CLASSIFICATION AND REPRESENTATION OF COMMONLY USED ROOFING MATERIAL USING MULTISENSORIAL AERIAL DATA

R.Ilehag[a,c*], D. Bulatov[b,c], P. Helmholz[c], D. Belton[c]

[a] Institute of Photogrammetry and Remote Sensing, Karlsruhe Institute of Technology, Germany - rebecca.ilehag@kit.edu
[b] Fraunhofer IOSB, Ettlingen, Germany, dimitri.bulatov@iosb.fraunhofer.de
[c] Department of Spatial Sciences, Curtin University, Perth, WA, Australia - (petra.helmholz, d.belton)@curtin.edu.au

Commission I, WG I/6

KEY WORDS: Multispectral, Thermal, High-resolution RGB, LiDAR, Building outlines, Classification, Image segmentation

ABSTRACT:

As more cities are starting to experience the urban heat islands effect, knowledge about the energy emitted from building roofs is of primary importance. Since this energy depends both on roof orientations and materials, we tackled both issues by analysing sensor data from multispectral, thermal infrared, high-resolution RGB, and airborne laser datasets (each with different spatial resolutions) of a council in Perth, Australia. To localise the roofs, we acquired building outlines that had to be updated using the normalised digital surface model, the NDVI and the planarity. Then, we computed a semantic 3D model of the study area, with roof detail analysis being a particular focus. The main objective of this study, however, was to classify three commonly used roofing materials: *Cement tiles*, *Colorbond* and *Zincalume* by combining the multispectral and thermal infrared image bands while the high-resolution RGB dataset was used to provide additional information about the roof texture. Three types of image segmentation approaches were evaluated to assess any differences while performing the material classification; pixel-wise, superpixel-wise and building-wise image segmentation. Due to the limited amount of labelled data, we extended the dataset by labelling data ourselves and merged *Colorbond* and *Zincalume* into one separate class. The supervised classifier Random Forest was applied to all reasonable configurations of segmentation kinds, numbers of classes, and finally, keeping track of the added value of principal component analysis.

1. INTRODUCTION

The land use coverage in cities affects the quality of living and as cities grow, the amount of vegetation is often reduced to make space for buildings. Due to the decreased amount of vegetation, areas within cities that have an increased temperature compared to their surrounding rural areas, so called urban heat islands (Oke, 1982), are on the rise in major cities. Urban heat islands can appear due to the decline of vegetation in urban areas (Weng et al., 2004; Chen et al., 2006) and as these areas are being heated up during the day, the natural cooling system from the surrounding vegetation is reduce or removed during the night. As the global temperature is increasing, such areas rub danger to be even more vulnerable in the future.

Knowledge about materials found on buildings is valuable information for municipalities and authorities when dealing with urban planning. Such information is crucial for city models when a high level of detail is needed, but also for estimates of the anthropocentric inventory and as input to models of the built environment. Information about building materials is also of interest as more cities are trying to reduce the building energy consumption. Additionally, by knowing which roofing materials that are present in areas located within urban heat island, municipalities and authorities can create strategies to counter the effect. This kind of information have a spatial aspect and are therefore often based on remote sensing datasets. Combining data from multiple sensors can provide with further information for urban studies (Dimmeler et al., 2013; Kumar et al., 2015).

The City of Melville, a local council located in Perth, Australia, disposes of multi-sensor aerial data, such as multispectral imagery, a thermal infrared imagery and a Light Detection and Ranging (LiDAR) point cloud, allowing to perform studies closely related to the urban heat island effect and urban vegetation. Studies they wanted to perform included determining the correlation between heat and the land usage, predicting future temperature changes using time series, analysing the shading and cooling properties of both native and exotic tree species, and localising commonly used roofing materials (Council of City of Meville, 2017). This study is closely related to the aforementioned topic and will be presented in this paper.

Virtual simulations enjoy increasing popularity since they represent a cheap way for predicting the effects of development and change for a scene of interest. For the study presented in this paper, simulation would mean user- and situation-friendly representation of knowledge about geometry and material of building roofs. This knowledge can be derived from the sensor data and can help to estimate the locations which in the future may face heating-related challenges due to non-favourable inclination angles and materials of roofs, positions of trees, etc. Therefore, conducting this study can help the council to avoid the undesirable scenario of urban heat islands and help mitigate their current effects.

In short, this study aims to classify commonly used roofing materials in the council of the City of Melville using three types of aerial acquired datasets and to create visually appealing, semantic and three-dimensional model representation of the study area. Additionally, this paper studies different segmentation approaches to evaluate the effect those bring to the classification. The paper is structured as followed; we present our dataset in

*Corresponding author

Section 2, that follows by the proposed method in Section 3 consisting of pre-processing, 3D reconstruction, segmentation, texture analysis and finally classification. Lastly, the results are described in Section 4, while the final remarks and suggestions for future work are provided in Section 5.

2. DATASETS

To conduct the study, we received three datasets from the City of Melville that had been acquired during a 5-day data campaign in February 2016 and consisted of a multispectral and a thermal infrared imagery in addition to a LiDAR point cloud. The City of Melville also provided us with building outlines and labelled data. Additionally, a high-resolution RGB imagery was acquired. What follows is a detailed description of each utilised dataset.

Multispectral The multispectral imagery was acquired from an unmanned aerial vehicle system using the sensor *Micasense Rededge* (MicaSense, 2017). This sensor has a global shutter and a field of view of 47.2 °. Furthermore, it has five narrow bands; blue, green, red, red-edge and near-infrared within the spectral range of 400 - 900 nm. The achieved Ground Sample Distance (GSD) during the campaign was 0.41 m.

Thermal infared The thermal infrared imagery was acquired from an airborne system using the sensor *FLIR A615* (FLIR, 2018) thermal. The sensor has an uncooled microbolometer with a spectral range of 7.5 -13.5 μm. The thermal infrared imagery was acquired with a GSD of 0.6 m during the night because of the desire to reduce the impact from the thermal radiance caused by the sun.

LiDAR The LiDAR point cloud was acquired from an airborne system using the scanner *Riegl VZ - 1000* (Riegl, 2017). The point cloud was acquired with approximately 2 - 4 points / m^2 and with an accuracy of 0.1 m.

High-resolution RGB A high-resolution RGB aerial imagery dataset that was acquired May of 2016 with a GSD of 0.1 m was provided by the company Spookfish (Spookfish, 2018).

Building outlines A shapefile containing the outlines for the buildings in our study scene was acquired from the City of Melville, dated from 2012.

Labeled training data We received labelled training data from the City of Melville for 10 buildings of three commonly used roofing materials in Western Australia; *Cement tiles*, *Colorbond* and *Zincalume*. *Colorbond* and *Zincalume* are both metallic steel roofs consisting of a mixture of mainly aluminum and zinc, whereas *Colorbond* is prepainted steel and *Zincalume* is coated steel. All three roofing materials come in different colours, that is, they can all appear in different colour coatings.

3. PROPOSED METHOD

For the data stemming from different sources, as in our case, preprocessing is usually performed to correct systematic errors and reference all data into a common coordinate system. After a description of the most important pre-processing modules in subsection 3.1, we will refer in subsection 3.2 to the 3D reconstruction. In subsections 3.3 and 3.4, we will describe the procedures of building segmentation and texture analysis, respectively. These are the necessary ingredients to perform the material classification, which is described in subsection 3.5.

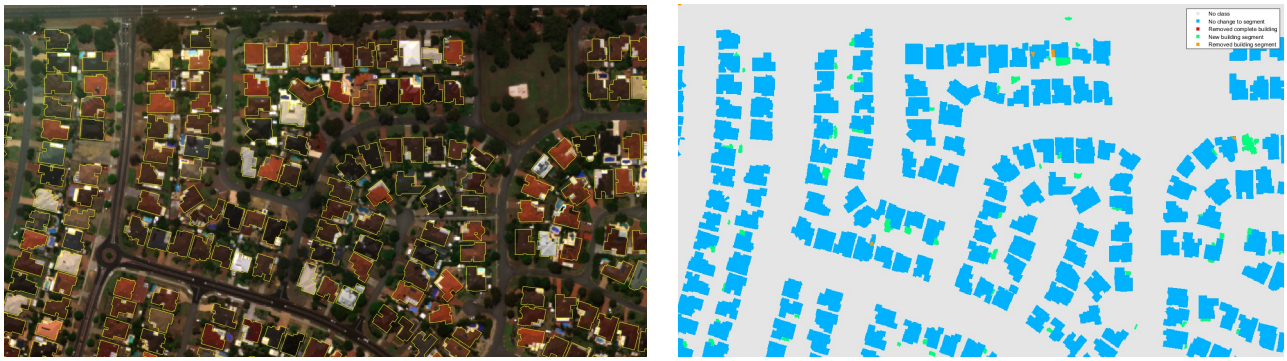
3.1 Data Pre-Processing

Our first step was the geometric and radiometric correction of the multispectral image. In order to correct the systematic errors caused by the non-nadir aspect of the orthophoto, the rectifying 2D homography best suiting the area of interest was retrieved interactively. For radiometric correction of the orthophoto, we determined the 0.01th and 0.99th quantiles of the intensities of red, green and blue values and then rescaled this image between these values. The resulting image contains almost the same information but has a much higher contrast and is therefore more suitable for texturation. To detect and analyse urban objects captured from the air, 2.5D representation of the terrain is mostly sufficient. We sampled the laser points to a Digital Surface Model (DSM) using the natural neighbour interpolation. Since the LiDAR point cloud has the point density of 2 - 4 points / m^2, the resolution of the DSM as well as of all the output results was set to 0.5 m. Starting at the DSM and using the method of Bulatov et al. (2014), we computed the Digital Terrain Model (DTM), and visualised it as a triangle mesh textured by the corrected multi-spectral image. The difference between the DSM and DTM, defined as Normalised Digital Surface Model (nDSM) (Weidner and Förstner, 1995) is now used for object detection and analysis. Two important measures, planarity (West et al., 2004) and the Normalised Difference Vegetation Index (NDVI) were additionally computed from nDSM and the multispectral image, respectively. NDVI is widely used in remote sensing, however, with respect to building detection, it occasionally fails. Therefore, planarity measure, originally proposed in West et al. (2004) and implemented by Gross and Thönnessen (2006) is additionally taken into account.

Even despite the geometric correction of the multi-spectral images, only systematic, translational errors were taken into account and not the non-nadir character of the image. Here, the deviations vary depending on the objects elevation. These deviations, mostly below 3 pixels, are not less significant, but, if an accurate analysis of elevated objects (buildings) is required, they may be quite disturbing, leaving aside the texturation aspect. Fortunately, the availability of the shapefile and the works accomplished by (Penney et al., 1998) offer a good strategy for building outlines correction. We search for a translation $\Delta u, \Delta v : -u_{\max} \leq \Delta u \leq u_{\max}, -v_{\max} \leq \Delta v \leq v_{\max}$, where $\cdot_{\max}$ denotes the maximum offset. An energy function is made up by a convex combination of two functions introduced by (Penney et al., 1998). The first is mutual information, encouraging the colour within the building outline to be constant. The second is the gradient correlation, encouraging high gradient values near outlines and also outside the building mask, since cars, house entries, and trees are usually situated next buildings.

The energy function is thus minimised using the gradient-free Nelder-Mead method implemented by Lagarias et al. (1998). This procedure allows computing the translational offsets $\Delta u, \Delta v$ for every building, whereby it is worth to note that it the method can be extended to determine unknown similarity or even affine transformation (with 4 or 6 degrees of freedom). We can use the rasterised outlines for assessing building-wise features based on elevation (that is, nDSM and planarity) while for those based on the multispectral image, the offsets are taken into account.

The shapefile containing the building outlines was from 2012 while the remaining datasets were dated from 2016. Thus, it was important to check whether buildings or building parts have been removed or newly added. To detect any completely removed buildings, we computed the building-wise median value of the



a b

Figure 1. Visualising the building outlines in our study area. a) the original building outlines visualised on the study area, b) the updated buildings outlines, whereas blue is representing unchanged building outlines, red completely removed buildings (none in this study area), green new building segments and orange removed building segment.

relative elevation. Buildings that were below the threshold were analysed. Mostly, the alarm was confirmed, however, roof made of dark (probably tarred) materials had extremely low point densities because of poor reflectance properties (Hebel, 2012). During DSM-sampling, these building masks were thus contaminated with ground points and exhibited low medians of relative elevations. Additionally, as prescribed by the regulations defined by the City of Melville, an area threshold of 10 m^2 was implemented to track the newly removed or added building parts. Removed segments were detected by thresholding the nDSM and by comparing the colour of the segment to its building’s mean colour. New building segments were detected by thresholding the nDSM, NDVI and the planarity measurement (using the implementation of Gross and Thönnessen (2006)) followed by connected component analysis. Once the building outlines were up-to-date, it was possible to extract the pixels within the building outlines and proceed to roof detail analysis and material classification. Figure 1 shows the updated building outlines for our study area. Additionally, to remove noisy and irregular shapes of the newly added or removed segments, we performed two morphological operators; *erosion* followed by *dilation* with a kernel of one pixel.

3.2 3D Scene Reconstruction

The crucial part of the scene reconstruction is the roof detail analysis of buildings. This procedure comprises two main substeps: dominant planes extraction and roof modelling. The first step presupposes that the building roof is piecewise planar. We used the J-Linkage algorithm (Toldo and Fusiello, 2008) to find planes in point clouds for every single building. A fixed number of planes is acquired from random triples of points, however preferring neighbours. The indicator vector, stating which point lies in which plane, is collected. By Jaccard distance over indicator vectors, points are clustered. We compared the results of J-Linkage algorithm with an alternative, also a very popular method, called Region Growing (Rabbani et al., 2006). The J-Linkage method performs better for our dataset. After the parameters for both approaches were fixed in order to produce possibly no undersegmentations, analysis of several test buildings showed that Region Growing produced by far more oversegmentations and, as the green roof of the building in Figure 2 shows, still some undersegmentations. On the negative side, J-Linkage produces sometimes ghost planes (dispersed inliers in different building parts) for complex buildings, and therefore some post-processing is necessary. This includes morphological operations, connected component analysis and filtering components by area and eccentricity. A more advanced approach, such as non-local optimisation proposed by Rothermel et al. (2014), was not performed because of the risk to lose important, but small connections of roof structures. Also, points on the margin of the building and not outliers of any dominant plane are mostly extremely important to comply with building outlines. They are added to the dominant planes using a 3D dilatation approach: the binary mask specifying assigned pixels is dilated. For every non-assigned point in the narrow band stemming from the dilatation, the closest (in 3D) assigned neighbour is computed and assigned a label. These steps (dilatation to obtain the narrow band, nearest neighbour computation and label propagation) are repeated until no non-assigned point remain.

What we achieved by the dominant planes extraction (visualised in Figure 2) is already enough to assess the inclination angles of roofs and with it, information about heat received when they are illuminated. However, the question *when* they are illuminated can only be assessed if occlusion analysis can be performed from an arbitrary viewpoint. This is the main reason why we strive for a polyhedral three-dimensional model representation. The second reason is that a realistic, easily recognisable and visually appealing model is conducive for promoting the awareness of non-expert residents for problems related to over-heating and energy-saving measures. Overall, the second step of creating of watertight building models comprises intersection of planes between each other and the building outlines. This is basically an extension of the algorithm of Xiong et al. (2014), however, experiments are currently being carried out to improve the handling of the step-lines.

After the roof detail analysis, walls are represented as vertical trapeziums by projecting the endpoints of border- or step-edges of the roofs to the ground. Intersection points with the ground are defined by the DTM values. Walls resulting from roof cut-edges are not required, since they are not visible in the model. To make the visual aspect of the output 3D model more appealing, building roofs were textured using the high-resolution images and the corrected multi-spectral image, following the method of Bulatov et al. (2014), without the geo-referencing step since everything was provided. Since the high-resolution images were only extracted from a portion of the scene and because of problems with rendering a scene with more than 1000 buildings and many trees, roofs of the buildings outside of the area were coloured in one of

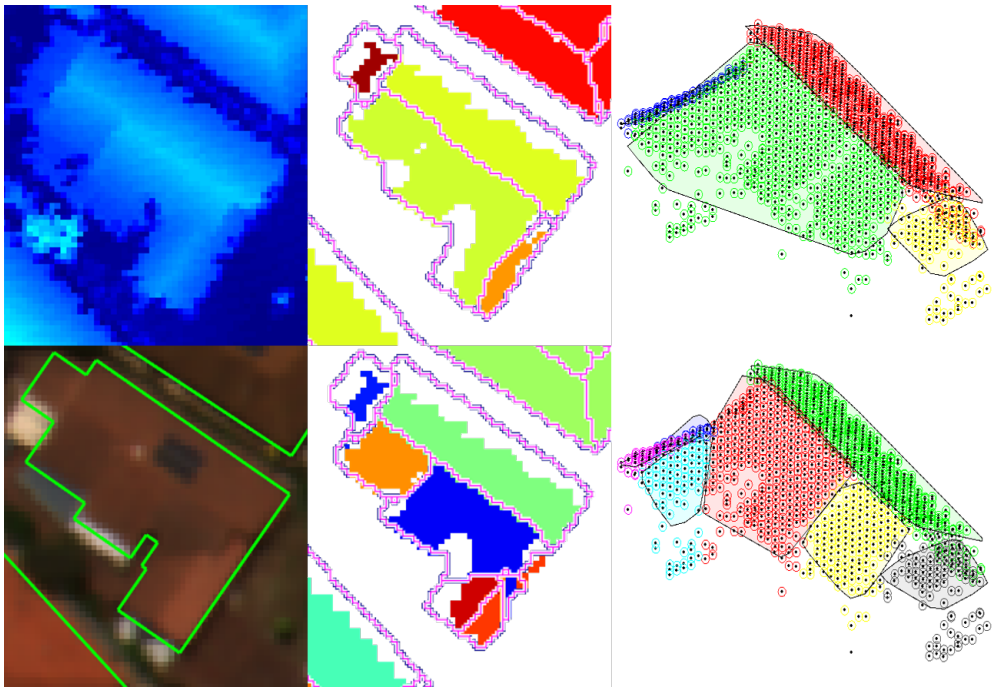


Figure 2. Comparison of performance of Region Growing and J-Linkage algorithms (middle and right in the top and bottom rows, respectively) for dominant plane extraction. On the left, the input data (fragment). In the middle, grouping of dominant plane extraction *before* post-processing and polygonisation *after* post-processing. On the right, 3D polygons after post-processing are specified by different colours and bounded by convex hulls.

two standard colours, and hence, building walls were treated. A screenshot showing the textured model is shown in Figure 3.

Lastly, we carried out the tree detection. The module for tree detection is based on the procedure of Bulatov et al. (2016). First, almost-circular components corresponding to the building size were detected in the binary image representing tree class. Next, watershed transformation is applied to the smoothed and inverted nDSM image. The watershed components are intersected with the tree class. Then, they are filtered with respect to their area and shape. Each of these components represents an individual tree. However, there are some larger components left, which could not be delineated and for which only an approximate number of trees and position of crown can be determined. Starting with the highest point of such a component, we delete from the component all pixels in a circular region of constant diameter around this point. The procedure is repeated for the remaining pixels of the component until it is empty. The height of a tree, its vertical position and diameter are extracted from the nDSM, the DTM, and the area of the component, respectively. A tree model is a standard object already employed in (Bulatov et al., 2016).

3.3 Segmentation

We considered three segmentation approaches for determining the most suitable for material classification: pixel-, superpixel- and building-wise. Pixel-wise approach is the most common for classification but may become intractable for larger scenes. Additionally, neighbouring pixels of the same roof might not be classified as the same material due to noise. On the other hand, a building-wise approach assumes that the entire roof consists of the same material. This is not always the case either, but a building-wise approach is useful for texture analysis since it considers the complete roof area. Additionally, the labelled data is already applied for a building-wise approach. Lastly, a superpixel-wise approach can divide a roof into several subparts based on the pattern differences. As a result, a roof can be divided into different material classes.

The superpixels were generated using the simple linear iterative clustering algorithm (Achanta et al., 2012) that groups a grayscaled RGB image into regions of similar pixel values. We assumed that one building contained a maximum of 10 superpixels, based on expert knowledge for the study area. Furthermore, we assigned every computed segment, that is, a building or a superpixel, a calculated mean value for that particular segment derived from the multispectral and the thermal infrared dataset.

3.4 Texture analysis

To add an additional band and additional information to perform the material classification, we computed texture features of the roofs using the high-resolution RGB dataset. Due to the desire to use different segmentation approaches, we had to perform two kinds of texture analyses approaches to suit the type of segmentation, i.e. the texture analysis of the pixel-wise segmentation had to be different to the superpixel- and the building-wise segmentations since the two aforementioned consisted of large segments and not only one pixel. For the pixel-wise segmentation, we determined the entropy, a statistical measure of randomness, and received a unique value for each pixel (Haralick et al., 1973). For the superpixel- and building-wise segmentations, the *Canny* edge detection algorithm (Canny, 1986) was utilised to determine the amount of pixels belonging to edges in each segment. For each segment, we calculated the percentage of determined edges as a signature of the roof texture.

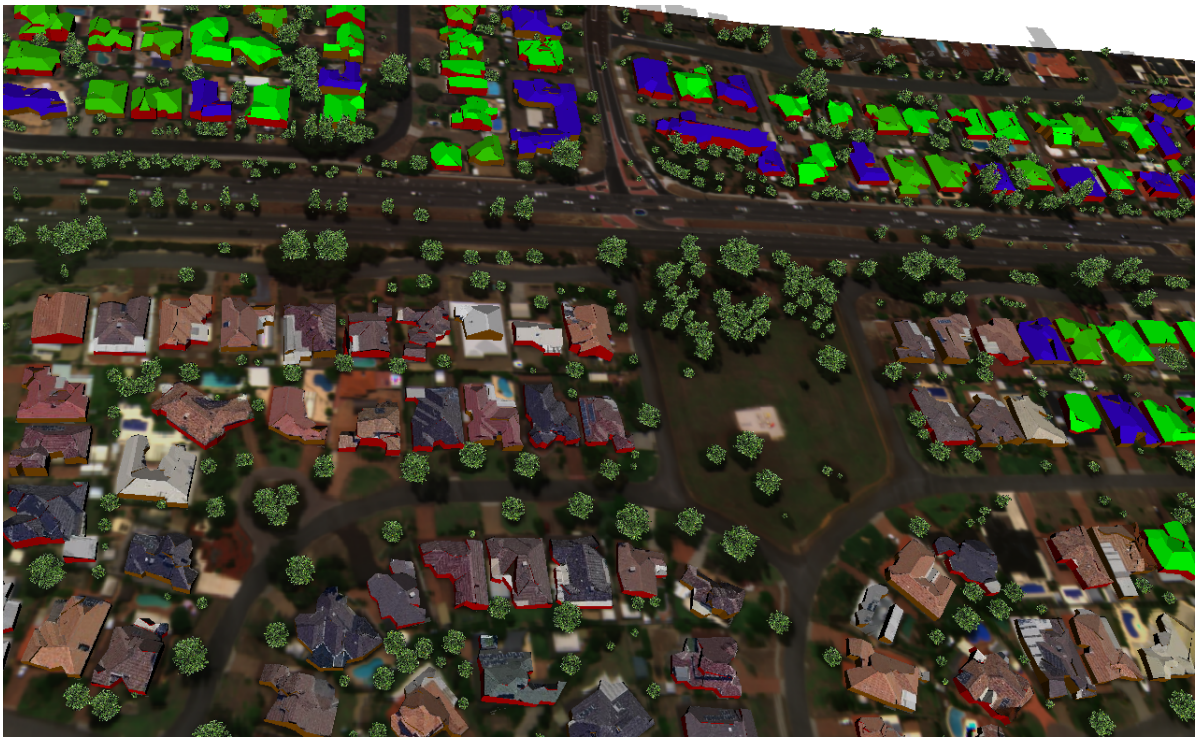
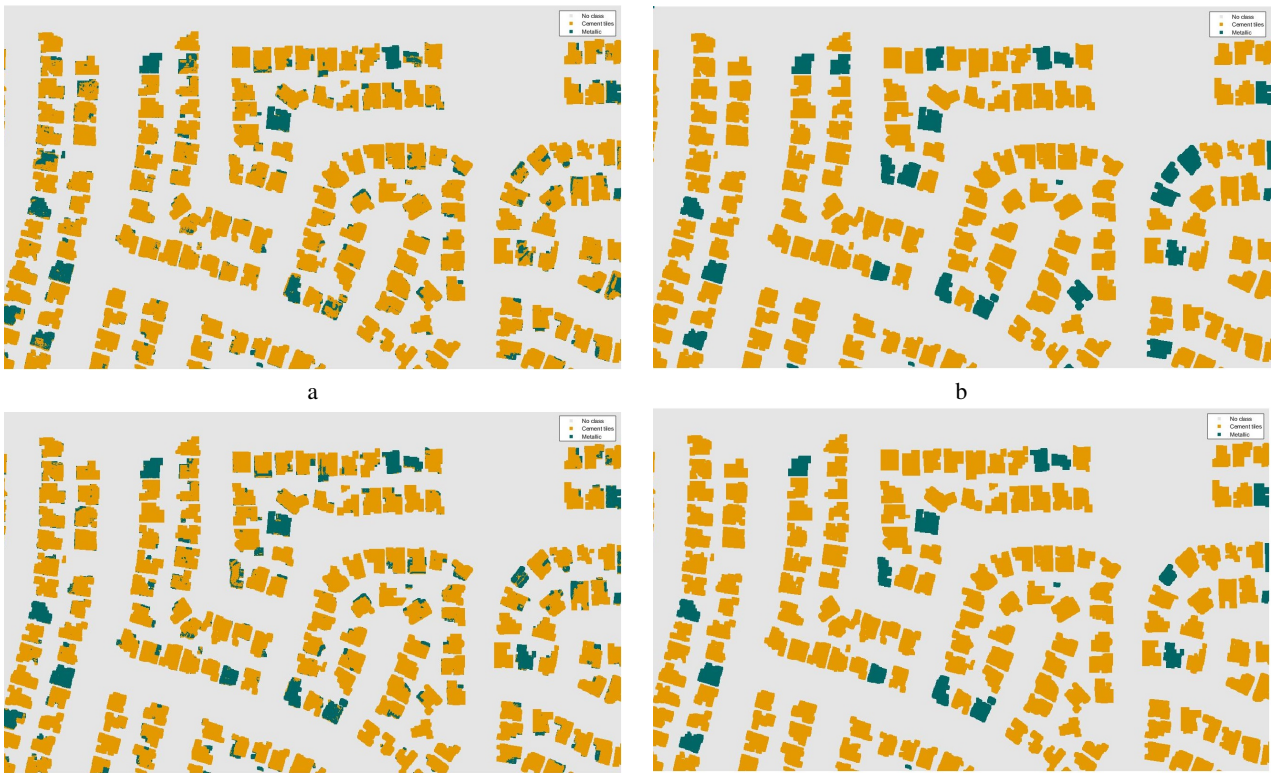


Figure 3. 3D model of the area of interest and surroundings.



a b c d

Figure 4. Visualised classification results using the original labelled dataset a) and b), and the extended labelled dataset c) and d). a) and c) display the classification results of pixel-wise segmentation, while b) and d) display the results of the building-wise segmentation.

3.5 Material classification

For the material classification, we implemented the supervised classifier Random Forest (RF) from the toolbox by Dollár (2016). To evaluate the performance of our three segmentation approaches, we generated an evaluation matrix consisting of the overall accu-

racy OA, which indicates the overall performance, the κ-value, which indicates how good the classes could can be separated from each other, and for each material class to detect any issues; the recall $\bar{R}$, the precision $\bar{P}$ and the F_1-score.

To determine correlation among our collected and computed bands, we performed a dimensionality reduction using a standard Principal Component Analysis (PCA), also implemented in the toolbox provided by Dollár (2016). By an orthogonal transformation, the data (features) is decorrelated and sorted in descending order with respect to its variability allowing to determine correlated bands. The first principal component covers the highest variability of the data while each following principal component covers the highest possible variability under the constraint generated by all previous principal components. PCA assumes that there is no significant information loss by discarding all other principal components by focusing on the first few principal components. For our study, we used the principal components which cover 99.9% of the variability of the given training data.

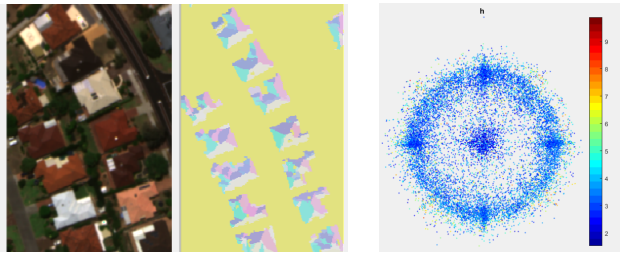
Finally, the originally labelled dataset from the City of Melville consisted of 10 buildings, whereas 7 were *Cement tiles*, 2 were *Colorbond* and 1 was *Zincalume*. Thus we merged the two classes *Colorbond* and *Zincalume* from the original dataset into one, denoted *Metallic*, and run additional experiments for two classes using all three types of segmentation approaches. Since we need training and testing data, it was not possible to perform building-wise classification for three classes. For all other configurations of number of classes and segmentation type, classification was carried out with and without PCA. Since we were classifying roofing materials using three different types of segmentations, we needed to adapt the number of training samples for each segmentation type. Logically, each segmentation approach utilised a different amount of training samples.

Due to the limited amount of labelled training data, we extend the dataset by labelling additional buildings on our own. We randomly chose 50 pixels in our study area and labelled the corresponding buildings. The labelled dataset was extended to 56 buildings, instead of the original 10 buildings.

By having three segmentation types and using both the original labelled dataset consisting of 10 buildings and the extended one comprising 56 buildings, we could determine differences between the various combinations. To have all bands in the same range, we normalised all data to have unity-based normalisation and unity standard deviation.

4. RESULTS AND DISCUSSION

Since the core of this work was to classify the most common roof materials, we will first briefly report the qualitative results on updating the building masks and derivation of roof orientations. In Figure 1, we can see that one residential building and at least three building-like smaller structures have appeared in the small fragment of the dataset. All these changes were confirmed by interactive intervention. By tracking removed buildings, we encountered holes in the nDSM that had occurred during the DSM sampling from the LiDAR point cloud and had caused false negatives. As for 3D modelling, Figures 3 and 5 suggest that the dominant orientations of the roof planes have been correctly estimated and, moreover, in most cases building models are indeed *watertight* despite extremely challenging roof structures. One can conclude from Figure 5 that there are many flat roofs in the dataset and



a b

Figure 5. Visualised statistics acquired about the inclination of the roofs. a) shows the colour-coded orientations (yellow: flat roof, pink: towards east, etc.) of dominant planes for a part of the dataset and b) shows the cosines of angles to x and y axes, coloured by the roof height (in the future, colour could refer to the material) once the dominant planes extraction has been performed.

that many roofs have approximately the same inclination angle, however, at different orientations (subfigure (b)).

Turning our attention to the classification results, presented in Table 1 and in Figure 4, a glance at the configuration with the original labelled data, 10 buildings and three classes (combinations a. 1-2 and b. 1-2) makes it clear that we obtained the best classification results at a pixel-wise level with an OA up to 96.7%. For instance, one building consists of enough pixel to generate a high accuracy. The various scores achieved using the aforementioned approach are in general high and indicate a good classification. However, due to the small amount of actual labelled data, the results should be taken in moderation. In contrast, the superpixel-wise segmentation approach appears to receive poorer results, in particular in the distinction between the three classes, as the κ-value ranges between 50.3 - 61.9%.

Moving on to studying the original labelled dataset with only two classes (combinations c. 1-2, d. 1-2, e. 1-2), the original dataset appears once again to generate good classification results with an OA ranging between 83.3 - 97.8%. However, the distinction between the two classes is again poor, as the κ-value for both the superpixel- and building-wise approaches ranges between 53.2 - 63.1%. The original labelled data consisted of 10 buildings and it is therefore inevitable that we did not have enough training data for separating the two roofing materials, *Cement tiles* and *Metallic*.

Finally, by studying the classification results using our extended labelled dataset (combinations c. 3-4, d. 3-4 and e. 3-4), it appears that we receive poorer classification results compared to only using the original 10 buildings. In comparison to the original dataset, we retrieved a poorer distinction between the classes as the κ-value indicates with, in general, lower values (ranging between 33.2 - 89.7%). This could be caused by incorrect labelling and by a larger variation of materials. For instance, an increased number of varied colours and different conditions (such as the age and the quality of building maintenance) contributed to misclassifications. However, these results are more realistic, since a natural variance does exist in terms of colouring and conditions.

Overall, the application of PCA does not seem to have improved the classification results (combinations using the datasets 2 or 4). Since we were only using 7 bands for the actual classification, the correlation between the bands might not have been large enough

		3 classes		2 classes		
		a.pixel	b.superpixel	c.pixel	d.superpixel	e.building
OA [%]	1.original	96.7	80.0	97.8	92.7	83.3
	2.original$_{\mathrm{PCA}}$	96.6	83.1	97.6	92.7	-
	3.extended	-	-	94.3	85.2	97.8
	4.extended$_{\mathrm{PCA}}$	-	-	94.0	87.5	86.7
κ [%]	1.original	91.4	50.3	94.3	53.2	57.1
	2.original$_{\mathrm{PCA}}$	91.3	61.9	93.7	63.1	-
	3.extended	-	-	79.6	33.2	89.7
	4.extended$_{\mathrm{PCA}}$	-	-	78.9	39.6	42.5
R [%]	1.original	96.6/96.8/99.8	82.4/66.7/100	97.7/98.2	94.7/66.7	88.9/66.7
	2.original$_{\mathrm{PCA}}$	96.1/98.2/99.8	84.3/75.0/100	97.5/97.9	92.1/100	-
	3.extended	-	-	94.0/95.7	86.6/68.0	97.5/100
	4.extended$_{\mathrm{PCA}}$	-	-	93.6/96.2	88.8/72.0	90.0/60.0
P [%]	1.original	99.3/89.1/90.1	91.3/57.1/40.0	99.4/93.4	97.3/50.0	100/50.0
	2.original$_{\mathrm{PCA}}$	99.6/88.3/87.3	100/81.8/18.2	99.3/92.8	100/50.0	-
	3.extended	-	-	99.2/73.3	97.2/28.3	100/83.3
	4.extended$_{\mathrm{PCA}}$	-	-	99.3/72.0	97.6/33.3	94.7/42.9
F_1 [%]	1.original	97.9/92.8/94.7	86.6/61.5/57.1	98.6/95.7	96.0/57.1	88.9/66.7
	2.original$_{\mathrm{PCA}}$	97.8/93.0/93.1	91.5/78.3/30.8	98.4/95.3	95.9/66.7	-
	3.extended	-	-	96.6/83.0	91.6/40.0	98.7/90.9
	4.extended$_{\mathrm{PCA}}$	-	-	96.4/82.4	93.0/45.6	92.3/50.0

Table 1. Classification results achieved for the 8 different combinations (segmentation with dataset). Displaying the overall accuracy OA and the κ-value in addition to the recall $\bar{R}$, the precision $\bar{P}$ and the F_1-score for each class.

to reduce the number of used bands. Indeed, when we are evaluating the number of principal components that were used for the PCA, the components that cover 99.9% of the variability, the number varies between 4 and 6 components. That implies that the bands do not correlate in most cases. While determining the importance each band had during the classification, it appears that the thermal infrared band and that proceeding from texture analysis are often the least important bands. Hence, the classification could have been performed by the multispectral imagery only and the results would not suffer tremendously.

Furthermore, since the labelled data was done on a building-wise level, that is, the labelling assumed that one building could only consist of one material, the pixel- and the superpixel-wise segmentation could also consist of unlabelled materials. Several buildings had either solar panels or solar water heaters on the roof and since none of those two classes had been labelled beforehand, these elements were classified into one of the existing labelled materials. This could explain the poorer classification results for the superpixel-wise segmentation while studying the different score retrieved from the evaluation matrix.

5. CONCLUSIONS AND OUTLOOK

In this work, we have classified commonly used roofing materials in the City of Melville using three imagery datasets, a multispectral, a thermal infrared and high-resolution RGB imagery. Additionally, we have used a LiDAR point cloud in combination with building outlines to localise the roofs and created 3D models of the buildings and the surrounding terrain. We utilised three labelled datasets; the original provided by the City of Melville that consisted of the three roofing materials *Cement tiles*, *Colorbond* and *Zincalume*, the original but consisting of the two roofing materials *Cement tiles* and *Metallic*, and finally an extended dataset consisting the two aforementioned materials. We used three different segmentation approaches; pixel-, superpixel- and building-wise segmentation. We classified 8 different combinations using the commonly used classifier RF, in combination with and without PCA.

By studying the evaluation matrices for the different combinations, it appears that we retrieve the best OA results using the pixel-wise segmentation approach. The κ-values indicate that we retrieve the best material classification using the original labelled dataset, receiving similar κ-values for using either two or three classes. The superpixel-wise approach generates the worst results, indicated particular by the κ-values. Furthermore, the building-wise approach appears to receive good results as well, but with a bit poorer κ-values. PCA does not seem to have improved the classification results implying that the bands do not correlate.

As future work, we would like to study the correlation between urban heat islands and the amount of vegetation and the roofing materials in the City of Melville. As stated, knowledge about commonly used materials in urban areas is valuable information for local councils, such as the City of Melville. By considering the classification results, it can help the council with upcoming planning in the area. Therefore, we would need to extend our labelled datasets to expand and increase our study areas.

Finally, we emphasise that the two accomplished steps, material classification and derivation of roof orientation are merely preparation steps for the localisation of the urban heat island effect. What should come at the end of a long-term research is a combination of both ingredients into a thermo-dynamical simulation. This would not only be a visually appealing representation of the scene, but also a more accurate assessing of surface temperature when it comes to tracing sunrays.

6. ACKNOWLEDGMENT

We would foremost like to thank the City of Melville, in particular Janine Ahola, for providing us with the necessary data to perform this study. We would also like to thank Spookfish for providing us with the high-resolution RGB imagery. Lastly, we would also like to thank Jochen Meidow and colleagues from Fraunhofer IOSB, (Ettlingen, Germany) for helping us with modelling building roofs.

The corresponding author would like to thank the graduate school GRACE for financially supporting the research stay at Department of Spatial Sciences, Curtin University.

References

Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P. and Süsstrunk, S., 2012. SLIC superpixels compared to state-of-the-art superpixel methods. *IEEE transactions on pattern analysis and machine intelligence* 34(11), pp. 2274–2282.

Bulatov, D., Häufel, G., Meidow, J., Pohl, M., Solbrig, P. and Wernerus, P., 2014. Context-based automatic reconstruction and texturing of 3D urban terrain for quick-response tasks. *ISPRS Journal of Photogrammetry and Remote Sensing* 93, pp. 157–170.

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

Canny, J., 1986. A computational approach to edge detection. *IEEE Transactions on pattern analysis and machine intelligence* (6), pp. 679–698.

Chen, X.-L., Zhao, H.-M., Li, P.-X. and Yin, Z.-Y., 2006. Remote sensing image-based analysis of the relationship between urban heat island and land use/cover changes. *Remote sensing of environment* 104(2), pp. 133–146.

Council of City of Meville, 2017. Urban forest strategic plan 2017-2036: Plan A: City-controlled plan. Technical report, City of Melville, Perth, Australia.

Dimmeler, A., Schilling, H., Shimoni, M., Bulatov, D. and Middelmann, W., 2013. Combined airborne sensors in urban environment. In: *Electro-Optical Remote Sensing, Photonic Technologies, and Applications VII; and Military Applications in Hyperspectral Imaging and High Spatial Resolution Sensing*, Vol. 8897, p. 88970U.

Dollár, P., 2016. Piotr's Computer Vision Matlab Toolbox (PMT). `https://github.com/pdollar/toolbox`. Last accessed 29.06.2018.

FLIR, 2018. FLIR A615 specifications. `https://www.flir.com/products/a615/`. Last accessed 06.07.2018.

Gross, H. and Thönnessen, U., 2006. Extraction of lines from laser point clouds. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* 36 (Part 3/W49), pp. 86–91.

Haralick, R. M., Shanmugam, K., Dinstein, I. et al., 1973. Textural features for image classification. *IEEE Transactions on systems, man, and cybernetics* 3(6), pp. 610–621.

Hebel, M., 2012. Änderungsdetektion in urbanen Gebieten durch objektbasierte Analyse und schritthaltenden Vergleich von Multi-Aspekt ALS-Daten. PhD thesis, Technical University of Munich.

Kumar, U., Milesi, C., Nemani, R. R. and Basu, S., 2015. Multi-sensor multi-resolution image fusion for improved vegetation and urban area classification. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* 40(7), pp. 51.

Lagarias, J. C., Reeds, J. A., Wright, M. H. and Wright, P. E., 1998. Convergence properties of the Nelder–Mead simplex method in low dimensions. *SIAM Journal on Optimization* 9(1), pp. 112–147.

MicaSense, 2017. RedEdge Specifications. `https://support.micasense.com/hc/en-us/articles/225950667-RedEdge-Manual-Specifications`. Last accessed 06.07.2018.

Oke, T. R., 1982. The energetic basis of the urban heat island. *Quarterly Journal of the Royal Meteorological Society* 108(455), pp. 1–24.

Penney, G. P., Weese, J., Little, J. A., Desmedt, P., Hill, D. L. et al., 1998. A comparison of similarity measures for use in 2-D-3-D medical image registration. *IEEE transactions on medical imaging* 17(4), pp. 586–595.

Rabbani, T., Van Den Heuvel, F. and Vosselmann, G., 2006. Segmentation of point clouds using smoothness constraint. *International archives of photogrammetry, remote sensing and spatial information sciences* 36(5), pp. 248–253.

Riegl, 2017. Riegl VZ-1000. `https://www.3dlasermapping.com/wp-content/uploads/2017/10/DataSheet_VZ-1000_2017-06-14.pdf`. Last accessed 06.07.2018.

Rothermel, M., Haala, N., Wenzel, K. and Bulatov, D., 2014. Fast and robust generation of semantic urban terrain models from UAV video streams. In: *Proc. International Conference on Pattern Recognition*, pp. 592–597.

Spookfish, 2018. Spookfish Product SiteView. `https://www.spookfish.com/`. Last accessed 06.07.2018.

Toldo, R. and Fusiello, A., 2008. Robust multiple structures estimation with J-Linkage. In: *Proc. of European Conference on Computer Vision, Marseille, France*, Vol. 1, pp. 537–547.

Weidner, U. and Förstner, W., 1995. Towards automatic building extraction from high-resolution digital elevation models. *ISPRS Journal of Photogrammetry and Remote Sensing* 50(4), pp. 38–49.

Weng, Q., Lu, D. and Schubring, J., 2004. Estimation of land surface temperature–vegetation abundance relationship for urban heat island studies. *Remote sensing of Environment* 89(4), pp. 467–483.

West, K. F., Webb, B. N., Lersch, J. R., Pothier, S., Triscari, J. M. and Iverson, A. E., 2004. Context-driven automated target detection in 3d data. In: *Automatic Target Recognition XIV*, Vol. 5426, International Society for Optics and Photonics, pp. 133–144.

Xiong, B., Elberink, S. O. and Vosselman, G., 2014. Building modeling from noisy photogrammetric point clouds. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 2(3), pp. 197.