Occlusion-free Orthophoto Generation for Building Roofs Using UAV Photogrammetric Reconstruction and Digital Twin Data

Jiajun Li[1], Frédéric Bosché[1], Chris Xiaoxuan Lu[2] and Lyn Wilson[3]

[1]School of Engineering, University of Edinburgh, UK
[2]School of Informatics, University of Edinburgh, UK
[3]Historic Environment Scotland, UK

J.Li-301@sms.ed.ac.uk, f.bosche@ed.ac.uk, xiaoxuan.lu@ed.ac.uk, lyn.wilson@hes.scot

Abstract -

Building roofs are a primary barrier of the building fabric to protect building interiors and their occupants from the effects of the weather. Effectively monitoring their condition and maintaining them is thus critical, and even more so in light of the increasing risks induced by climate change. Ideally, this requires frequent and careful surveying, as well as recording of the results in a way that facilitates monitoring over time. Traditional manual data collection and defect detection and recording methodologies have been used mainly on building facades. Orthophotos are commonly used for recording. However, (1) current approaches to generate orthophotos cannot ensure the photos are occlusion-free; and (2) the results cannot be efficiently recorded in a structured, digital way with semantic relationship to the corresponding elements in the building Digital Twin (DT). In this paper, a method is proposed to automatically generate high-resolution orthophotos of roof panels from UAV data, given an existing building DT. The resulting orthophotos have high-resolution (e.g. 5mm per pixel). They are occlusion-free and linked semantically to the building elements in the DT. The method is demonstrated using a real case study of a complex pitched slated roof of a traditional building.

Keywords -

Roof; UAV; Orthophoto; Photogrammetry; Occlusion; Projection.

1 Introduction

As a key building element, a roof is one of the most exposed elements to the environment. Elevated temperature, moisture, material soiling can result in staining, weathering and even damage to different materials [1]. Over time, this gradually leads to roof fragility, component failure, and load redistribution, which can create further risks to the building and its occupants [2]. Extreme climate events, such as frequent high summer temperatures, heavier winter precipitation, unpredictable wind speed, intense storms, etc, will aggravate and accelerate deterioration to all kinds of buildings. For traditional buildings, preventing failure and conservation measures have become prime considerations [3]. Therefore, building roofs should be routinely monitored, so that defects can be detected and remedied promptly.

The first difficulty of maintaining building roofs is capturing roof data, which is commonly undertaken manually. However, with the development of digital photography, Unmanned Aerial Vehicles (UAV) is increasingly used, as it reduces the cost of access provision and the risks to surveyors [4]. Meanwhile, digital photogrammetry has been an active topic for generating 3D models of buildings due to its high level of automation and the fact that it can generate 3D data at low cost [5]. The combination of these two approaches, *UAV-based photogrammetry*, has been applied to several contexts, such as archaeology, disaster monitoring, construction monitoring, building surveying, or land mapping [6].

In the Facility Management (FM) industry, Digital Twins of building assets — derived from Building Information Modelling (BIM) models — can help facility managers make and record decisions related to operation and maintenance [7]. To ease the generation of semantically-rich 3D models which are commonly the basis of Digital Twins, Scan-to-BIM methodologies can be applied to survey data acquired by laser scanning or photogrammetry. For example, Rocha et al. [8] employ a manual modelling workflow based on photogrammetry and laser scanning data to develop a BIM model of a historic house in Lisbon for conservation and restoration. In [9], Adán et al. present an integrated system to reconstruct large scenes at five semantic levels, addressing issues of automated data collection, to architectural building element modelling, to small accessory component recognition. And, Valero et al. [10] propose a Terrestrial Laser Scanning (TLS) data-processing pipeline for producing semantic 3D models of furnished office and home interiors.

Orthophoto imagery has been used for many decades in FM, most commonly in the context of (traditional) building facade surveying. The advantage is that they provide a uniform scale across the entire image, which makes them measurable (with dimensions comparable across the or-

thophoto) and more suitable for detecting, reporting and assessing building defects [11]. The generation of orthophotos is achieved by an orthographic projection on vertical planes [12, 13]. Recently, high-resolution orthophotos of facades have been used as input data for automated detection of facade defects [14, 15, 16]. These examples show that orthophotos are convenient for monitoring buildings, and accessible for actual surface measurement and (automated) analysis.

However, compared to building facades, roofs are more easily occluded by other building elements, such as chimney, dormers, aerials, satellite dishes, etc [17]. As a result, building facade orthophotos may not enable a full view of roofs. Figure 1 shows an example of orthophoto commonly generated for a whole building. One can see that, while the stone facade(s) appear(s) fully, the middle roof panel is occluded by the balustrades and chimneys. These problems challenge the generation of clear orthophotos (without occlusion and with limited blur), and ultimately the development of automatic defect detection models, as the models need to be agnostic to different image acquisition situations (e.g. angle, distance).

Figure 1. Orthophoto for the entire front elevation of Duff House.

Therefore, in this study, a method is proposed for the generation of orthophotos of pitched building roofs based on Scan-vs-BIM context afforded by an existing building DT. The approach generates vertical orthophotos that are free of occlusions from other building elements and with reduced blur.

2 Methodology

The ortho-projection method realises transformation from a dense set of images taken by UAV and the DT of the building to a set of orthophotos corresponding to all roof panels defined in the DT.

This method contains 3 steps: 1. As-is 3D reconstruction; 2. Roof panel mask generation; 3. Orthophoto generation.

2.1 Use Case

The methodology is simultaneously explained and validated using data from Duff House, a traditional building located in Banff, Aberdeenshire, Scotland. Duff House is a category A listed, early Georgian mansion under the care of Historic Environment Scotland (HES)[1]. Note that this methodology is in fact generic, and not specific to Duff House.

2.2 Step 1: Reconstructing building's as-built 3D model

Duff House was fully surveyed using a UAV equipped with a Sony ILCE-7R digital camera, with a fixed focal length of 35 mm. The UAV survey acquired 859 images containing the roof, and with over 75% overlapping, enabled a detailed and accurate 3D reconstruction of the roof (and upper parts of the facades). A full 3D reconstruction of the house exterior (facades, entrance stair case, etc.) was also obtained by merging the UAV reconstruction with a Terrestrial Laser Scanning (TLS) point cloud. The complete 3D model of Duff House is shown in Figure 2, while the roof reconstruction from the UAV data only is shown in Figure 3. The UAV data reconstruction contains 2,104,179 points registered in the world coordinate system. The roof is 34.2m in length and 25.7m in width. Therefore, the average point cloud density is approximately 2,400 points/m^2.

Figure 2. Duff House complete 3D reconstruction.

[1]https://www.historicenvironment.scot/archives-and-research/publications/publication/?publicationId=d752300f-2266-4574-8334-a57000ca8ed8/

The complete roof contains 36 slated panels, 20 of which are planar and 16 are curved. Aside from the slated roof panels, there is also significant leadwork surrounding the panels, as well as chimneys and decorative components nearby, like stone ornaments and balustrades, which can cause shadow and occlusions when taking images with UAV. Panel '4' in Figure 3 will be used to illustrate the processing for occlusion handling, while all four representative panels highlighted in the figure will be shown in the section reporting the final results.

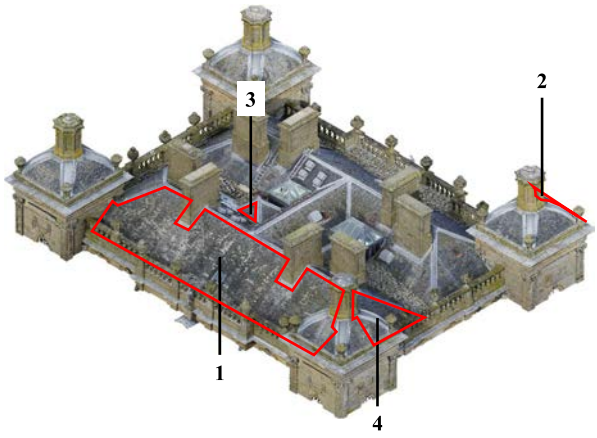


Figure 3. Reconstructed 3D point cloud of Duff House's roof. Four representative roof panels are highlighted that are used to illustrate and validate the proposed method.

Based on the input DT, a more granular model was built with an Industry Foundation Classes (IFC) object for each roof panel. This model simply needs to contains the geometry and location of each roof. As shown in Figure 4, all the panel models were built for representing only the slate part, without components such as leadwork and windows — for a reason that falls outside the scope of this paper. The geometry of each panel is defined as panel boundary with extrusion. As input to the proposed process, geometries of each roof were then defined as triangle meshes.

2.3 Step 2: Generating roof panel masks

The purpose of this step is to define, for each input UAV image, the parts of the image corresponding to a given roof panel. The boundary of each roof panel model is defined in world coordinate system. According to the input image camera external calibration in this world coordinate, this boundary can be transferred into the camera's 3D coordinate system, using equation 1.

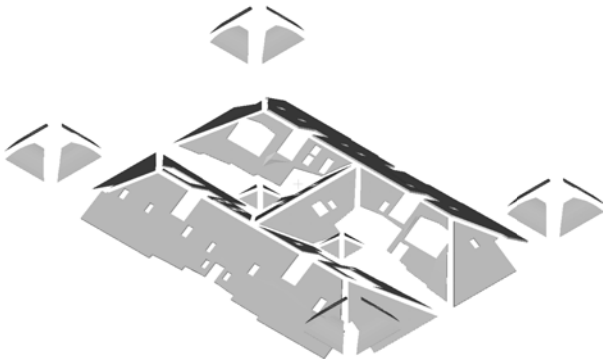


Figure 4. Roof panel models.

$$\mathbf{P_c} = \begin{bmatrix} X_c \\ Y_c \\ Z_c \end{bmatrix} = \begin{bmatrix} \mathbf{R} & \mathrm{t} \end{bmatrix} \begin{bmatrix} \mathbf{P}_w \\ 1 \end{bmatrix} = \begin{bmatrix} \mathbf{R} & \mathrm{t} \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} \quad (1)$$

where $\mathbf{P_c}$ and $\mathbf{P_w}$ are the point coordinates in camera system and world system respectively. $\mathbf{R}$ represents the 3×3 rotation matrix in the world system, defined by the three camera rotation angles around x, y, and z axes. t represents the 3×1 camera translation vector.

Some camera system would contain a polynomial distortion transformation, which is optional and varies with different camera settings. Such processing should be applied to normalize $\mathbf{P}_c$ by dividing its first two components, X_c and Y_c, by its last one Z_c.

Next, each 3D point $\mathbf{P_c}$ can be projected to the 2D image coordinate system, where the origin is the image centre, using equation 2. The output results u and v, in range [-0.5, 0.5], are relative coordinates in the image system (centre as the origin).

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \frac{1}{\mathbf{Z_c}} \mathbf{K} \mathbf{P_c} \quad (2)$$

where $\mathbf{K}$ is the 3×3 camera's internal calibration matrix, containing focal length, camera skew, aspect ratio, and principle point offset. Finally, the above output results, u and v, should be scaled to the image pixels and transformed to where the top left corner is the origin of image coordinate system, as defined in Equation 3.

$$\begin{cases} \text{scale} = \max(\text{w}, \text{h}) \\ \text{x} = \text{scale} * u + w/2 \\ \text{y} = \text{scale} * v + h/2 \end{cases} \quad (3)$$

where w and h are the width and height pixels of the image. x and y compose the pixel coordinate in the image.

With this process, the boundary polygon of the target roof panel can be projected onto the image. The mask image for the input UAV image and for the given roof panel is then generated by setting the RGB value of all the pixels outside the polygon as [0, 0, 0], and those inside the polygon as [1, 1, 1].

By applying this process to all UAV pictures, the subset of pictures that captures at least part of a given roof panel is identified and, for each of those pictures, the set of pixels corresponding to that panel are also identified.

2.4 Step 3: Generating orthophotos

After generating all the mask images containing the given target roof panel, the masked UAV images can be converted into an orthophoto. For this, all the pixels in the masked area are projected on the model of the panel boundary in the world coordinate system, and these projections are then projected on the pre-defined orthophoto.

The plane of projection of each orthophoto is defined by computing the principal axes of the roof panel using Principal Component Analysis (PCA) [18]. A rotation matrix $\mathbf{R_{PCA}}$ derived in this process defines the projection plane of the orthophoto in the world coordinate system of the reconstructed DT mesh data. As illustrated in Figure 5, the plane YOZ is the orthophoto projection plane for the roof panel '4'.

The orthophoto generation process then includes 3 steps: 1. (2D-3D) All the coloured 2D pixels in original masked image are firstly projected to the DT mesh data, with corresponding 3D points as output; 2. (3D-2D) All the 3D points derived are then projected onto the orthophoto plane. 3. The orthophoto boundary in the projection plane is then set based on the projection of the panel geometry and the resolution of the orthophoto is set to the required value — we use $5mm \times 5mm$ per pixel.

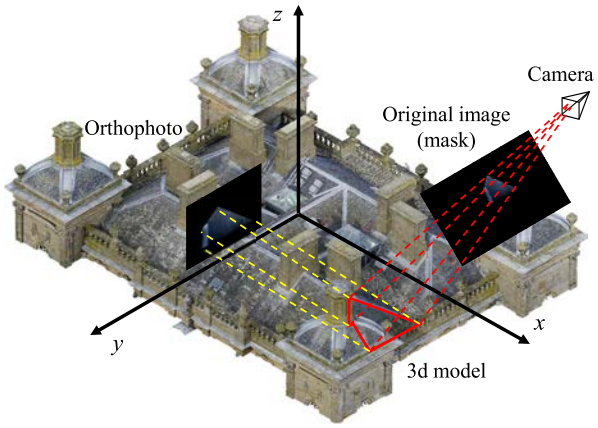


Figure 5. Orthophoto generation.

The outcome of the above process is an orthophoto generated from each of the input UAV picture and for each given roof panel. These orthophotos can be used as-is for analysis. However, because of the fact that the original UAV images were taken from different angles and locations, pixels from occluders still exist and resolution artefacts can appear. The orthphoto in Figure 6a is a typical example: some pixels remain black because no original pixel reprojects on them, because the input image was too far and oblique. Also, the orthophoto contains occlusions by a nearby chimney.

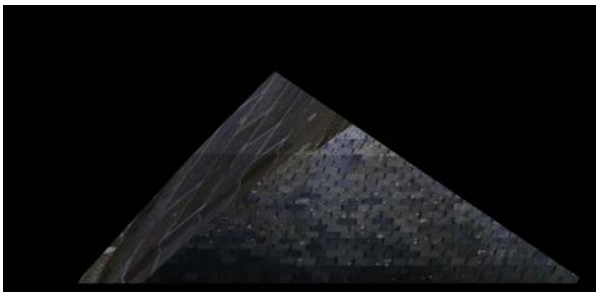
We address these issues by merging all orthophotos for one roof panel into one. The images can be merged into one, by calculating the median value of the RGB for each orthophoto pixel. If, as we have observed, occlusions appear in fewer than 50% of the images, merging the orthophotos could help eliminate occlusions. Figure 6b shows the result of merging the 45 orthophotos obtained from the 45 UAV images in which roof panel '4' appears.

As can be seen, the resulting orthophoto is free of pixels from occluders and also does not contain any hole (i.e. empty pixel) in the roof panel area. However, careful observation shows that these merged images often contain significant amount of blur, which is attributed to small errors from the photogrammetric process.

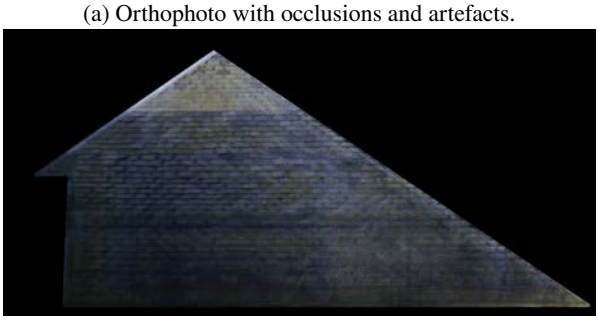
To reduce the influence of these errors, the small shifts between images are detected and a correction applied. For this, Scale-Invariant Feature transform (SIFT) features are detected and matched across images [19]. Figure 7 shows the SIFT feature matching between two orthophotos. The green lines show SIFT matches between them two. Based on this, the homography can be computed and applied to the right orthophoto before merging it with the left one. This process is applied by selecting one orthophoto as the baseline, and shifting all the other ones to align them to it. The baseline orthophoto is ideally the one with best visual quality, and is selected by choosing the smallest values of the filtering parameters introduced below.

Afterwards, the median value for each pixel can be calculated taking into account all shifted orthophotos, to generate a merged orthoimage. As can be seen in Figure 6c, the resulting median orthophoto has less blur than that of the initial median orthophoto.

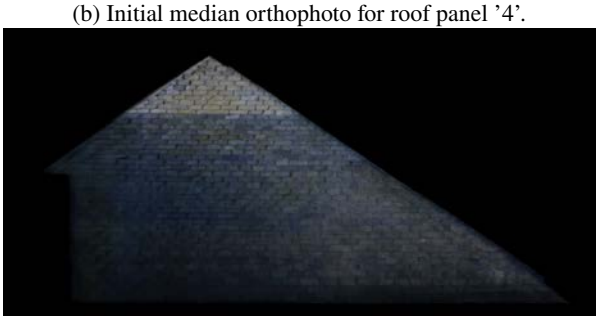
However, due to the fact that images were taken from various distances and with various incidence angles, not all the images contribute well to the merged orthophoto. Therefore, image filtering should first be carried out based on the angle and distances in order to remove those images that are just taken from too far or with angles that are too oblique. As illustrated in Figure 8, we define the ray shot from the pinhole camera O_i through the image centre C_i, and the intersection point with the roof plane is named as C_i'. Let's call C_p the centre of the target panel, and $\vec{n_p}$ its normal. Therefore, the first filtering parameter in



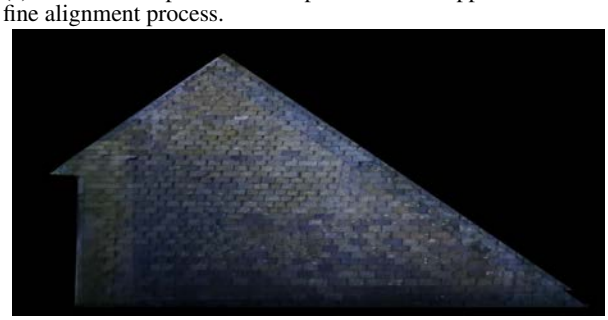
(a) Orthophoto with occlusions and artefacts.



(b) Initial median orthophoto for roof panel '4'.



(c) Median orthophoto for roof panel '4' after application of the fine alignment process.



(d) Median orthophoto for roof panel '4' after application of the fine alignment and filtering processes.

Figure 6. Merging result comparison between different datasets.

this study is the angle α_{pi} between $\overrightarrow{C_i'O_i}$ and $\vec{n_p}$, which captures the picture incidence angle between respect to the roof panel. The second filtering parameter d^1_{pi} is the distance $\|\overrightarrow{O_iC_i'}\|$, which broadly captures the distance from the camera to the roof panel along the view direction. Finally, the third filtering parameters d^2_{pi} is the distance

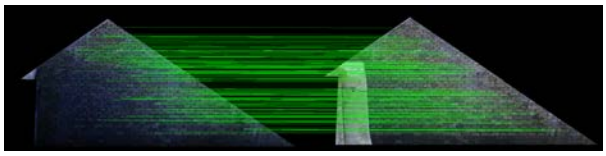


Figure 7. Pixel shifting.

$\|\overrightarrow{C_pC_i'}\|$ normalised by the maximum dimension of this panel d_p^{max}, which can be used to measure whether the camera is pointing at the target panel or away from it. The smaller α_{pi}, d^1_{pi} and d^2_{pi} are, the better the capture of the roof panel in the image should be.

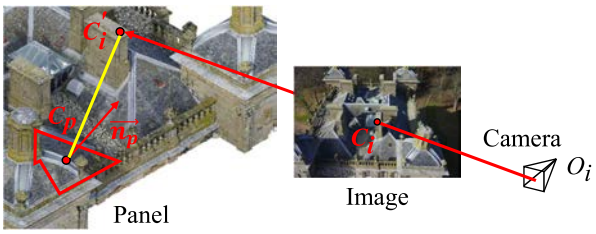


Figure 8. Explanation of different parameters for screening.

In the Duff House example, as mentioned earlier, initially 45 images contain the target roof panel '4'. Their α_{pi}, d^1_{pi} and d^2_{pi} values are presented in the 3D scatter plot in Figure 9. For the target panel here, the thresholds for α_{pi}, d^1_{pi} and d^2_{pi} are: 45°, 34m and 2. The angle 45° is defined somewhat arbitrarily but is set because beyond that images would be too oblique. The distance 34m is set as twice the value of the average UAV flying height from the roof. The relative distance 2, means twice the value of the largest dimension of the roof. In this case, 12 images simultaneously meet all these three requirements. Figure 6d shows the median orthophoto obtained using only the filtered set of images.

The final orthophoto in Figure 6d presents significantly less blur than the median orthophoto in Figure 6c obtained without image filtering. Furthermore, the proposed image dataset filtering appears to help balance the lighting condition in the whole image, with Figure 6b and Figure 6c both showing a brighter section at the top of the panel, while Figure 6d does not show it.

3 Further Results

The application of the proposed method to four representative panels in Figure 3 are reviewed here. Panel '1' is the largest panel within the whole roof, placed on the front and back side of the house (easily occluded by balustrades, chimneys and towers in some images, it doesn't appear

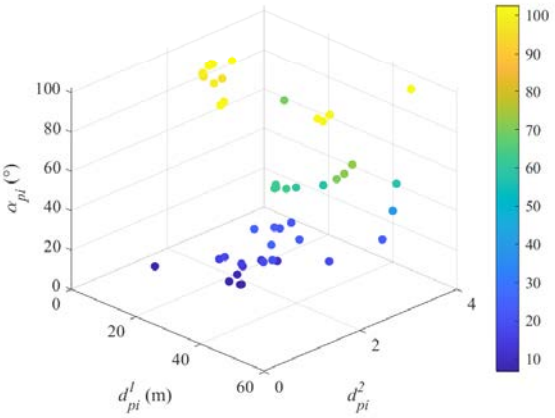


Figure 9. α_{pi}, d^1_{pi} and d^2_{pi} in different images.

completely in any single image). Panel '2' is a curved panel placed on the one of the towers of the house (easily occluded by stone ornaments). Panel '3' is one of the triangular panels placed near the middle of the roof with almost no occlusion. Finally, Panel '4' is one of the panels placed on the side of the house (easily occluded by the towers and balustrades on the edges).

The results obtained for panel '1' to '3' are shown in Figure 10. The results are evaluated qualitatively and quantitatively, considering: 1. whether the occlusions from other components have been removed; 2. the general level of blur the orthophoto presents (which can be evaluated by assessing the sharpness of the boundary of each slate); and 3. the mean percentage of matched features $\bar{a}$ described below.

To quantify the quality of the final merged orthophoto for each panel, we propose to use the *mean percentage of matched features*, $\bar{a}$, calculated as:

$$\bar{a} = \frac{\sum_{i \in O \setminus \{b\}} |\mathcal{F}_i^b|}{|\mathcal{F}_b|} \tag{4}$$

where O is the set of orthophotos associated to the given panel, b is the baseline orthophoto, and $\mathcal{F}_i^b$ is the set of SIFT features from orthophoto i matched to the baseline orthophoto, and $\mathcal{F}_b$ is the set of SIFT features from orthophoto b.

Panel '1' presents the worst result, with the most significant and uneven level of blur. For this panel, $\bar{a}$ = 5%, which is the lowest among all the panels. This is likely due to the large size of the panel, which results in the whole panel not being seen completely in any single image meeting the acquisition minimum requirements in terms of the parameters α_{pi}, d^1_{pi} and d^2_{pi}. As a result, the alignment process described in Section 2.2 starts with one image that only covers a part of the roof. Notwithstanding, in comparison to Figure 1, all the potential occlusions for this panel are successfully removed. The result for the curved Panel '2' is better, with $\bar{a}$ = 16%. This is likely because it is smaller and not as occluded, so that the baseline orthophoto has more matches with the other orthophotos have the most matches with the baseline. In the smallest Panel '3', the details are more clear under magnification: defects like the shallow organic growth appear very clearly. In this case again, $\bar{a}$ = 12%. Note that, for Panel '4', $\bar{a}$ = 10%. In general, The $\bar{a}$ values of panel '2' to '4' are high.

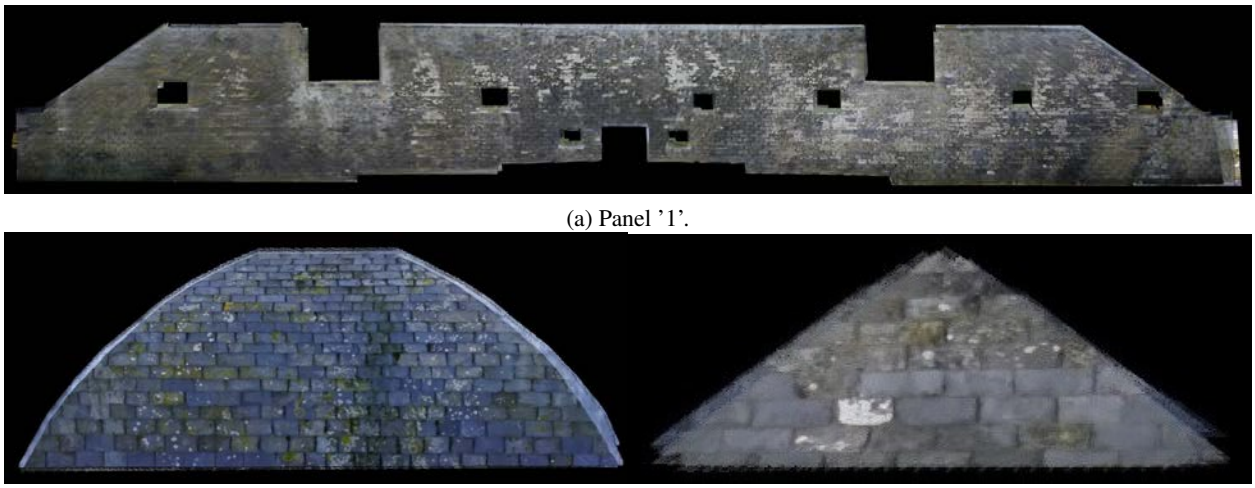
The orthophoto examples in Figure 6d and Figure 10 illustrate two advantages: the slated roofs are clearly and precisely extracted and may thus be well suited for the application of automated defect detection models (e.g. deep learning based models). Besides, no matter which angle was the image taken or where the panel is located, the final orthophotos are all vertical, with the slates appearing as parallel horizonal rows in all images. This should further ease the development and application of robust automated defect detection models — and such representation also remains suited for manual defect annotation using current manual methodologies.

4 Discussion

The novel method for generating vertical orthophotos free of occlusions shows good results. Occlusions are currently handled by analysing the frequency of RGB and selecting the median value. While the method works reasonably well in the four representative examples presented, in some extreme cases, occlusions could appear in a large(r) number of images, which would defeat the proposed approach and cause shadows in the final merged orthophotos. For example, when some decorative components are placed right above or too close to the target panel, only when the UAV is flown exactly at the right location can the entire panel be captured without occlusion, which would require great skills as well as meticulous planning by the operator.

An alternative, more robust approach to deal with occlusions is to use depth information to refine the image masks (see section 2.3). Indeed, if the panel view is unobstructed, the depth associated to a pixel should be the range of the ray going through the pixel and intersecting the roof panel plane. But, if the ray intersects the overall reconstruction mesh (see Figure 2) at an occluding point, then the range will be significantly shorter, and the pixel should be labelled occluded and added to the mask. The remaining steps of the proposed process would then be applied as described.

From the final orthophoto of the curved panel, the slates closer to the top appear narrower than the ones near the bottom, due to the curvature of the panel. This illustrates the limitation of using vertical orthophotos. To solve such



(a) Panel '1'.

(b) Panel '2'. (c) Panel '3'.

Figure 10. Generated orthophotos of typical panels.

a problem, an alternative approach to create orthophotos could be to define the projection plane perpendicular to the mean normal of the roof panel, i.e. along the two principal axes of the panel. Note that this would benefit the generation of orthophotos of planar roof panels as well. The reason why we have focused on generating vertical orthophotos is that this is in line with current practice by surveying professionals (as shown in Figure 1).

5 Conclusion

Current approaches to generate orthophotos for entire building facades have two main limits: lacking semantic link to the BIM model, and unavoidable occlusions in specific building area. To deal with these problems, this study proposes a new approach that generates an orthophoto for each building element, and illustrates the method specifically in the challenging context of roofs. The proposed method combines building reconstruction, camera projection, and orthophoto image generation with occlusion removal. The main benefits of the proposed approach are:

1. Orthophotos contain data for all parts of the roof elements and occlusions by other building elements are eliminated through the proposed merging mechanism that calculates the median orthophoto.
2. By introducing image filtering parameters (i.e. incidence angle and distances), the quality of the final merging orthophoto is further improved, with less blur and seemingly improved colour balancing.
3. Given a 3D model, the orthophoto generation process is fully automatic, and could bring convenience to current surveying practice, as well as emerging digitalised and automated methodologies, with the orthophotos (and any result that would be obtained from their analysis) semantically linked to elements in the Digital Twin or Digital Logbook — of the building.
4. The proposed method can be applied to drone survey data acquired (at frequent intervals in the case of pro-active maintenance schemes) over the life of the building. The orthophotos acquired at each epoch can be used to detect defects and the comparison of the results across epochs be used to assess the evolution of defects over time.

6 Acknowledgements

This paper was made possible thanks to research funding from Historic Environment Scotland (HES) and the Engineering and Physical Sciences Research Council (EPSRC) [grant reference EP/W524384/1]. The views and opinions expressed in this article are those of the authors and do not necessarily reflect the official policy or position of HES and EPSRC. The authors would also like to acknowledge the HES Digital Documentation team for providing us with both the image and point cloud data used in the experiments reported in this paper. The authors appreciate the help from members of Cyberbuild group, including Dr Martin Bueno, Dr Shirin Malihi and Mr Boan Tao. For the purpose of open access, the authors have applied a Creative Commons Attribution (CC BY) licence to any Author Accepted Manuscript version arising from this submission.

References

[1] Paul Berdahl, Hashem Akbari, Ronnen Levinson, and William A. Miller. Weathering of roofing materials – An overview. *Construction and Building*

Materials, 22(4):423–433, 2008. ISSN 0950-0618. doi:10.1016/j.conbuildmat.2006.10.015.

[2] Mark G. Stewart, John D. Ginger, David J. Henderson, and Paraic C. Ryan. Fragility and climate impact assessment of contemporary housing roof sheeting failure due to extreme wind. *Engineering Structures*, 171:464–475, 2018. ISSN 0141-0296. doi:10.1016/j.engstruct.2018.05.125.

[3] C. H. Sanders and M.C. Phillipson. Uk adaptation strategy and technical measures: the impacts of climate change on buildings. *Building Research & Information*, 31(3-4):210–221, 2003. doi:10.1080/0961321032000097638.

[4] Anna Banaszek, Sebastian Banaszek, and Anna Cellmer. Possibilities of use of uavs for technical inspection of buildings and constructions. *IOP Conference Series: Earth and Environmental Science*, 95(3):032001, dec 2017. doi:10.1088/1755-1315/95/3/032001.

[5] Anestis Koutsoudis, Fotis Arnaoutoglou, and Christodoulos Chamzas. On 3D reconstruction of the old city of Xanthi. A minimum budget approach to virtual touring based on photogrammetry. *Journal of Cultural Heritage*, 8(1):26–31, 2007. ISSN 1296-2074. doi:10.1016/j.culher.2006.08.003.

[6] J. Unger, M. Reich, and C. Heipke. Uav-based photogrammetry: monitoring of a building zone. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XL-5:601–606, 2014. doi:10.5194/isprsarchives-XL-5-601-2014.

[7] De-Graft Joe Opoku, Srinath Perera, Robert Osei-Kyei, and Maria Rashidi. Digital twin application in the construction industry: A literature review. *Journal of Building Engineering*, 40:102726, 2021. ISSN 2352-7102. doi:https://doi.org/10.1016/j.jobe.2021.102726.

[8] Gustavo Rocha, Luís Mateus, Jorge Fernández, and Victor Ferreira. A scan-to-bim methodology applied to heritage buildings. *Heritage*, 3(1):47–67, 2020. ISSN 2571-9408. doi:10.3390/heritage3010004.

[9] A. Adán, B. Quintana, S.A. Prieto, and F. Bosché. An autonomous robotic platform for automatic extraction of detailed semantic models of buildings. *Automation in Construction*, 109:102963, 2020. ISSN 0926-5805. doi:10.1016/j.autcon.2019.102963.

[10] Enrique Valero, Antonio Adán, and Frédéric Bosché. Semantic 3d reconstruction of furnished interiors using laser scanning and rfid technology. *Journal of Computing in Civil Engineering*, 30(4):04015053, 2016. doi:10.1061/(ASCE)CP.1943-5487.0000525.

[11] Ayman F Habib, Eui-Myoung Kim, and Chang-Jae Kim. New methodologies for true orthophoto generation. *Photogrammetric Engineering & Remote Sensing*, 73(1):25–36, 2007. doi:https://doi.org/10.14358/PERS.73.1.25.

[12] Gary S Smith. Digital orthophotography and gis. In *Proceedings of the 1995 ESRI User Conference*, pages 22–26, 1995.

[13] Yu Liu, Xinqi Zheng, Gang Ai, Yi Zhang, and Yuqiang Zuo. Generating a high-precision true digital orthophoto map based on uav images. *ISPRS International Journal of Geo-Information*, 7(9), 2018. ISSN 2220-9964. doi:10.3390/ijgi7090333.

[14] Enrique Valero, Alan Forster, Frédéric Bosché, Ewan Hyslop, Lyn Wilson, and Aurélie Turmel. Automated defect detection and classification in ashlar masonry walls using machine learning. *Automation in Construction*, 106:102846, 2019. ISSN 0926-5805. doi:https://doi.org/10.1016/j.autcon.2019.102846.

[15] Koubouratou Idjaton, Xavier Desquesnes, Sylvie Treuillet, and Xavier Brunetaud. Stone-by-stone segmentation for monitoring large historical monuments using deep neural networks. In *Pattern Recognition. ICPR International Workshops and Challenges*, pages 235–248, Cham, 2021. doi:10.1007/978-3-030-68787-8_17.

[16] Koubouratou Idjaton, Xavier Desquesnes, Sylvie Treuillet, and Xavier Brunetaud. Transformers with yolo network for damage detection in limestone wall images. In *Image Analysis and Processing. ICIAP 2022 Workshops*, pages 302–313, 2022. doi:/10.1007/978-3-031-13324-4_26.

[17] Philipp Meixner, Franz Leberl, and Mathieu Brédif. 3d roof details by 3d aerial vision. In *2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops)*, pages 212–218, 2011. doi:10.1109/ICCVW.2011.6130245.

[18] Chi Yuan, Xiaoqing Yu, and Ziyue Luo. 3d point cloud matching based on principal component analysis and iterative closest point algorithm. In *2016 International Conference on Audio, Language and Image Processing (ICALIP)*, pages 404–408, 2016. doi:10.1109/ICALIP.2016.7846655.

[19] David G. Lowe. Distinctive image features from scale-invariant key points. *International Journal of Computer Vision*, 60(2):91–110, 2004. doi:10.1023/B:VISI.0000029664.99615.94.