Using Passive Multi-Modal Sensor Data for Thermal Simulation of Urban Surfaces

Dimitri Bulatov[1, *], Dirk Frommholz[2], Benedikt Kottler[1], Kevin Qui[1], Eva Strauss[1]

[1] Fraunhofer IOSB, Ettlingen, Germany - (dimitri.bulatov, kevin.qui, benedikt.kottler, eva.strauss)@iosb.fraunhofer.de
[2] DLR Institute of Optical Sensor System, Berlin, Germany - dirk.frommholz@dlr.de

KEY WORDS: 3D Reconstruction, Digital Twin, Land-cover Classification, Multi-sensor Data, Oblique Imagery, Urban Heat

ABSTRACT:

This paper showcases an integrated workflow hinged on passive airborne multi-modal sensor data for the simulation of the thermal behavior of built-up areas with a focus on urban heat islands. The geometry of the underlying parametrized model, or digital twin, is derived from high-resolution nadir and oblique RGB, near-infrared and thermal infrared imagery. The captured bitmaps get photogrammetrically processed into comprehensive surface models, terrain, dense 3D point clouds and true-ortho mosaics. Building geometries are reconstructed from the projected point sets with procedures presupposing outlining, analysis of roof and façade details, triangulation, and texturing mapping. For thermal simulation, the composition of the ground is determined using supervised machine learning based on a modified multi-modal DeepLab v3+ architecture. Vegetation is retrieved as individual trees and larger tree regions to be added to the meshed terrain. Building materials are assigned from the available visual, infrared and surface planarity information as well as publicly available references. With actual weather data, surface temperatures can be calculated for any period of time by evaluating conductive, convective, radiative and emissive energy fluxes for triangular layers congruent to the faces of the modeled scene. Results on a sample dataset of the Moabit district in Berlin, Germany, showed the ability of the simulator to output surface temperatures of relatively large datasets efficiently. Compared to the thermal infrared images, several insufficiencies in terms of data and model caused occasional deviations between measured and simulated temperatures. For some of these shortcomings, improvement suggestions within future work are presented.

1. INTRODUCTION

Urban heat islands (UHI) are a well-known phenomenon of larger cities. They can be characterized as zones of a higher temperature compared to the rural surroundings or, on a smaller scale, the non-built-up parts within the urban area itself. UHIs are significantly caused by the artificial materials used for man-made installations and their geometry. These factors affect the thermal properties compared to natural lands such as heat absorption, conductivity and emissivity often boosting temperatures locally. Further, worldwide and regional climate prospects suggest extreme heatwave events to become increasingly frequent, more intense, and longer (Silva et al., 2022). Especially within the UHIs, the health of (human) lifeforms will be likely impacted due to heat exhaustion in the warm season. Some heat mitigation measures, such as installing energy-consuming air conditioning units, would exacerbate the energy balance, other, like resorting to blue and green infrastructure, would be, without any doubt, costly and not easily implementable.

In order to assess the UHI potential of an urban site, mapping and simulation turn out to be helpful tools to review the current thermal response of the area under examination and predict its temperature characteristics in the future. With recent advances in remote sensing, these analyses could be performed without the installation of in-situ sensors and monitoring their outputs over a long period of time. Instead, digital twins, i.e., comprehensive virtual models of buildings, their surroundings, and even objects that do not exist yet will provide relevant numerical output depending on input parameters. In case of the thermal behavior, the location of the scene, its geometry, materials, and environmental conditions could be used to estimate the local temperature for a dense set of positions. Still, obtaining the input quantities will only be possible through multi-source data processing that yields the necessary terrain slope, land cover, building models, and their surface materials.

1.1 Previous research

There is plenty of research going on that deals with modeling the thermal performance of city areas. The work of (Bruse and Fleer, 1998) on which the commercial software ENVI-met is based on, solves multiple differential equations of physical models for the temperature, humidity, air flow, soil, radiative fluxes, surfaces etc. to extensively simulate the environment of urban scenes. Buildings and vegetation are represented through box models with manually assigned materials and customizable plant prototypes respectively. These items have to be either manually digitized and placed on a grid map using true-ortho mosaics as backdrop images or semi-automatically derived from geo-referenced 2.5D vector shapes. In (Guo et al., 2018), the scene geometry for thermal simulation is obtained from multi-view stereo (MVS) imagery. The produced meshes are classified, turned into planar proxies, and subsequently used to erect simplified 3D models in different levels of detail (LOD) depending on the object semantics. Further, the triangular model faces are assigned materials. To eventually obtain the thermal distribution for the area, both general and polygon-specific layered heat balance schemes are evaluated including sun exposure. The results are represented as thermal textures attached to the simplified object geometry of the chosen LOD. Thus, the multiresolution approach should work in small and large urban setups although simulation accuracy will suffer when scene details are successively left out. Similarly to the previous concept, (Xiong et al., 2016) produce simplified 3D models from MVS data of comprehensive scenes to determine their thermal beha-

* Corresponding author

vior. However, instead of utilizing planar proxy objects, the authors directly apply topology-based simplification to obtain sparse watertight meshes from the original detailed polygonal scene geometry. Radiative simulation is then conducted on the elements of the subdivision surfaces of the 3D models, and the resulting temperatures are stored as texture bitmaps to facilitate visualization. In (Kottler et al., 2019), thermal flux simulation for large areas is founded on digital elevation models and land cover classification bitmaps for the terrain material properties which both are derived from heterogeneous remote sensing data. The 3D surface geometry of the scene to be investigated is adaptively reduced by a restricted quadtree triangulation algorithm as an intrinsic LOD mechanism. Buildings are separately represented by prismatic shapes. The prisms result from the extrusion of publicly available 2D vector data intersected with the height information. Temperatures over time are eventually computed for a stack of thermal layers assigned to each mesh triangle and output as texture maps. The underlying physical model also considers environmental factors such as solar irradiation, air humidity and atmospheric damping. Aside from built-up areas and UHI phenomena, (Maréchal et al., 2010) focus on rendering realistic winter sceneries over time. The landscape to be visualized is modeled as a sparse set of discrete volumes, i.e., voxels, for which the radiative heat transfers and phase change processes like freezing and melting are evaluated. The final scene is retrieved from the voxel grid information as a textured mesh to which a vertical displacement is added to account for snowfall. Realistic visual appearance is increased through the application of a perturbance function.

1.2 Contribution

This paper presents a comprehensive processing pipeline from raw aerial sensor data to the thermal simulation of an urban scene considering potential urban heat islands. Its particular contribution is that only passive nadir and oblique cameras in both the visual and infrared spectral ranges have been used to capture the scene and retrieve important parameters for temperature analysis over time. While land cover classification relies on the combination of photogrammetrically derived 2D and 2.5D datasets, dense 3D point clouds also for the façades are used to reconstruct buildings including wall surfaces with window openings and airborne textures. Further, single trees are detected and separately represented, and forests get approximated to a higher degree in order to reduce simulation complexity. Given the digital twin of the target area, a computationally-efficient thermal model is employed to calculate the surface temperatures on the desired temporal resolution, usually in the range of seconds. For the sample scene of Moabit, a district of the German capital Berlin, simulation outputs were stored for every 60 minutes within a time-frame of 33 hours while they revealed a reasonable deviation compared to the available long-wave infrared reference. In this spirit, the workflow to be described picks up ideas from previous contributions. However, it differs in the special sensor setup, degree of automatization, geometric level of detail and the selected deep learning approaches applied for land cover classification.

2. DATASET

For the setup of the digital twin for thermal simulation, an airborne dataset of the Moabit district of Berlin, Germany has been selected. The urban target area spans roughly 500 by 720 m (0.36 km^2) and is located in the temperate climate zone. The scene comprises multi-storey apartment blocks, commercial buildings such as shopping centers and a few industrial compounds with a low proportion of green space. Most structures were built throughout the 20th century and - due to the eventful history of the German capital - expose various architectural styles. Because of these factors, Moabit cannot be considered to be adequately prepared to prolonged heat waves or periods of drought and hence qualifies for thermal and UHI analyses over time. In fact, phases of excessive temperatures occurred recently during the summer months of the years 2019 to 2022 (Helmholtz UFZ, 2023). For the Berlin metropolitan area, July 2019 exposed peak temperatures that were almost 5 °C above the average of the hottest days from 1980 to 2010 (Bissolli et al., 2019). Because of high financial costs of protection against extreme weather events, mapping the local urban thermal response is of utmost importance.

2.1 Aerial survey

Data for the chosen scene of Berlin-Moabit originates from two aerial surveys that were conducted on August 14, 2017 and March 5-6, 2019. For the summer campaign, RGB and near-infrared (NIR) imagery were simultaneously captured with the DLR MACS-HALE camera (Brauchle et al., 2015). Effective resolutions for the 16 bit nadir and oblique bitmaps taken at an off-nadir angle of 35 degrees to the left and right to capture the building façades vary from roughly 8 to 50 cm per pixel. These values do not only depend on the sensor tilt but also on the current angle of attack of the aircraft and the orientation of the photographed surface relative to the optical axis of the sensor.

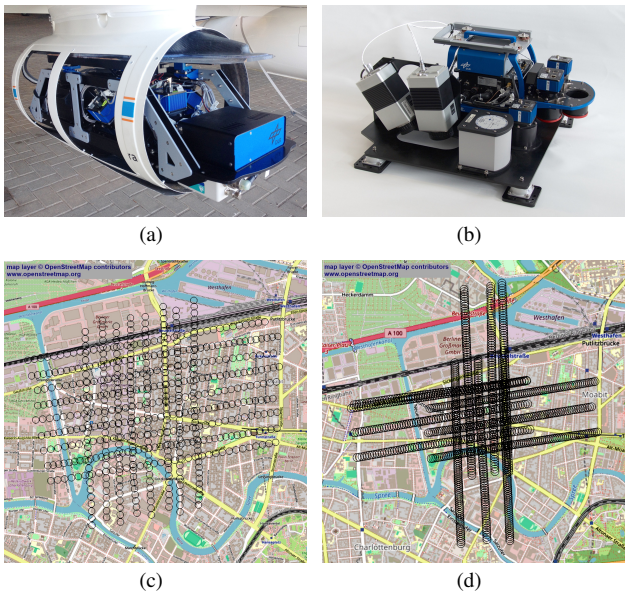
The winter survey was carried out with two Infratec VarioCam HD thermal infrared (TIR) sensors around midnight under partly cloudy conditions to assess the building insulation status in the first place. However, the obtained data may also serve as a baseline for thermal simulation because temperature disturbances due to sunshine, air flow, running combustion engines, heat radiators inside the buildings (which could be expected to have cooled down already) and the infrared albedo are minimal. The two sensors were pointing downwards and forward at an off-nadir angle of 30 degrees to capture the thermal signature of the façades like for the RGB data. The actual on-surface resolutions of the recorded 16 bit graylevel bitmaps storing absolute temperatures range from approximately 35 cm to 2 m per pixel. Figure 1 depicts the cameras deployed and the cross-style flight patterns which allow to capture any side of the buildings given the available oblique viewing directions.

2.2 Image preprocessing

Both datasets were preprocessed regarding demosaicking, dark signal non-uniformity correction, photo response non-uniformity (PRNU) compensation, and geometric distortion adjustments. While the necessary dark images could be simply recorded with lens covers on, the inhomogeneity shots for the PRNU were taken in front of a flatfield. Radial distortion coefficients got derived with a laser-based calibration system (Dahlke et al., 2019). Further, the initial image orientation records from the GPS receivers and inertial measurement units (IMU) built into the cameras underwent semi-automatic bundle adjustment. Preselection to meet the desired frame overlap of roughly 80% and reduce redundancies resulted in 2329 RGB/NIR and 1669 TIR pinhole bitmaps for the photogrammetric processing pipeline.

2.3 Elevation model and ortho generation

Photogrammetric processing starts with semi-global matching (SGM) (Hirschmüller, 2005) to regain dense height information from the two-dimensional imagery. Pairs of bitmaps from



(a) (b) (c) (d)

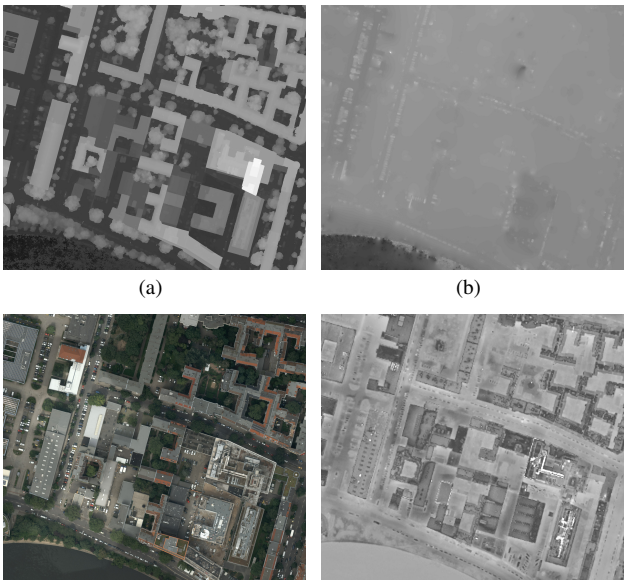
Figure 1. Aerial survey setup (a) DLR MACS HALE (b) TIR cameras, (c) RGB/NIR flight pattern, (d) TIR flight pattern

both the RGB and TIR datasets were analyzed for pixel correspondences. For the low-contrast thermal data, a linear intensity stretch was applied in advance to increase the robustness of the SGM algorithm. Triangulation of the output disparity maps yielded nadir and oblique 3D point clouds. The nadir point clouds further were projected onto the xy plane and binned at ground sampling distances (GSDs) of 10 cm for the RGB and 1 m for the TIR datasets respectively to produce digital surface model (DSM) bitmaps. The DSMs were used to perpendicularly reproject the nadir footage into true-ortho mosaics (TOMs) matching the DSM pixel pitch.

For thermal simulation with separately modeled buildings and vegetation, the surface bitmaps need to be stripped from elevated items to solely retain the terrain information. Raised object detection was performed exclusively for the high-resolution RGB-based DSM according to the algorithm from (Piltz et al., 2016). The method requires only the minimum height of an object to be classified as elevated and its maximum width as parameters to be interactively chosen. Retrieval of the terrain mask is achieved through calculating the normalized volume above ground (NVAG) metric for each pixel of the DSM along horizontal and vertical scanlines. The NAVG of the elevated segments found gets maximized via dynamic programming to yield elevation profiles. The set of profiles is combined into a binary image that identifies the parts to be removed from the surface model. Voids get interpolated from surrounding support pixels with the inverse distance weighting method (Shepard, 1968) to obtain a watertight digital terrain model (DTM) at the resolution of the underlying DSM.

Having the surface and terrain information, the normalized digital surface model (NDSM) is eventually computed by subtracting the DTM from the DSM. The NDSM provides information about the vertical dimensions of structures, which is useful for building reconstruction and land cover classification.

Figure 2 displays geographically aligned samples of the elevation data derived from the high-resolution visual imagery together with the corresponding RGB and TIR true-ortho mosaic tiles.



(a) (b) (c) (d)

Figure 2. Elevation data products and true-ortho mosaics (a) DSM, (b) DTM, (c) RGB true-ortho, (d) TIR true-ortho

2.4 Building reconstruction

Individual structures of the scene were reconstructed with the automated data-driven method as proposed by (Frommholz et al., 2016). This is to yield more recent data than the publicly available city models from 2009 to 2014 (Berlin Partner für Wirtschaft und Technologie GmbH, 2020). In the workflow, the projection of the previously generated oblique 3D point clouds onto the regularly subdivided ground plane undergoes linear regression. The resulting strokes of adjacent grid cells are grouped, extrapolated, and intersected to obtain the building contours. Likewise, reconstruction of the roof planes is accomplished with an incremental two-dimensional principal component analysis on the grid projection of the nadir point clouds preserving their height values. Similarly oriented cell planes are consolidated into larger ones through region growing. The spatial relationship between façade segments and roof planes is established by testing the origin of the items inside a small window shifted over the projection grid cells. To produce the 3D geometry of the structures, the individual façade outlines are extruded perpendicularly to the ground according to the NDSM height values. The resulting wall surfaces are further intersected with the roof planes according to the building topology.

For the windows, the polygonal building hulls get textured from the RGB and TIR imagery. However, due to the higher resolution, only the generated ortho-rectified texture atlases of the visible spectrum were used to detect the openings. The detection process itself exploits empirically chosen color features and the rectangular shape that windows typically come with. By tracing the contours of the window patches found inside the texture bitmaps and mapping them onto the respective building façades, their polygonal representation can be retrieved and added to the structure geometry as overlay elements. Together with the surface types obtained from topology analysis, semantically annotated building models were eventually stored in both CityGML format and as Alias/Wavefront OBJ representations. While the former suits geographic information systems (GIS), the latter facilitates parsing and serves as the input for thermal simulation.

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-X-2-2024-17-2024 | © Author(s) 2024. CC BY 4.0 License.

Figure 3 shows sample buildings of the Berlin-Moabit scene as produced with the reconstruction workflow. For this specific target, a modeling completeness of 97% in the number of structures was achieved. The correctness of the obtained building outlines compared to the footprints of the cadastral map reached a value of 71% in terms of the matching corner points. This is mostly due to contour simplification that goes along with line segment grouping when the façades are determined.

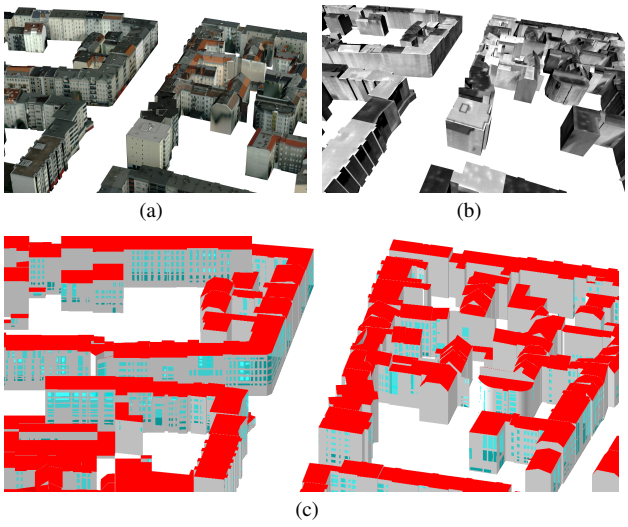


Figure 3. Polygonal building models with (a) RGB textures, (b) TIR textures, (c) surface semantics in different colors

3. METHODOLOGY

This section outlines the necessary parametrization of the remote sensing data to obtain the digital twin in preparation for thermal simulation. Also, it provides the description of the heat transfer model to eventually deliver surface temperatures.

3.1 Preparation of parametric meshes

The task of this subsection is to convert the available remote sensing data together with the results of the building reconstruction procedure into a semantic parametric mesh of the entire scene. This mesh is constituted by three submeshes for the ground, trees, and buildings. To retrieve relevant physical parameters for the thermal simulation of the ground model, a land-cover, or even better, a material class has to be assigned to every ground element. To perform land-cover classification, the machine-learning (ML) pipeline presented in (Qiu et al., 2022) is applied to two radiometrically downscaled 8-bit color images: The first is the RGB orthophoto, and the second is the synthesized image with channels made up by truncated NDSM, NDVI, and planarity map computed from the DSM according to (Gross and Thoennessen, 2006). The contents of the latter two channels are rescaled to integer values between 0 and 255.

Both images are processed by the approach of (Qiu et al., 2022), which is based on the DeepLab v3+ method of (Chen et al., 2018). The ResNet (He et al., 2016) backbone constitutes the encoder. After passing both images through their own first residual block, both output features are averaged and processed by the remaining ResNet blocks together. Both ResNet branches are initialized using the pre-trained weights of ImageNet (Deng et al., 2009). The standard cross-entropy function was chosen as the loss function.

To perform land-cover classification in the supervised way, specification of sufficient data for each of the classes in question is indispensable. In order to spare the sometimes time-consuming task of retrieving training data for a completely new data set, it is often possible to train an ML model on an annotated, freely available dataset such as the ISPRS Potsdam 2D labeling contest dataset (Rottensteiner et al., 2014) and apply it to the new one. Since the NDSMs are derived from absolute, metric units and both NDVI and planarity are indices, one could expect a certain transferability from the second branch. Nevertheless, we annotated several patches of the Moabit dataset to fine-tune the Potsdam model. This will allow us to track the performance gap during evaluation. The classes were the same as in the Potsdam dataset: Building, tree, grass, road, vehicle and clutter. We refer to (Qiu et al., 2022) for further details.

To convert the DTM into triangles, the land cover classes vehicle, tree and clutter are suppressed, since they do not constitute parts of the ground. The pixels corresponding to their classes are inpainted. Buildings are suppressed in an analogous way if their models exist and modeled as brick piles otherwise, mostly on the border of the dataset. The clutter class is subdivided into classes water (the Spree river, crossing Moabit and shown in the bottom-left corner of Figure 2, is segmented manually), earth (larger connected components) and noise (smaller connected components, suppressed and inpainted). Before and after these steps, morphological processing takes place to suppress too small regions and smooth the classes’ boundaries. To this post-processed result, the *restricted (top-down) quadtree triangulation* (RQT or RTDQT) is applied. It presupposes an iterative subdivision of too large and inhomogeneous (in terms of either elevation or classes) triangles. To guarantee a consistent mesh, adjacent triangles must also be subdivided, which sometimes leads to the necessity to add new vertices. Differently to (Bulatov et al., 2020), the number of iterations, or pyramids, for subdivision is set to four (a lower number) to avoid unrealistic shadow courses.

Regarding the preparation of tree models, the tree detection workflow of (Bulatov et al., 2016) has been employed to detect single trees by means of connected component analysis and watershed transformation, as well as to derive their heights and positions. Individual trees and bushes have been modeled using several predefined watertight models with ≈ 150 faces whereby detection of adjacent trees took only place in connected regions corresponding to the tree class whose areas did not exceed 640 m^2. For larger regions, the technique involving forest boxes has been employed, similarly to (Kottler et al., 2019).

In order to prepare the building models for the simulation, we need to guarantee that the building façades consist of adjacent but spatially disjoint triangles that can be assigned a unique material. Since the façade components are given in the form of overlaying planar polygons, they initially will be processed by the Constrained Delaunay Triangulation (CDT) approach (Domiter and Žalik, 2008). The CDT preserves the boundary edges of the building surfaces and openings. Further, the algorithm combines the separately modeled shapes into sets of three-sided faces while maximizing their minimum angle and maintaining the original semantic designation. Prior to triangulation, two important steps must be carried out. First, since the CDT takes place in a plane, the 3D coordinates of vertices of every single wall must be transformed to the xy plane using a suitable embedding transformation. Second, façade elements (e.g. windows) incident to this wall are assigned. Hereby, not only the corresponding plane equations of a façade and window

must coincide, but the window must also be completely contained in the wall. Figure 4 exemplifies a situation where the first condition is insufficient. We see three advanced wall units that have the same plane. Each of these units has a few façade elements within the same plane. Assigning façade elements to a wall unit only because they have the same plane and performing CDT would result in many spurious edges.

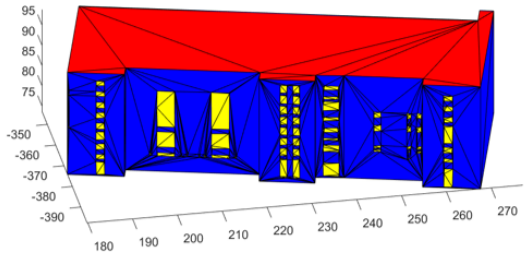


Figure 4. Building model prepared for simulation with different colors for different semantic classes. As a particularity, the plane of all windows is incident with those of more than one wall.

To assign materials of building roofs and walls, all available raster data were used. For the roofs, segment-based features were collected from the sensor data: raw image channels, including near-infrared, entropy over two different kernels, and planarity. For each of the three classes *metal*, *tile* and *concrete*, five to ten roof samples were selected interactively. The classifier employed was Random Forest with five decision trees (Breiman, 2001). The color needed to compute the albedo values was determined using kMeans with $k = 4$ clusters. Also for the façades, only an annotation of materials of a few selected buildings took place. In the absence of ground measurements, automatic annotation would be impossible, and interactive too cumbersome. Finally, the necessary over-simplification – one building, one wall material – was made. The materials in question are *stucco* and *firewall*; Façade elements like doors and windows are assumed to be made of glass.

To all classes, namely, asphalt (road), water, soil, grass, tree trunks and crowns, as well as the aforementioned roof and wall materials, we assigned the physical parameters: Density, thermal conductivity, specific heat capacity, emissivity and albedo. Thus, there is only one parameter vector for one material, leading to a quite simplified but efficiently prepared model.

3.2 Temperature simulation

Following previous work from (Strauss and Bulatov, 2024), the temporal evolution of the surface temperature T is calculated for each triangle of the scene using the heat transfer equation

$$\rho c_v d \frac{\partial T}{\partial t} = q(T) = A + R + S + D + I, \qquad (1)$$

with the temperature T of the triangle, the density ρ and specific heat capacity c_v of the material assigned to the triangle, the virtual depth of the triangle d and the heat flux $q(T)$, i.e., heat per triangle area, which needs to be modeled as described in the next paragraph. Note that differential equation (1) basically represents the one-dimensional case and is solved with the Finite Volume Method. The reason is to avoid time-consuming computations of temperature and wind velocity fields on volumetric grids, as it is the case for many RANS simulations (Afshari and Ramirez, 2021) or similar computational fluid dynamics approaches. Instead, we pursue temperature calculation along the surface. However, the surface itself is three-dimensional, which is important for shadow analysis. Besides, with the virtual depth, the surface triangle is virtually expanded in inverse surface normal direction to convert it into a volumetric element.

The right-hand side of (1), $q(T)$, is the sum of conductive, convective, and radiative heat transfers, which we review here briefly. Conduction along the surface normal (I) implicates defining the inner boundary condition depending on the semantic class: building, tree, and earth. Except for buildings involving insulation, i.e. zero heat flux as boundary condition, constant core temperatures of these elements are determined empirically. The second type of conduction (D), perpendicular to the surface normal, requires the knowledge about triangle neighborhoods, stored in the so-called adjacency matrix. With neighborhoods, the computation burden becomes more expensive. Therefore, the term D is often dropped, as it was in our case. Convection (A) is the transfer of heat from one place to another due to particle motion in fluids or gases. Here, important parameters are wind velocity and air temperature. Simplifications are made so these parameters are constant locally while varying temporarily. The radiative transfer includes solar heating (S) and atmospheric heat exchange (R). The former considers the date, time, and location of the scenery, allowing to compute the relative position of the sun and to perform occlusion analysis. Occlusion analysis is performed during the day and for non-overcast sky only. Cloud coverage is retrieved from historic weather data of the period being simulated together with the air temperature and wind speed needed for convection. The latter depends on the material-dependent thermal emissivity and the temperature of the atmosphere, which under a few simplifying assumptions, is a function of air temperature.

4. RESULTS

4.1 Land-cover classification

Regarding the performance evaluation of land-cover classification, both the original (i.e., purely image-based) and the modified DeeplabV3+ models were used. Also, the effects of fine-tuning these two models were investigated using 28 labeled image patches of 512×512 pixels for training and 12 tiles for validation. Subdivision was performed with an empirically chosen but fixed setting throughout the series of experiments. As performance metrics, overall accuracy (OA), Kappa coefficient (κ), and the average F1-score (F1) were considered. The original method of (Chen et al., 2018) using the Potsdam semantic labeling model yielded 71.10% on OA, 60.49% on κ and 61.22% on F1. Employing the second branch which also incorporates elevation data improved the results only insignificantly. The increase on κ barely exceeded 1% (61.81% in total). Once fine-tuning took place, there were considerable improvements with 89.76/91.96% on OA, 84.90/88.13% on κ and 78.42/81.24% on F1 for the image-only path and second branch respectively.

These results promote the interpretation that training data from the dataset to be investigated tend to be more crucial for proper classification than sophisticated features. The minor influence of the 3D information could be due to different acquisition and processing techniques employed for the Potsdam benchmark. For the fine-tuned models, the additional data shows some improvement in the performance readings, however, to a less significant extent than expected. Thus, the receptive field of the neural network and the training data are apparently sufficient to retrieve the most relevant information from images only.

Figure 5 depicts qualitative observations. Since the slightly red-tinted asphalt in the Moabit scene resembles open earth regions in the Potsdam dataset, the clutter class appears to be quite dominant for the old model as shown in the rightmost image. Clutter regions are also found on building roofs although building detection mostly is successful. Trees preserve their 2D shapes, too. The model trained with the few Moabit patches does not exhibit significant misclassifications and maintains correct results even in shadow areas. As shortcomings, building boundaries appear noisy and should be generalized. In the upper image row, clutter did not get adequately predicted. In the bottom strip, larger vehicles were detected despite the presence of occlusions by trees, but oversmoothing appears for the numerous bikes. This glitch however will probably not impact the temperature distribution calculation hinged on the digital twin.

4.2 Temperature distribution

The parametrized reconstructed scene consists of 959730 triangles of seven semantic classes where some semantic classes (buildings, trees) can be assigned multiple material classes. A building, for example, may come with several roof and wall materials while a tree consists of a trunk and a crown. To reflect the Berlin scenery adequately, new materials had to be added to the existing material database, particularly for the openings and roof covering. This includes glass (windows), wood (doors), and concrete (roof) as a generalization since the mentioned previous studies have confirmed that thermal model still will produce feasible outputs.

Given weather data for the period of the remote sensing campaign, i.e., from March 3rd, 2019, at 11 p.m. till March 6th, 2019, at 8 a.m., thermal simulation was run with a temporal sampling of one hour to obtain surface temperatures. For this time-frame, point measurements based on a handheld thermal camera and a contact thermometer as well as the aerial thermal infrared images encoding absolute temperatures in Kelvin are available for a comparison. To check for urban heat islands in Berlin-Moabit, a second simulation was carried out for the period from August 14th, 2023 at 12 a.m. till August, 16th, 2023 at 12 a.m., i.e., two days during European summer, and contrasted to official temperature measurements.

The temporal resolution for both simulations was set to 0.1 s, and cubic spline interpolation was applied to match the hourly raster of weather data. Aside from physical quantities, weather data also includes an informal description of the cloud cover, e.g. ’clear’, ’cloudy’, or ’overcast’. Since the simulation approach requires a numerical representation of this terminology, percentages reflecting the conditions have been assigned with 0 denoting clear skies and 1 indicating a closed cloud cover. The latter value will also be taken when precipitation is present. To overcome thermal inertia, a precalculation of temperatures per material using the given weather data was performed following (Strauss and Bulatov, 2024) for both time intervals.

To evaluate the results for the winter scenario simulation against the measured temperature image, the three-dimensional simulation results were orthogonally projected as shown in Figure 6. The final mean absolute deviation is 9.35 K, which is a somehow discouraging result at the first glance. There are several reasons that contribute to these deviations. Firstly, the registration of the orthophoto, and thus the final simulation results, and the thermal infrared image encoding the temperatures in Kelvin, yielded a non-negligible deviation, causing a systematic error in the average deviation between measured and simulated temperatures. Secondly, the measured thermal imagery shows some artifacts caused by the image generation process (stitching artifacts, upsampling), hot spots (e.g. street lights, ventilation exhausts), or change in cloud coverage. The latter strongly affects the surface temperature as well as the radiative temperature measured by infrared cameras. That the effect of infrared cameras itself is not considered within the simulation, contributes to the general tendency of temperature overestimation. Thirdly, the vegetative and convective models seem to be over-simplified so that, during the night, the temperature basically becomes a function of the material class. As a consequence of few materials, the temperature distribution is quite homogeneous, as well. Finally, as in the top right area of the simulated images, the deviations from the infrared signature are particularly large, one could have a reason to assume that either the material classification of roofs or its level of abstraction (one material per roof) are not sufficiently accurate because in reality, chimneys, solar-panels, and other roof elements are abundant. At the second glance, we see reasonably similar temperatures in the areas of water, metallic and concrete roofs, and paved roads.

To compare the 3D simulation results to the point measurements, first their UTM coordinates were transferred into the local Cartesian coordinate system of the simulation model. Due to uncertainties in the coordinates, an exact replication of the locations in the simulation was not possible. Additionally, the two temperature measurements at the ground locations differed strongly with a mean difference of 8.84 K. Thus, when comparing against the simulation results, the mean deviation is 13.6 K (radiative temperature of handheld camera) and 4.8 K (contact thermometer), respectively. The latter is in line with findings of previous studies (Strauß and Bulatov, 2022).

Two 3D views of simulation results of the summer and winter scenario are shown in Figure 7. During summer, surface temperatures of pavement and concrete roofs remain relatively hot even in the evening hours although the air temperature falls by 4.5 K. In the winter scenario, surfaces cool down more quickly despite a lower decrease of air temperature of only 1.3 K.

5. CONCLUSION

This paper has outlined an integrated workflow on how to combine multi-modal remote sensing data for thermal simulation of large urban areas. The input data was entirely acquired by passive airborne sensors. From the nadir and oblique imagery, 3D point clouds, DSMs, and true-orthomosaics were obtained. Building reconstruction included outlining, analysis of roof details and façade elements, as well as tessellation using the CDT and assigning roof materials and colors. The orthophoto was subject to land-cover classification, whereby encouraging results of over 90% on overall accuracy could be obtained. However, we had to use a few annotated patches in the dataset at hand because the accuracy obtained after direct application of the previously trained model was significantly lower. The rasterized result was compressed and decomposed into triangles. To triangle meshes from building and terrain, meshes of individual trees and larger tree regions were added and the corresponding physical parameters, such as density or emissivity, were retrieved. With actual weather data, temperatures of every triangle could be simulated with a temporal discretization of one second using an efficient realization of the one-dimensional heat transfer equation.

One of the directions of future research is to tackle the performance gap for land-cover classification. Once a suitable technique for domain adaptation has been identified, much better

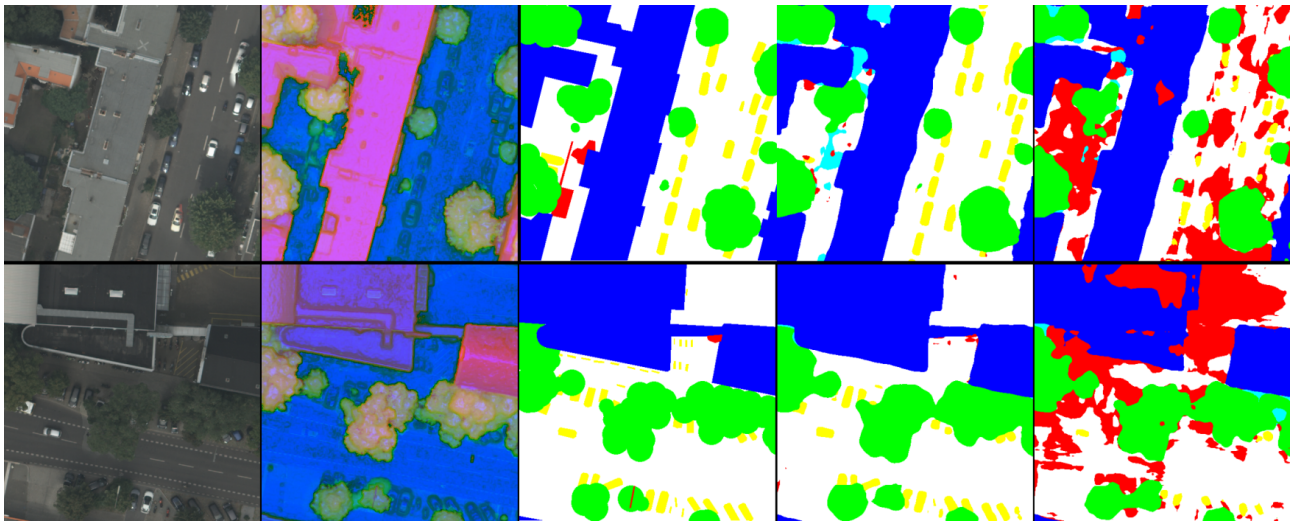


Figure 5. Input and output of land cover classification for two Berlin-Moabit fragments. Left to right: RGB TOM (first branch), NDSM-NDVI-planarity image (second branch), interactively labeled ground truth, result of classification with and without referenced patches of the Moabit dataset.

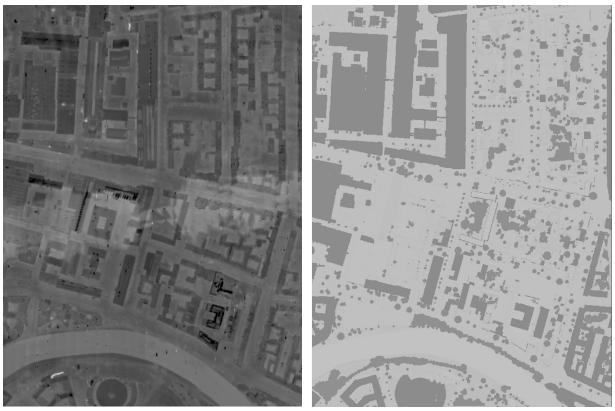


Figure 6. Measured thermal infrared image (left) and corresponding simulation result (right).

results on land-cover classification can be obtained without the need for a labor-intensive data annotation process. Besides, in future versions of the algorithm, we plan to assess the temperatures of the façade triangles using the texture maps obtained from oblique infrared images. One problem to address here is the presence of long and narrow triangles regularly resulting from the CDT. Finally, for the thermal simulation, improved models for convection and vegetation would contribute to a better identification of areas where the air gets trapped and UHIs are likely to appear.

AKNOWLEDGEMENTS

The authors would like to thank Aleksandar Atanasov from Fraunhofer IOSB for programming the interface for weather data extraction. On the part of DLR Optical Sensor Systems, kudos goes to Magdalena Linkiewicz, Dennis Dahlke and Sebastian Pless for their contribution to the photogrammetric processing workflow, Matthias Geßner for his nightshift as the TIR camera operator and the MACS hardware development group.

REFERENCES

Afshari, A., Ramirez, N., 2021. Improving the accuracy of simplified urban canopy models for arid regions using site-specific prior information. *Urban Climate*, 35, 100722.

Berlin Partner für Wirtschaft und Technologie GmbH, 2020. LoD2 Gebäudedaten Berlin. https://daten.berlin.de/datensaetze/lod2-gebäudedaten-berlin (11.4.2024).

Bissolli, P., Deutschländer, T., Imbery, F., Haeseler, S., Lefebvre, C., Blahak, J., Fleckenstein, R., Breyer, J., Rocek, M., Kreienkamp, F., Rösner, S., Schreiber, K.-J., 2019. Hitzewelle Juli 2019 in Westeuropa – neuer nationaler Rekord in Deutschland. https://www.dwd.de/DE/leistungen/besondereereignisse/temperatur/20190801_hitzerekord_juli2019.html (11.04.2024).

Brauchle, J., Hein, D., Berger, R., 2015. Detailed and highly accurate 3D models of high mountain areas by the MACS-Himalaya aerial camera platform. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XL-7/W3, 1129–1136.

Breiman, L., 2001. Random forests. *Machine learning*, 45(1), 5–32.

Bruse, M., Fleer, H., 1998. Simulating surface–plant–air interactions inside urban environments with a three dimensional numerical model. *Environmental Modelling & Software*, 13(3-4), 373–384.

Bulatov, D., Burkard, E., Ilehag, R., Kottler, B., Helmholz, P., 2020. From multi-sensor aerial data to thermal and infrared simulation of semantic 3D models: Towards identification of urban heat islands. *Infrared Physics & Technology*, 105, 103233.

Bulatov, D., Wayand, I., Schilling, H., 2016. Automatic tree-crown detection in challenging scenarios. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, 41, 575–582.

Chen, L., Zhu, Y., Papandreou, G., Schroff, F., Adam, H., 2018. Encoder-Decoder with atrous separable convolution for semantic image segmentation. *CoRR*, abs/1802.02611. http://arxiv.org/abs/1802.02611.

Dahlke, D., Geßner, M., Meißner, H., Stebner, K., Grießbach, D., Berger, R., Börner, A., 2019. Calibrating photogrammetric airborne camera systems with diffractive optical elements. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLII-2/W13, 1637–1642.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., Fei-Fei, L., 2009. Imagenet: A large-scale hierarchical image database. *IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, 248–255.

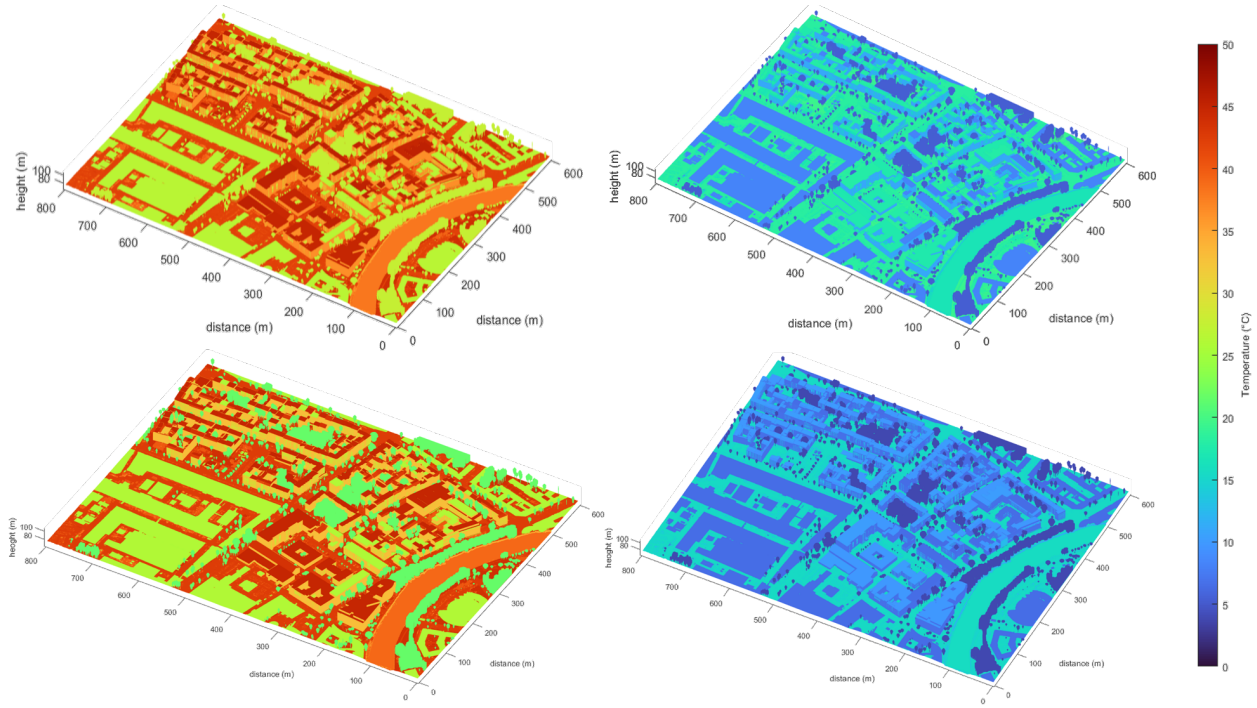


Figure 7. Simulation results for summer (left) and winter weather (right), 5 p.m. (top) and 9 p.m. (bottom) respectively.

Domiter, V., Žalik, B., 2008. Sweep-line algorithm for constrained Delaunay triangulation. *International Journal of Geographical Information Science*, 22(4), 449-462.

Frommholz, D., Linkiewicz, M., Poznanska, A. M., 2016. Inlining 3D reconstruction, multi-source texture mapping and semantic analysis using oblique aerial imagery. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLI-B3, 605–612.

Gross, H., Thoennessen, U., 2006. Extraction of lines from laser point clouds. *Symposium of ISPRS Commission III: Photogrammetric Computer Vision PCV06. Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, 36(Part 3), 86–91.

Guo, S., Xiong, X., Liu, Z., Bai, X., Zhou, F., 2018. Infrared simulation of large-scale urban scene through LOD. *Optics express*, 26(18), 23980–24002.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 770–778.

Helmholtz UFZ, 2023. Dürremonitor Deutschland. https://www.ufz.de/index.php?de=37937 (11.04.2024).

Hirschmüller, H., 2005. Accurate and efficient stereo processing by semi-global matching and mutual information. *IEEE Conference on Computer Vision and Pattern Recognition*, 2, IEEE, 807–814.

Kottler, B., Burkard, E., Bulatov, D., Haraké, L., 2019. Physically-based thermal simulation of large scenes for infrared imaging. *International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*, INSTICC, 53–64.

Maréchal, N., Guérin, E., Galin, E., Mérillou, S., Mérillou, N., 2010. Heat transfer simulation for modeling realistic winter sceneries. *Computer Graphics Forum*, 29(2), 449–458.

Piltz, B., Bayer, S., Poznanska, A. M., 2016. Volume based DTM generation from very high resolution photogrammetric DSMs. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLI-B3, 83–90.

Qiu, K., Budde, L. E., Bulatov, D., Iwaszczuk, D., 2022. Exploring fusion techniques in U-Net and DeepLab V3 architectures for multi-modal land cover classification. *Earth Resources and Environmental Remote Sensing/GIS Applications XIII*, 12268, SPIE, 190–200.

Rottensteiner, F., Sohn, G., Gerke, M., Wegner, J., Breitkopf, U., Jung, J., 2014. Results of the ISPRS benchmark on urban object detection and 3D building reconstruction. *ISPRS Journal of Photogrammetry and Remote Sensing*, 93, 256-271.

Shepard, D., 1968. A two-dimensional interpolation function for irregularly-spaced data. *Proceedings of the 1968 23rd ACM National Conference*, ACM ’68, Association for Computing Machinery, New York City, New York, USA, 517–524.

Silva, R., Carvalho, A. C., Pereira, S. C., Carvalho, D., Rocha, A., 2022. Lisbon urban heat island in future urban and climate scenarios. *Urban Climate*, 44, 101218.

Strauß, E., Bulatov, D., 2022. A region-based machine learning approach for self-diagnosis of a 4D digital thermal twin. *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci.*, 10, 265–272.

Strauss, E., Bulatov, D., 2024. Between gaming and microclimate simulations: Temperature estimation of an urban area. *19th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2024)*.

Xiong, X., Zhou, F., Bai, X., Xue, B., Sun, C., 2016. Semi-automated infrared simulation on real urban scenes based on multi-view images. *Optics Express*, 24(11), 11345–11375.