in terms of the eigenvalue:

$$\mathbf{n}^T\mathbf{X}'^T\mathbf{X}'\mathbf{n} = \lambda\, \mathbf{n}^T\mathbf{n} = \lambda. \tag{9}$$

Therefore, to minimize the cost we need to select the eigenvector (normal) corresponding to the minimum eigenvalue:

$$\mathbf{n}^* \leftarrow \underset{\mathbf{n}}{\arg\min}\, \lambda. \tag{10}$$

The plane equation is then calculated as:

$$\mathbf{n}^*(\mathbf{x} - \mathbf{x}_0) = ax + by + cz + d. \tag{11}$$

where $\mathbf{n}^*$ is the normal vector corresponding to the plane, $\mathbf{x}_0$ is a known fixed point on the plane and $\mathbf{x}$ is any point on the plane. Hence, the parameters a, b, c and d become:

$$a = \mathbf{n}_x^*, \quad b = \mathbf{n}_y^*, \quad c = \mathbf{n}_z^*, \quad d = -\mathbf{n}^{*T}\mathbf{x}_0, \tag{12}$$

where $\mathbf{n}_x$, $\mathbf{n}_y$ and $\mathbf{n}_z$ are the x, y and z components of vector $\mathbf{n}$.

2.3.2 Robust 3D structure estimation with geometric constraints (RSEC)

We propose an improvement over the baseline method in order to exploit the 3D structure of the room. To be specific, we assume that room surfaces meet at a right (90^o) angle, and plane estimation is performed with this as a constraint. Accordingly, RSEC requires the computation of three surface normal vectors that are orthogonal to each other. In order to provide this constraint, two different cost functions are required for the whole room (recall that, in BPE, each wall is handled independently); one for plane fitting and a second one for the surface orthogonality constraint. The first cost function is defined as follows:

$$J_{fit}(\mathbf{X}_1, \mathbf{X}_1, \dots \mathbf{X}_6; \mathbf{n}_1,\ \mathbf{n}_2,\ \mathbf{n}_3)$$
$$= \mathbf{n}_1^T \mathbf{X}_1^T \mathbf{X}_1 \mathbf{n}_1 + \mathrm{n}_2^T \mathbf{X}_2^T \mathbf{X}_2 \mathrm{n}_2 + \mathbf{n}_1^T \mathbf{X}_3^T \mathbf{X}_3 \mathbf{n}_1 + \mathbf{n}_2^T \mathbf{X}_4^T \mathbf{X}_4 \mathbf{n}_2 \tag{13}$$
$$+ \mathbf{n}_3^T \mathbf{X}_5^T \mathbf{X}_5 \mathbf{n}_3 + \mathbf{n}_3^T \mathbf{X}_6^T \mathbf{X}_6 \mathbf{n}_3,$$

which can be simplified as:

$$J_{fit}(\mathbf{X}_1, \mathbf{X}_1, \dots \mathbf{X}_6; \mathbf{n}_1,\ \mathbf{n}_2,\ \mathbf{n}_3) = \mathbf{n}_1^T \mathbf{A} \mathbf{n}_1 + \mathbf{n}_2^T \mathbf{B} \mathbf{n}_2 + \mathbf{n}_3^T \mathbf{C} \mathbf{n}_3, \tag{14}$$

where $\mathbf{A}, \mathbf{B}, \mathbf{C}$ are symmetric positive semi-definite matrices defined as $\mathbf{A} = \mathbf{X}_1^T \mathbf{X}_1 + \mathbf{X}_3^T \mathbf{X}_3$; $\mathbf{B} = \mathbf{X}_2^T \mathbf{X}_2 + \mathbf{X}_4^T \mathbf{X}_4$, and $\mathbf{C} = \mathbf{X}_5^T \mathbf{X}_5 + \mathbf{X}_6^T \mathbf{X}_6$. The three vectors $\mathbf{n}_1$, $\mathbf{n}_2$ and $\mathbf{n}_3$ can be stacked into a 3×3 matrix as $\mathbf{N} = [\mathbf{n}_1,\ \mathbf{n}_2,\ \mathbf{n}_3]$. The cost then can be formulated using the **N** matrix as:

$$J_{fit}(\mathbf{X}_1, \mathbf{X}_1, \dots \mathbf{X}_6; \mathbf{N}) = \mathbf{i}_1^T \mathbf{N}^T \mathbf{A} \mathbf{N} \mathbf{i}_1 + \mathbf{i}_2^T \mathbf{N}^T \mathbf{B} \mathbf{N} \mathbf{i}_2 + \mathbf{i}_3^T \mathbf{N}^T \mathbf{C} \mathbf{N} \mathbf{i}_3\,, \tag{15}$$

where $\mathbf{i}_1 = [1\ 0\ 0]^T$, $\mathbf{i}_2 = [0\ 1\ 0]^T$ and $\mathbf{i}_3 = [0\ 0\ 1]^T$.

The second cost function which measures orthogonality can be defined by exploiting the fact that $\mathbf{N}^T\mathbf{N}$ should be equal to the identity matrix **I** by the orthogonality principle as follows:

$$J_{ort}(\mathbf{N}) = \left|\left|\ \mathbf{N}^T\mathbf{N} - \mathbf{I}\ \right|\right|_F^2 = \mathrm{tr}\big((\mathbf{N}^T\mathbf{N} - \mathbf{I})^T(\mathbf{N}^T\mathbf{N} - \mathbf{I})\big), \tag{16}$$

where $\left|\left|\ \mathbf{D}\ \right|\right|_F = \sum_i \sum_j D_{ij}^2$ is the Frobenius norm.

As the last step, these two cost functions are merged using a penalty factor:

$$J(\mathbf{X}_1, \mathbf{X}_1, \dots \mathbf{X}_6; \mathbf{N}, \lambda) \; = \; J_{fit}\,(\mathbf{X}_1, \mathbf{X}_1, \dots \mathbf{X}_6; \mathbf{N}) + \; \lambda\, J_{ort}(\mathbf{N})\,. \tag{17}$$

The final cost function in Eq. 17 is minimized by using the "Nelder-Mead" method [58] with an initial **N** matrix provided by consecutive three normal vectors provided by our baseline method. Here, optimal **N** is computed as:

$$\mathbf{N}^* \; \leftarrow \; \arg\min_{\mathbf{N}} \; J(\mathbf{X}_1, \mathbf{X}_1, \dots \mathbf{X}_6; \mathbf{N}, \lambda). \tag{18}$$

We tested different λ values such as 0.1, 1.0 and 10.0 with 140 images from a Nikon D90 camera and we settled on λ to 1.0, which experimentally provided maximum geometric accuracy.

2.3.3 Window and Door Selection

To mark the windows and doors in 3D, the corners of the windows and the doors are selected by the users on the visible images. With this, we assume that the windows and doors need to be co-planar with the walls on which they are located. After this selection, the corresponding 3D point for each corner point is calculated by intersecting the ray passing from selected point on the 2D plane with the corresponding wall plane.

2.4 Fusion with Thermal Data

In this step, the 3D points are assigned thermal values from the thermal images to be able to calculate thermal transmittance values for the external walls in the next step. To this end, the (intrinsic) parameters of visible band and thermal cameras and the relative 3D rotation and 3D translation between them are used. This is different from Ham and Golparvar-Fard [32], who performed separate reconstructions for the visible band and thermal images and then combined them. In the case of calibrated pair of cameras, this is inefficient and unnecessary.

The steps of how we fuse the 3D model with the thermal images are as follows:

- After calibration, we have $\mathbf{R}_{RGB}$ and $\mathbf{t}_{RGB}$, the 3D rotation and translation of the visible (RGB) camera with respect to the calibration object.
- We then calculate the transformation from a 3D point **x** to the thermal camera as follows:

$$\mathbf{p} = \mathbf{P}_{thermal}\,\mathbf{x}, \tag{19}$$

$$\mathbf{P}_{thermal} = \mathbf{K}_{thermal}[\mathbf{R}_{thermal}\,,\mathbf{t}_{thermal}], \tag{20}$$

$$\mathbf{R}_{thermal} = \mathbf{R}_{relative}\mathbf{R}_{RGB}^{T}, \tag{21}$$

$$\mathbf{t}_{thermal} = \mathbf{R}_{RGB}^{T}\mathbf{t}_{RGB} + \mathbf{t}_{relative}, \tag{22}$$

where $\mathbf{R}_{thermal}$ is the rotation matrix of the thermal camera, $\mathbf{R}_{relative}$ is the rotation matrix of the rotation difference between the RGB and the thermal cameras (obtained via the calibration step), $\mathbf{t}_{thermal}$is the translation vector of the thermal camera, $\mathbf{t}_{relative}$ is the translation vector of the translation difference between the RGB and the thermal camera, $\mathbf{P}_{thermal}$is the projection matrix and $\mathbf{K}_{thermal}$ is the intrinsic matrix of the thermal camera.

- Each 3D point of the point cloud is projected onto thermal images' planes using $\mathbf{P}_{thermal}$ the previously calculated projection matrices. Afterwards, the depth (the 3rd value that we get from projection since we use homogenous coordinate system) and the projected coordinates are checked. If the depth and the projected coordinates are both positive and within the image's coordinates, they are accepted as point correspondences and the RGB color of the thermal image point is assigned to its corresponding 3D point (Figure 6).

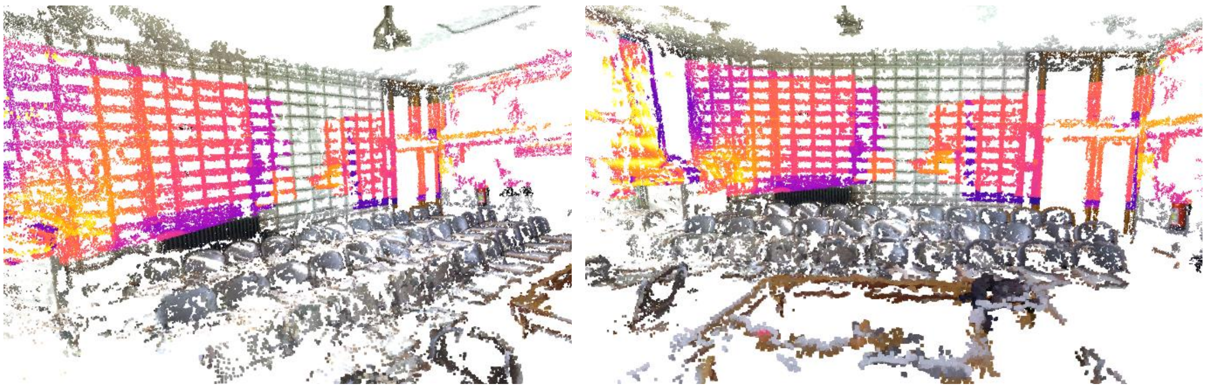


Figure 6 Juxtaposition of the thermal images on the point cloud.

2.5 Calculation of thermal transmittance

The thermal characterization of opaque external building surfaces is of key importance for energy simulations. Particularly, surface thermal transmittance, or U-value, is mainly used in the calculation of conductive heat transfer between the indoor and outdoor environments. U-value is dependent on the thickness and type of materials, which may be unknown for existing buildings. The heat flux method (HFM) is a non-destructive method that calculates R-value by measuring the temperature gradient and the direct heat-flux through the envelope with heat flux meters. However, the results can easily be affected by thermal bridges, humidity, mold and poor adhesion of the sensors [59]. As a robust and easy alternative to HFM, IR imaging can be used to estimate U-value using surface temperature values measured by a thermal camera. Instead of partially focusing on a limited number of measurements read from heat flux meters in HFM, the IR thermography method can estimate the average temperature and the overall R-value of a surface [60]. The main assumption behind the IR-based calculations is that the total heat transfer from the surface to the thermal camera is due to thermal radiation and thermal convection. Radiation heat transfer happens between two physically disconnected bodies with different temperatures. Convective heat transfer is transfer of energy between a moving gas or liquid phase and a solid phase, or, in this case, the building element and the indoor air.

U-values of the external walls are calculated using an existing infrared thermography method proposed by Albatici et al. [51]. This method assumes that heat passing through the element, dissipated from its surface and transferred to the IR thermal camera sensor (*P*), is the sum of heat dissipated by the element for radiation (*E*) and heat dissipated for convection (*H*). *E* is calculated using the Stefan–Boltzman Law for grey body radiation as:

$$E = \sigma \epsilon T_{s,in}^4 \tag{23}$$

where σ is the Stefan-Boltzmann constant of proportionality (5.67×10^{-8} [W/(m^{2}K^4]), ϵ is the thermal emissivity of the surface and $T_{s,in}$ is the surface temperature (K) of the external wall measured by the thermal camera. For the specific cases wherein a gray body is completely enclosed within a closed environment (i.e. a room), Eq. 23 can be modified to account for the net radiation exchange as:

$$E = \sigma \epsilon (T_{s,in}^4 - T_{refl}^4) \tag{24}$$

where T_{refl} is the measured dry bulb air temperature of the environment (K). T_{refl} is measured to exclude the impact of reflected radiation in the thermal image, and to acquire the surface's correct temperature information. T_{refl} is the average reflected temperature from a reflective mirror (i.e. a crumpled aluminum coil) that is placed at a short distance from the wall and measured by a thermal camera. *H* can be calculated as:

$$H = \alpha_{con} \left| T_{s,in} - T_{air,in} \right| \tag{25}$$

where α_{con} is the convective heat transfer coefficient [Wm^{-2}K] and $T_{air,in}$ is the measured air temperature of the indoor environment. Finally, *U-value* [W/m^{2}K] of a building surface is calculated by considering that *P* is the sum of *E* and *H*:

$$U = P/(T_{air,in} - T_{air,out}) \tag{26}$$

$$U = \frac{\sigma\epsilon\left|T_{s,in}^4 - T_{refl}^4\right| + \alpha_{con}\left|T_{s,in} - T_{air,in}\right|}{T_{air,in} - T_{air,out}} \tag{27}$$

where $T_{air,out}$ is the measured outdoor air temperature. The U-value can thus be calculated from the mean surface temperature of each external wall.

2.5.1 Data integration

Once the previous steps are completed, the generated data (corner points of planes, U-values, window and door points) is written in an XML file with an XML schema developed by us. This file is read by a parser to construct a surface model for 3D model validation, and by OpenStudio SDK, an open-source framework that provides access to EnergyPlus object attributes [61] for building energy modeling.

In EnergyPlus, the opaque room surfaces are instantiated as the *BuildingSurface:Detailed* object type, which specifies the surface type (wall, floor, ceiling or roof), the thermal zone that the surface is a part of, the outside boundary condition, sun and wind exposure, the construction name and the four vertices that define the planar surface. The transparent surfaces are defined in a similar way using the *FenestrationSurface:Detailed* object type. Both opaque and transparent surfaces are instantiated by automatically reading in the vertices of the previously calculated surfaces, and manually entering the remaining information. The external walls, for which the U-value is calculated using the IR thermography method, need to be associated with the *Material:NoMass* object, as only the U-value of the surface is known. When instantiating this object, it should be associated with the relevant surface, and the thermal resistance (the reciprocal of the calculated U-value) value needs to be automatically entered. For all the other surfaces, the standard *Material* object type can be used, which requires manual data input for material thickness, conductivity, density and specific heat.

2.6 3D Room Reconstruction Tool

Together with the described pipeline, a graphical user interface (GUI) was designed to allow the users select the geometric model extraction method, normal mode or test mode (where the tool calculates the geometric errors 100 times and calculates these error's mean and standard deviation) the penalty method parameters used in robust 3D geometry estimation part, the optical and thermographic images for 3D reconstruction, manually mark the surfaces of walls on optical images for plane construction and mark the boundaries of windows and doors to define these elements in the model.

3 Results

In this section we present the results of the two experiments that aim to validate the developed pipeline. The first experiment aims to assess the performance of the pipeline and the two surface construction methods (BPE and RSEC) in constructing precise surface models subject to different input datasets. The second experiment is a comparative analysis between the calculated energy use of two energy simulation models: the first based on an existing high-precision surface model with theoretical thermal transmittance values, and the second using the proposed method that constructs an energy model with calculated thermal transmittance.

3.1 Experiment setup

The experiments were conducted in a classroom in an educational building (Figure 7**Error! Reference source not found.**). The room has an area of 46.48 m^2 and a volume of 171.85 m^3. The external wall material is reinforced concrete with 0.25 cm thickness. Previous to this study, a ground truth model of the room had been obtained through 3D laser scanning, using a high-precision laser scanner (Faro Focus 120) that is registered by FARO-Scene. The resulting 3D point cloud model was used to manually generate a surface model. The room is a rectangular prism, with a door and a window. The electro-optic images of the room were captured with a digital single-lens reflex camera, Nikon D90. Nikon D90 has a 12.3-megapixel resolution and a built-in autofocus motor. To capture the thermal images of the external wall, a FLIR E60 infrared camera was used. FLIR 60 has a thermal sensitivity of <0.05°C, with an 800-pixel resolution for infrared images (320×240).

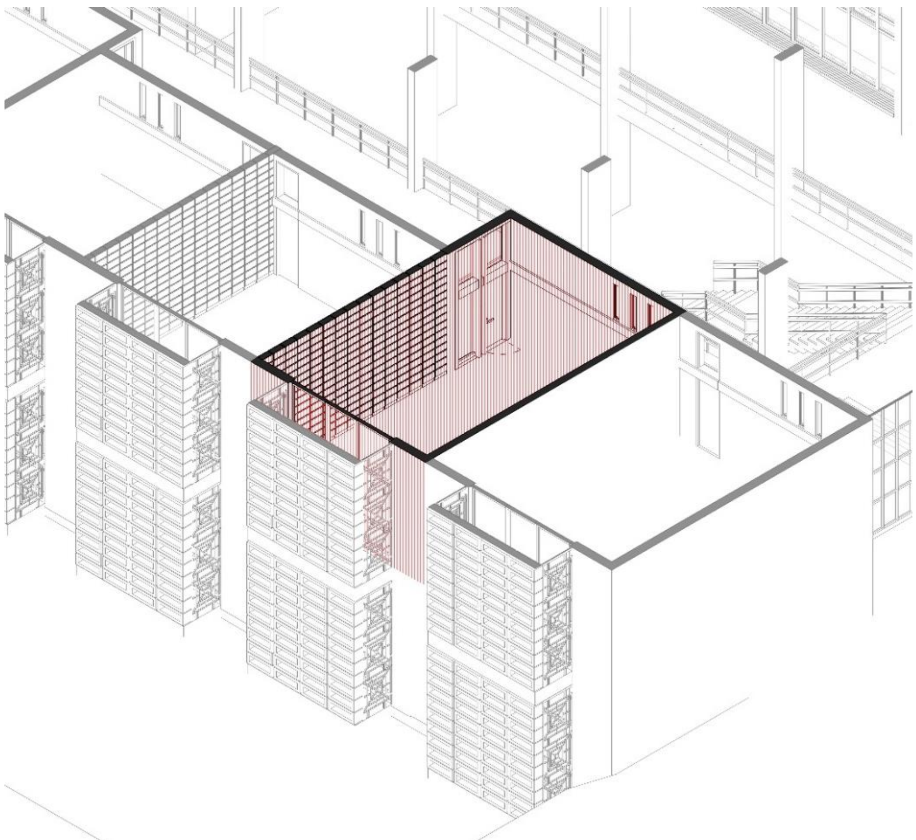


Figure 7 Isometric view of the selected classroom

The experiments aim to demonstrate the viability and precision of the proposed pipeline and the two surface construction methods (BPE and RSEC, as described in Section 2.3) when subjected to different input dataset conditions. The input datasets are representative of dataset sizes and image qualities (Table 1). The experimental setup considers a various number of images from 140 to 35, with both high- and low-quality images. To generate the latter, we decreased the pixel size of the original images and applied Gaussian noise with 5.0 standard deviation to lower the image quality. Each combination of image count and image resolution is called as an experiment case in the rest of the section.

Table 1 Experiment setup. Each combination of image count and resolution is used as an experimental case for evaluation in the rest of the paper.

Number of images	140p (140 photos) 105p (105 photos) 70p (70 photos) 35p (35 photos)
Image resolution	high resolution (4288×2848 pixel) low resolution (2144×1424 pixel, with Gaussian noise added)

For each setup case, the following steps are followed:

a. 8 image datasets are considered for 3D reconstruction, with different number of images and image resolutions, to generate the corresponding 3D point clouds (Figure 8-A).

b. The scale and pose of the point cloud model are adjusted through user interaction. Three points are selected in the physical environment, and their physical distances to a selected origin point are measured manually. The same points are selected on the images by the user using the developed tool, and the measurements are manually entered to the corresponding points (Figure 8-B).

c. The user selects the images that will be used to calculate the wall surfaces and marks the walls. This step can be realized using one of the methods that we developed, BPE or RSEC. As a result, the surfaces of the room are generated for each method (Figure 8-C and D).

d. The user the selects the images that will be used to calculate the window and doors, and outlines these surfaces on the images, as described in the previous section (Figure 8-E).

e. The resulting surface model is then converted to XML format to be read into a 3D modeler or the energy simulation tool.

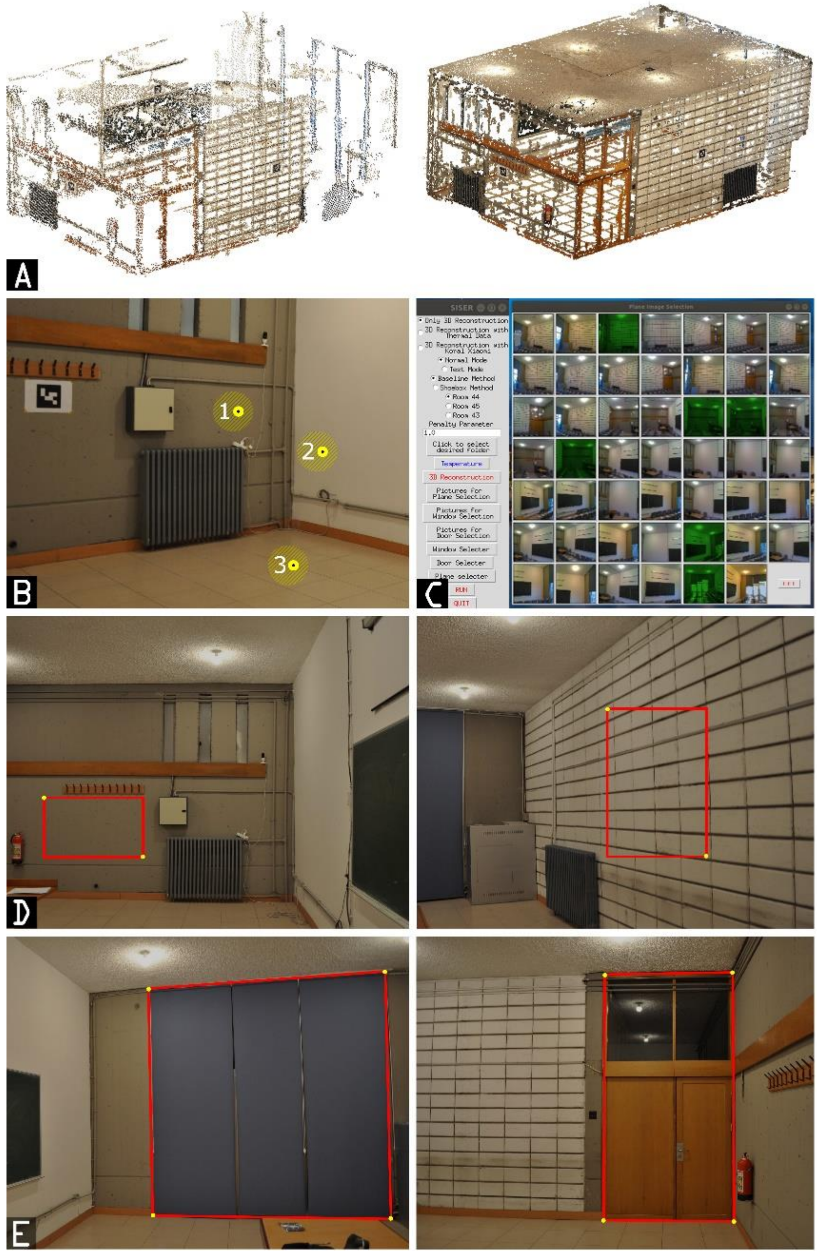


Figure 8 A. The sparse 3D model (left) and the densified 3D point cloud (right)surfaces, B. Adjusting scale and pose of the model, C. The main GUI and image selection for walls, D. Identification of the surfaces on the selected photos for each wall, E. Identification of the window and door surfaces by selecting a certain area on the desired window (left) or door (right).

3.2 Evaluation measures

The comparative analysis between the two 3D models are based on the evaluation measures (**Error! Reference source not found.**). The results of these measures facilitate the benchmarking between the computed models and the ground truth model. The evaluation measures include:

a. <u>The difference between room volumes</u>:

$$d_V(G,C) = |V_G - V_C|, \tag{28}$$

where $d_V(G,C)$ is the volume difference, V_G and V_R are the volume of the ground truth model and the calculated model respectively.

b. <u>The cumulative Euclidean distances between surface vertices</u>:

$$d_E(G,C) = \sum_{i=0}^{n} d\left(\mathbf{x}_i^g, \mathbf{x}_i^c\right), \tag{29}$$

where $d_E(G,C)$ is the total Euclidean distance error between all the vertices in the model including the boundary surfaces, windows and doors; G is the ground truth model; C is the calculated model; *n* is the total number of measured points; $\mathbf{x}_i^g \in G$ is the i^{th} point of the ground truth model; $\mathbf{x}_i^c \in C$ is the corresponding point of the computed model where correspondence is manually determined; and $d(\cdot,\cdot)$ is the Euclidean distance between two vectors.

c. <u>The cumulative angle differences between the surface normal vectors of the walls</u>:

$$d_\theta(G,C) = \sum_{j=0}^{n} \left|\mathbf{n}_i^g - \mathbf{n}_i^c\right|, \tag{30}$$

where $d_\theta(G,C)$ is the total angular distance error; n is the total number of wall surfaces; $\mathbf{n}_i^g$ and

$\mathbf{n}_i^c$ respectively are the surface normal of the i^{th} ground truth surface and the corresponding estimated surface.

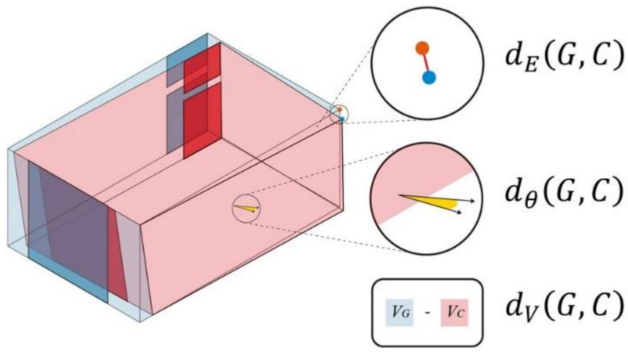


Figure 9 The evaluation measures. The blue box (G) and the red box (C) represent the ground truth geometry and the calculated geometry respectively.

3.3 Comparative analysis

The first experiment aims to assess the performance of the proposed pipeline subject to different datasets in constructing precise surface models, focusing particularly the two surface construction methods (BPE and RSEC). **Error! Reference source not found.** shows the change in computation time and point cloud density for eight point-cloud models generated for each setup case. The results depict an expected decreasing trend in the time cost and point cloud density with decreasing dataset sizes (Figure 11). The ratio of decrease in computational time from 140p to 35p are 73% and 67.1% for high- and low-resolution images respectively. The ratio of decrease in point cloud density from 140p to 35p are 58.3% and 58% for high- and low-resolution images respectively.

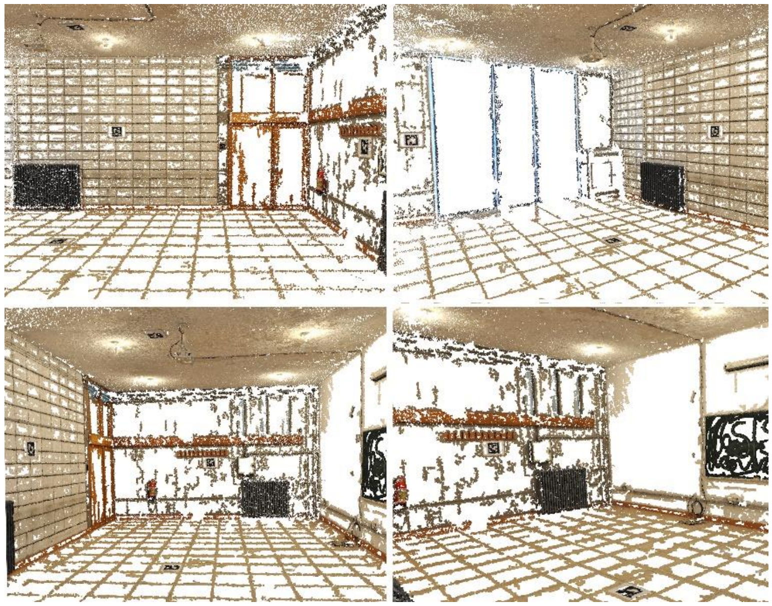
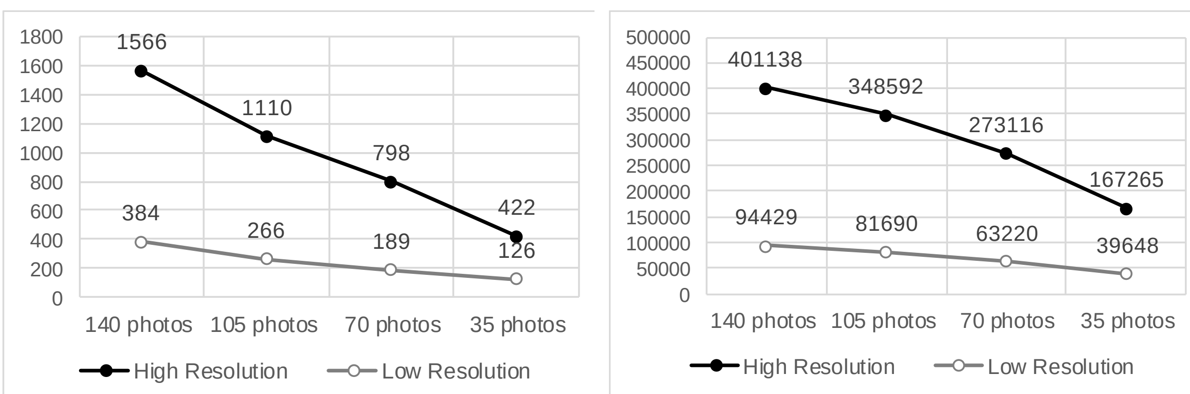


Figure 10 High resolution point cloud, interior views



(a) (b)

Figure 11 (a) Computational time cost for 3D point cloud generation in seconds and (b) 3D point cloud density (the number of points) for each setup. High resolution is 4288×2848 and low resolution is 2144×1424. The results are obtained on a PC with Intel® Core™ i7-7700 Processor (3.60GHz 8MB) with a 16 GB DDR4 RAM.

The results of the benchmark metrics can be found in **Error! Reference source not found.**. As expected, dataset size (the number of images) has the most significant impact on model accuracy. In all metrics, the magnitude of error consistently increases from high to low number of images. However, the error difference between high and low number of images is more prominent in the BPE method. In other words, model precision diminishes more rapidly for BPE as compared to RSEC. For instance, the absolute difference for ΔV amounts to 5.55 and 5.89 m^2 for BPE^{high} and BPE^{low}, while for $RSEC^{high}$ and $RSEC^{low}$ these values are 2.16 and 3.35 m^2 (**Error! Reference source not found.**-a). Moreover, models constructed with 140p using $RSEC^{high}$ and $RSEC^{low}$ start off with lower precision (d_V = 1.43 m^2 and 1.35 m^2) as compared to BPE^{high} and BPE^{low} (d_V = 0.07 m^2 and 0.83 m^2), but outperform BPE^{high} and BPE^{low} as the number of images drop to 35. This underlines that the RSEC method can be more viably and easily used in practical settings. The volume error is most determinant for energy simulations, as the heating / cooling energy consumption is directly proportional to the room volume. This places additional emphasis on the selection of the correct method and underlines the importance of an accurate understanding of the tradeoffs between the use of different datasets.

Table 2 The results of the benchmark metrics. The best values for each error measure are marked boldface.

Metric	Method	*140p*	*105p*	*70p*	*35p*
Volume difference error ($\boldsymbol{d_V(G,C)}$)	BPE^{high}	**0.07**	**1.08**	3.16	5.62
	$RSEC^{high}$	1.43	1.80	**2.65**	**3.59**
	BPE^{low}	0.83	1.55	4.46	6.73
	$RSEC^{low}$	1.35	2.21	3.66	4.70
Euclidean distance error ($\boldsymbol{d_E(G,C)}$)	BPE^{high}	3.35	3.15	3.08	4.92
	$RSEC^{high}$	**2.30**	**2.03**	**2.11**	**2.33**
	BPE^{low}	3.52	3.64	4.24	5.79
	$RSEC^{low}$	2.64	3.04	3.84	4.46
Angle difference error ($\boldsymbol{d_\theta(G,C)}$)	BPE^{high}	6.56	7.52	9.24	13.56
	$RSEC^{high}$	**0.09**	**0.09**	**0.09**	**0.09**
	BPE^{low}	9.45	9.24	14.74	23.13
	$RSEC^{low}$	3.92	5.62	8.48	8.04

The results of d_E and d_θ (**Error! Reference source not found.**-b and **Error! Reference source not found.**-c) show a similar trend to d_D, such that the magnitude of error increases with lower image number. The exception to this is the results of $RSEC^{high}$, wherein a insignificant change d_E is observed, and d_θ is zero. *BPE* performs poorly in d_θ with 140p for both high- and low-resolution images (d_θ = 6.56° and 9.45° respectively), which degrades to intolerable results with 35p (13.56° and 23.13°). In contrast, d_θ results with $RSEC^{high}$ are negligible, indicating that the constraints could ensure the orthogonality of the room geometry. $RSEC^{low}$, on the other hand, showed poorer performance in d_θ due to the lower resolution point cloud.

3.4 Building energy modeling

In this section, we assess the viability of the use of the proposed method in building energy modeling. In the previous section, we tested the geometric precision of the constructed geometries through metrics that quantify the magnitude of error between the ground truth geometry and the calculated geometries. In this section, we comparatively assess the difference in building performance metrics between a theoretical energy model and the calculated energy model.

140 images were captured each from electro-optic and IR cameras mounted on a FLIR E60. The IR images had 76,800-pixel (320×240) and the electro-optic images had 3.1-megapixel (2048×1536) resolution. While capturing images, measurements were taken for U-value calculations, on 24 December 2018 at 6:00am, to achieve a quasi-steady-state condition of heat transfer. The variables in Eq. 27 were measured as $T_{air,in}$= 23.1 C°, $T_{air,out}$= 1.0 C° and T_{refl} = 27.0 C°.

The images were processed through the pipeline shown in Figure 4. The resulting point cloud had a density of 403,861, and the calculated V = 164.38 m^3. The average U-value of the 0.25 m reinforced concrete external wall was calculated as 2.0 W/m^2-K. This result is consistent with a previous study that conducts in-situ IR Thermography measurements of the same building wall from the external environment by [62], which measured the external wall average U-value as 2.07±0,38 W/m^2K.

Following, two EnergyPlus models of the classroom was built for the ground truth (theoretical) model and the newly calculated building model. The theoretical U-value and the newly calculated U-values was used for the theoretical and calculated models respectively. The surfaces other that external surfaces were modeled as adiabatic surfaces to exclude heat transfer with other indoor spaces. Other material thermal characteristics used in both models can be found in Table 3. The windows are modeled as double-glazed windows (U= 2.6 W/m^2K, SHGC = 0.75, VT = 0.8). The standard templates defined in DesignBuilder for people, lighting and equipment are used as internal loads. The heating setpoint and setback temperatures are set to 21 C° and 15.5 C° respectively. Infiltration is set to 25ACH at 50Pa. The simulation run period was set to 24-30 December, which was the week that the actual measurements were taken.

Table 3 The thermal characteristics of the opaque materials in the energy models

Element	**Material**	**Thickness (mm)**	**Theoretical U value (W/m^2-k)**	**Calculated U value (W/m^2-k)**
Wall	Concrete, reinforced	250	**3.7**	**2.0**
Roof	Concrete, Reinforced	130	0.577	- (the theoretical value is used in simulations)
	Waterproofing membrane	-		
	XPS - CO2 blowing	50		
	Roofing felt	4		
	Stone chipping	10		

The performance metrics are as follows:

a. Conductive heat loss through the external wall (Q_C) occurs as a result of temperature difference as well as the thermal properties of the wall.
b. Operative temperature (*OT*) is a metric for thermal comfort, which is defined as the average of indoor air dry bulb temperature and mean radiant temperature in a room. In cases that a room's boundary surfaces are different from the room temperature, significant changes to *OT* and occupant discomfort can be observed.
c. Heating energy use (Q_H) is the amount of energy to maintain the room temperature at the determine setpoint temperature.

According to the simulation results (

Table 4, Figure 12), the difference in total heat loss through the external from the theoretical model to the calculated model is -41.32 %. This decrease is due to the lower heat transfer rate through the concrete wall in the calculated model. Because of the increased thermal performance of the external wall in the calculated model, a decrease of -19.17 % in the heating energy use was observed. Hourly results also show that the results of the theoretical model are consistently offset from the calculated model. The maximum difference in heat loss is 0.30 kW (at 02/11, 15:00), operative temperature is 0.88 C° (at 02/09, 13:00), heating energy use is 1.28 kW (at 02/06, 13:00).

Table 4. The results of the energy simulations for the two models

	Total heating energy use (kW)	**Total heat loss through external walls (kW)**	**Average Operative Temperature (C°)**
Simulation with theoretical model	617.56	65.51	18.17
Simulation with generated model	499.18	38.44	18.37
% change	-19.17 %	-41.32 %	1.1 %

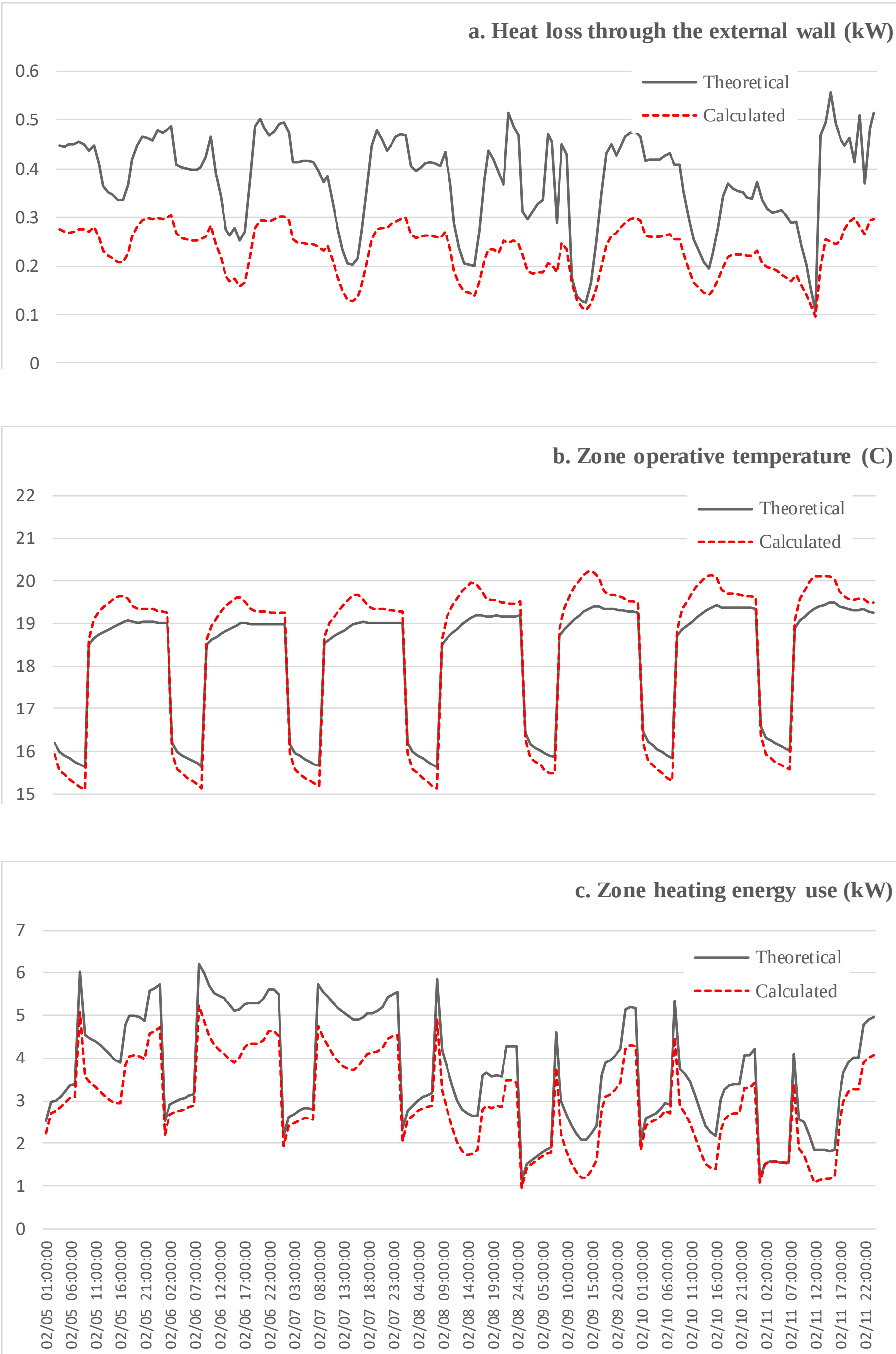


Figure 12 The hourly results of the energy simulations for the two models

4 Conclusion

In this paper, we propose a 3D reconstruction pipeline that semi-automatically merges digital 2D visible images and IR images of a room into a single 3D building energy model. Our approach facilitates the fast and easy modeling of buildings by also allowing users' interaction with the 2D images. During conversion of 3D point clouds into planar surface models, two methods are proposed for the robust estimation of planes for room surfaces. The first method, BPE, is based on RANSAC and SVD, which estimates planes corresponding to each wall, ceiling or floor independently. The second method, RSEC, is an improvement over BPE, exploiting the assumption of rectangularity of the room and considering surface orthogonality as a second constraint to plane estimation. For surface U-value estimation, an existing non-destructive method based on infrared thermography [51] was used, and a similar approach to [32] was developed in the registration of U-values to the 3D point clouds.

The proposed pipeline was evaluated in a classroom, wherein the electro-optic images, thermal images and the environmental conditions to calculate wall U-value were first captured. Two experiments were carried out. The first experiment assessed the performance of the pipeline subject to different input datasets representing different dataset sizes and image resolutions. The second experiment aimed to assess the viability of the pipeline in building energy modeling through a comparison of simulation results. The results of the first experiment has shown a consistent change in point data cloud density as a result of different datasets. While model precision reduces from high to low number of images for all evaluation measures, this change is more prominent for BPE as compared to RSEC. For all the evaluation measures, BPE calculates more accurate models for 140p, but RSEC outperforms BPE when the dataset scale is reduced towards 35p. This indicates RSEC's robustness to low image quality and reduced sizes of input datasets. Moreover, although the experiments were conducted on a simple geometry, the RSEC method is generalizable to rooms of arbitrary complexity, provided that the angular relationships are previously known and specified. Our RSEC method jointly optimizes the parameters of the planes for all surfaces at once. This aim is achieved by assuming some *a priori* angular difference between the wall normal vectors such that the overall solution satisfies these orientation constraints together. Provided that the orientation constraints are suitably adjusted, the RSEC method can be applied to any room with planar surfaces.

The proposed approach requires user interaction for the identification of building elements to be modeled. Different from the existing approaches, user interaction in our pipeline in realized using the 2D images through simple mouse clicks. This simple interaction routine does not necessitate any expertise on complex software tools or digital models that might be difficult to manage for the users. More specifically, our approach unburdens the user of complex operations performed on 3d point clouds, and instead allows interaction with easy-to-understand images. Our approach, therefore, can be said to be advantageous due to its ease of use and practicality.

The extension of our pipeline to buildings requires addressing several issues. The first issue is the reduction of user interaction for e.g. the definition of doors and windows. This problem is especially relevant for large-scale settings, wherein the definition of a number of building elements in the model might exceed the capacity of user interaction. In this case, an object detector (e.g. YOLO [63], RetinaNet [64]) can be used to detect doors and windows automatically, and the user could only correct mis-detections or errors in localization. The second problem is putting together 3D models of single rooms for large-scale buildings. Currently, the proposed pipeline currently supports the accurate 3D modeling of a single room. However, practical contexts that aim to assess whole building performance usually consider multiple rooms in a building. This issue can be addressed by making use of simultaneous localization and mapping (SLAM) techniques [65]. In SLAM, a moving camera continuously captures snapshots from an environment, from adjacent frames 3D scene is reconstructed and "appended" to the current 3D model of the whole environment by considering the 3D motion of the camera. With such an approach, the 3D spatial transformation between the rooms of a building can be estimated and therefore, 3D models of single rooms can be stitched together to form a single 3D model. We leave both aspects as future work.

Acknowledgements

This work was supported by an Institutional Links grant under the Newton-Katip Celebi partnership, Grant No. 217M519 by the Scientific and Technological Research Council of Turkey (TUBITAK) and ID [352335596] by British Council, UK. This is a substantially extended and enhanced version of the paper presented at the CIB W78 Annual Conference held at Northumbria University in Newcastle UK in September 2019. We would like to acknowledge the editorial contributions of Professor Bimal Kumar of Northumbria University and Dr. Farzad Rahimian of Teesside University in the publication of this paper. We also would like to thank METU Faculty of Architecture for their support during field tests. The LiDAR data of the test building was obtained during a previous research project on conservation planning, funded by the 'Keeping It Modern' Grant of the Getty Foundation. Their contribution is also gratefully acknowledged.

References

[1] R. Volk, J. Stengel, F. Schultmann, Building Information Modeling (BIM) for existing buildings - Literature review and future needs, Automation in Construction. 38 (2014) pp. 109–127. https://doi.org/10.1016/j.autcon.2013.10.023.

[2] J. Jung, S. Hong, S. Jeong, S. Kim, H. Cho, S. Hong, J. Heo, Productive modeling for development of as-built BIM of existing indoor structures, Automation in Construction. 42 (2014) pp. 68–77. https://doi.org/10.1016/j.autcon.2014.02.021.

[3] J.L.M. Hensen, R. Lamberts, Building performance simulation for design and operation, in: J.L.M. Hensen, R. Lamberts (Eds.), Building Performance Simulation for Design and Operation, Spon Press, 2012: pp. 1–14. https://doi.org/10.4324/9780203891612.

[4] A. Mahdavi, Simulation-based control of buildings systems operation, Building and Environment. 36 (2001) pp. 789–796. https://doi.org/10.1016/S0360-1323(00)00065-2.

[5] S. Wang, C. Yan, F. Xiao, Quantitative energy performance assessment methods for existing buildings, Energy and Buildings. 55 (2012) pp. 873–888. https://doi.org/10.1016/j.enbuild.2012.08.037.

[6] I. Brilakis, M. Lourakis, R. Sacks, S. Savarese, S. Christodoulou, J. Teizer, A. Makhmalbaf, Toward automated generation of parametric BIMs based on hybrid video and laser scanning data, Advanced Engineering Informatics. 24 (2010) pp. 456–465. https://doi.org/10.1016/j.aei.2010.06.006.

[7] F. Bosché, E. Guenet, Automating surface flatness control using terrestrial laser scanning and building information models, Automation in Construction. 44 (2014) pp. 212–226. https://doi.org/10.1016/j.autcon.2014.03.028.

[8] X. Xiong, A. Adan, B. Akinci, D. Huber, Automatic creation of semantically rich 3D building models from laser scanner data, Automation in Construction. 31 (2013) pp. 325–337. https://doi.org/10.1016/j.autcon.2012.10.006.

[9] L. Mahdjoubi, C. Moobela, R. Laing, Providing real-estate services through the integration of 3D laser scanning and building information modelling, Computers in Industry. 64 (2013) pp. 1272–1281. https://doi.org/10.1016/j.compind.2013.09.003.

[10] B. Akinci, F. Boukamp, C. Gordon, D. Huber, C. Lyons, K. Park, A formalism for utilization of sensor systems and integrated project models for active construction quality control, Automation in Construction. 15 (2006) pp. 124–138. https://doi.org/10.1016/j.autcon.2005.01.008.

[11] E. Valero, A. Adán, C. Cerrada, Automatic method for building indoor boundary models from

dense point clouds collected by laser scanners, Sensors (Switzerland). 12 (2012) pp. 16099–16115. https://doi.org/10.3390/s121216099.

[12] A. Dimitrov, M. Golparvar-Fard, Vision-based material recognition for automated monitoring of construction progress and generating building information modeling from unordered site image collections, Advanced Engineering Informatics. 28 (2014) pp. 37–49. https://doi.org/10.1016/j.aei.2013.11.002.

[13] F. Bosché, A. Guillemet, Y. Turkan, C.T. Haas, R. Haas, Tracking the built status of MEP works: Assessing the value of a Scan-vs-BIM system, Journal of Computing in Civil Engineering. 28 (2014). https://doi.org/10.1061/(ASCE)CP.1943-5487.0000343.

[14] P. Tang, D. Huber, B. Akinci, R. Lipman, A. Lytle, Automatic reconstruction of as-built building information models from laser-scanned point clouds: A review of related techniques, Automation in Construction. 19 (2010) pp. 829–843. https://doi.org/10.1016/j.autcon.2010.06.007.

[15] R. Hartley, A. Zisserman, Multiple View Geometry in Computer Vision, Cambridge University Press, 2003. https://doi.org/10.1016/S0143-8166(01)00145-2.

[16] M. Golparvar-Fard, J. Bohn, J. Teizer, S. Savarese, F. Peña-Mora, Evaluation of image-based modeling and laser scanning accuracy for emerging automated performance monitoring techniques, Automation in Construction. 20 (2011) pp. 1143–1155. https://doi.org/10.1016/j.autcon.2011.04.016.

[17] C. Strecha, W. Von Hansen, L. Van Gool, P. Fua, U. Thoennessen, On benchmarking camera calibration and multi-view stereo for high resolution imagery, in: 26th IEEE Conference on Computer Vision and Pattern Recognition, CVPR, 2008: pp. 1–8. https://doi.org/10.1109/CVPR.2008.4587706.

[18] A. Dimitrov, M. Golparvar-Fard, Segmentation of building point cloud models including detailed architectural/structural features and MEP systems, Automation in Construction. 51 (2015) pp. 32–45. https://doi.org/10.1016/j.autcon.2014.12.015.

[19] H. Kim, K. Kim, H. Kim, Data-driven scene parsing method for recognizing construction site objects in the whole image, Automation in Construction. 71 (2016) pp. 271–282. https://doi.org/10.1016/j.autcon.2016.08.018.

[20] J. Park, H. Cai, D. Perissin, Bringing Information to the Field: Automated Photo Registration and 4D BIM, Journal of Computing in Civil Engineering. 32 (2018). https://doi.org/10.1061/(ASCE)CP.1943-5487.0000740.

[21] S. Tang, Y. Zhang, Y. Li, Z. Yuan, Y. Wang, X. Zhang, X. Li, Y. Zhang, R. Guo, W. Wang, Fast and automatic reconstruction of semantically rich 3D indoor maps from low-quality RGB-D sequences, Sensors (Switzerland). 19 (2019) pp. 553. https://doi.org/10.3390/s19030533.

[22] I. Brilakis, H. Fathi, A. Rashidi, Progressive 3D reconstruction of infrastructure with videogrammetry, Automation in Construction. 20 (2011) pp. 884–895. https://doi.org/10.1016/j.autcon.2011.03.005.

[23] C.A. Balaras, A.A. Argiriou, Infrared thermography for building diagnostics, Energy and Buildings. 34 (2002) pp. 171–183. https://doi.org/10.1016/S0378-7788(01)00105-0.

[24] S. Ribarić, D. Marčetić, D.S. Vedrina, A knowledge-based system for the non-destructive diagnostics of façade isolation using the information fusion of visual and IR images, Expert Systems with Applications. 36 (2009) pp. 3812–3823. https://doi.org/10.1016/j.eswa.2008.02.043.

[25] M. Der Yang, T.C. Su, H.Y. Lin, Fusion of infrared thermal image and visible image for 3D thermal model reconstruction using smartphone sensors, Sensors (Switzerland). 18 (2018) pp.

2003. https://doi.org/10.3390/s18072003.

[26] D. González-Aguilera, P. Rodriguez-Gonzalvez, J. Armesto, S. Lagüela, Novel approach to 3D thermography and energy efficiency evaluation, Energy and Buildings. 54 (2012) pp. 436–443. https://doi.org/10.1016/j.enbuild.2012.07.023.

[27] P. Merchán, A. Adán, S. Salamanca, V. Domínguez, R. Chacón, Geometric and colour data fusion for outdoor 3D models, Sensors (Switzerland). 12 (2012) pp. 6893–6919. https://doi.org/10.3390/s120606893.

[28] A. Costanzo, M. Minasi, G. Casula, M. Musacchio, M.F. Buongiorno, Combined use of terrestrial laser scanning and IR Thermography applied to a historical building, Sensors (Switzerland). 15 (2014) pp. 194–213. https://doi.org/10.3390/s150100194.

[29] S. Lagüela, J. Armesto, P. Arias, J. Herráez, Automation of thermographic 3D modelling through image fusion and image matching techniques, Automation in Construction. 27 (2012) pp. 24–31. https://doi.org/10.1016/j.autcon.2012.05.011.

[30] A. Adan, T. Prado, S.A. Prieto, B. Quintana, Fusion of thermal imagery and LiDAR data for generating TBIM models, in: Proceedings of IEEE Sensors, 2017: pp. 1–3. https://doi.org/10.1109/ICSENS.2017.8234261.

[31] S. Schramm, J. Rangel, A. Kroll, Data fusion for 3D thermal imaging using depth and stereo camera for robust self-localization, in: 2018 IEEE Sensors Applications Symposium, SAS 2018 - Proceedings, 2018: pp. 1–6. https://doi.org/10.1109/SAS.2018.8336740.

[32] Y. Ham, M. Golparvar-Fard, An automated vision-based method for rapid 3D energy performance modeling of existing buildings using thermal and digital imagery, Advanced Engineering Informatics. 27 (2013) pp. 395–409. https://doi.org/10.1016/j.aei.2013.03.005.

[33] A. Rashidi, M.H. Sigari, M. Maghiar, D. Citrin, An analogy between various machine-learning

techniques for detecting construction materials in digital images, KSCE Journal of Civil Engineering. 20 (2016) pp. 1178–1188. https://doi.org/10.1007/s12205-015-0726-0.

[34] Q. Lu, S. Lee, L. Chen, Image-driven fuzzy-based system to construct as-is IFC BIM objects, Automation in Construction. 92 (2018) pp. 68–87. https://doi.org/10.1016/j.autcon.2018.03.034.

[35] C. Wang, Y.K. Cho, C. Kim, Automatic BIM component extraction from point clouds of existing buildings for sustainability applications, Automation in Construction. 56 (2015) pp. 1–13. https://doi.org/10.1016/j.autcon.2015.04.001.

[36] E. Valero, A. Adán, F. Bosché, Semantic 3D reconstruction of furnished interiors using laser scanning and RFID technology, Journal of Computing in Civil Engineering. 30 (2016). https://doi.org/10.1061/(ASCE)CP.1943-5487.0000525.

[37] C. Liu, S. Shirowzhan, S.M.E. Sepasgozar, A. Kaboli, Evaluation of classical operators and fuzzy logic algorithms for edge detection of panels at exterior cladding of buildings, Buildings. 9 (2019). https://doi.org/10.3390/buildings9020040.

[38] P. Martinez, R. Ahmad, M. Al-Hussein, A vision-based system for pre-inspection of steel frame manufacturing, Automation in Construction. 97 (2019) pp. 151–163. https://doi.org/10.1016/j.autcon.2018.10.021.

[39] S. Bruno, M. De Fino, F. Fatiguso, Historic Building Information Modelling: performance assessment for diagnosis-aided information modelling and management, Automation in Construction. 86 (2018) pp. 256–276. https://doi.org/10.1016/j.autcon.2017.11.009.

[40] F.P. Rahimian, S. Seyedzadeh, S. Oliver, S. Rodriguez, N. Dawood, On-demand monitoring of construction projects through a game-like hybrid application of BIM and machine learning, Automation in Construction. 110 (2020). https://doi.org/10.1016/j.autcon.2019.103012.

[41] H. Hamledari, B. McCabe, S. Davari, Automated computer vision-based detection of components of under-construction indoor partitions, Automation in Construction. 74 (2017) pp. 78–94. https://doi.org/10.1016/j.autcon.2016.11.009.

[42] E. Valero, F. Bosché, A. Forster, Automatic segmentation of 3D point clouds of rubble masonry walls, and its application to building surveying, repair and maintenance, Automation in Construction. 96 (2018) pp. 29–39. https://doi.org/10.1016/j.autcon.2018.08.018.

[43] W. Fang, L. Ding, B. Zhong, P.E.D. Love, H. Luo, Automated detection of workers and heavy equipment on construction sites: A convolutional neural network approach, Advanced Engineering Informatics. 37 (2018) pp. 139–149. https://doi.org/10.1016/j.aei.2018.05.003.

[44] C. Kropp, C. Koch, M. König, Interior construction state recognition with 4D BIM registered image sequences, Automation in Construction. 86 (2018) pp. 11–32. https://doi.org/10.1016/j.autcon.2017.10.027.

[45] H. Kim, H. Kim, Y.W. Hong, H. Byun, Detecting Construction Equipment Using a Region-Based Fully Convolutional Network and Transfer Learning, Journal of Computing in Civil Engineering. 32 (2018). https://doi.org/10.1061/(ASCE)CP.1943-5487.0000731.

[46] A. Adán, B. Quintana, S.A. Prieto, F. Bosché, Scan-to-BIM for ‘secondary’ building components, Advanced Engineering Informatics. 37 (2018) pp. 119–138. https://doi.org/10.1016/j.aei.2018.05.001.

[47] H. Fathi, F. Dai, M. Lourakis, Automated as-built 3D reconstruction of civil infrastructure using computer vision: Achievements, opportunities, and challenges, Advanced Engineering Informatics. 29 (2015) pp. 149–161. https://doi.org/10.1016/j.aei.2015.01.012.

[48] R. Richter, J. Döllner, Concepts and techniques for integration, analysis and visualization of massive 3D point clouds, Computers, Environment and Urban Systems. 45 (2014) pp. 114–124.

https://doi.org/10.1016/j.compenvurbsys.2013.07.004.

[49] B. Mičušík, J. Košecká, Multi-view Superpixel Stereo in Urban Environments, International Journal of Computer Vision. 89 (2010) pp. 106–119. https://doi.org/10.1007/s11263-010-0327-9.

[50] C. Mura, O. Mattausch, A. Jaspe Villanueva, E. Gobbetti, R. Pajarola, Automatic room detection and reconstruction in cluttered indoor environments with complex room layouts, Computers & Graphics. 44 (2014) pp. 20–32. https://doi.org/10.1016/j.cag.2014.07.005.

[51] R. Albatici, A.M. Tonelli, M. Chiogna, A comprehensive experimental approach for the validation of quantitative infrared thermography in the evaluation of building thermal transmittance, Applied Energy. 141 (2015) pp. 218–228. https://doi.org/10.1016/j.apenergy.2014.12.035.

[52] P. Moulon, P. Monasse, R. Perrot, R. Marlet, OpenMVG: Open multiple view geometry, in: B. Kerautret, M. Colom, P. Monasse (Eds.), Workshop on Reproducible Research in Pattern Recognition, Springer International Publishing, 2017: pp. 60–74. https://doi.org/10.1007/978-3-319-56414-2_5.

[53] D.G. Lowe, Distinctive image features from scale-invariant keypoints, International Journal of Computer Vision. 60 (2004) pp. 91–110. https://doi.org/10.1023/B:VISI.0000029664.99615.94.

[54] J. Cheng, C. Leng, J. Wu, H. Cui, H. Lu, Fast and accurate image matching with cascade hashing for 3D reconstruction, in: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2014: pp. 1–8. https://doi.org/10.1109/CVPR.2014.8.

[55] P. Moulon, P. Monasse, R. Marlet, Adaptive structure from motion with a contrario model estimation, in: K.M. Lee, Y. Matsushita, J.M. Rehg, Z. Hu (Eds.), Computer Vision – ACCV

2012. ACCV 2012. Lecture Notes in Computer Science, 2013: pp. 257–270. https://doi.org/10.1007/978-3-642-37447-0_20.

[56] Y. Furukawa, J. Ponce, Accurate, dense, and robust multiview stereopsis, IEEE Transactions on Pattern Analysis and Machine Intelligence. 32 (2010) pp. 1362–1376. https://doi.org/10.1109/TPAMI.2009.161.

[57] S. Umeyama, Least-Squares Estimation of Transformation Parameters Between Two Point Patterns, IEEE Transactions on Pattern Analysis and Machine Intelligence. 13 (1991) pp. 376–380. https://doi.org/10.1109/34.88573.

[58] J.A. Nelder, R. Mead, A Simplex Method for Function Minimization, The Computer Journal. 7 (1965) pp. 308–313. https://doi.org/10.1093/comjnl/7.4.308.

[59] I. Nardi, D. Paoletti, D. Ambrosini, T. De Rubeis, S. Sfarra, U-value assessment by infrared thermography: A comparison of different calculation methods in a Guarded Hot Box, Energy and Buildings. 122 (2016) pp. 211–221. https://doi.org/10.1016/j.enbuild.2016.04.017.

[60] P.A. Fokaides, S.A. Kalogirou, Application of infrared thermography for the determination of the overall heat transfer coefficient (U-Value) in building envelopes, Applied Energy. 88 (2011) pp. 4358–4365. https://doi.org/10.1016/j.apenergy.2011.05.014.

[61] L. Brackney, A. Parker, K. Macumber, D.Benne, The OpenStudio Software Development Kit, in: Building Energy Modeling with OpenStudio, Springer, Cham, 2018: pp. 287–314. https://doi.org/https://doi.org/10.1007/978-3-319-77809-9_9.

[62] M. Sayin, A. Tavukcuoglu, Cephelerin Isi Yalitimlilik Durumlarinin Isil Görüntüleme İle Değerlendirilmesi (Evaluation of Thermal Insulation of Facades Using Infrared Imaging), Yalitim, Is Dunyasi Yayincilik, Istanbul. 152 (2016) pp. 46–54. http://www.yalitim.net/edergi/18/152/index.html (Date of last access: December 26, 2019).

[63] J. Redmon, S. Divvala, R. Girshick, A. Farhadi, You only look once: Unified, real-time object detection, in: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2016: pp. 779–788. https://doi.org/10.1109/CVPR.2016.91.

[64] T.Y. Lin, P. Goyal, R. Girshick, K. He, P. Dollar, Focal Loss for Dense Object Detection, in: Proceedings of the IEEE International Conference on Computer Vision, 2017: pp. 2999–3007. https://doi.org/10.1109/ICCV.2017.324.

[65] H. Durrant-Whyte, T. Bailey, Simultaneous localization and mapping: Part I, IEEE Robotics and Automation Magazine. 13 (2006) pp. 99–108. https://doi.org/10.1109/MRA.2006.1638022.