INVESTIGATION OF JOINT VISIBILITY BETWEEN SAR AND OPTICAL IMAGES OF URBAN ENVIRONMENTS

L. H. Hughes[1], S. Auer[2], M. Schmitt[1]

[1] Signal Processing in Earth Observation, Technical University of Munich (TUM), Munich, Germany - (lloyd.hughes, m.schmitt)@tum.de
[2] Remote Sensing Technology Institute (IMF), German Aerospace Center (DLR), Oberpfaffenhofen, Germany - stefan.auer@dlr.de

Commission II, WG II/4

KEY WORDS: Synthetic aperture radar (SAR), optical remote sensing, feature visibility, data fusion

ABSTRACT:

In this paper, we present a work-flow to investigate the joint visibility between very-high-resolution SAR and optical images of urban scenes. For this task, we extend the simulation framework SimGeoI to enable a simulation of individual pixels rather than complete images. Using the extended SimGeoI simulator, we carry out a case study using a TerraSAR-X staring spotlight image and a Worldview-2 panchromatic image acquired over the city of Munich, Germany. The results of this study indicate that about 55% of the scene are visible in both images and are thus suitable for matching and data fusion endeavours, while about 25% of the scene are affected by either radar shadow or optical occlusion. Taking the image acquisition parameters into account, our findings can provide support regarding the definition of upper bounds for image fusion tasks, as well as help to improve acquisition planning with respect to different application goals.

1. INTRODUCTION

One of the most important examples for the exploitation of complementary information from different remote sensing sensors is the joint use of synthetic aperture radar (SAR) and optical data (Tupin, 2010, Schmitt et al., 2017). While SAR measures the physical properties of an observed scene and can be acquired independently of daylight and cloud coverage, optical sensors measure chemical characteristics, and require both daylight and clear environmental conditions. Nevertheless, optical data is significantly easier to interpret for human operators and usually provides more details at a similar resolution. In contrast to this, SAR data not only includes amplitude information, but phase too, which enables a high-precision measurement of three-dimensional scene topography and the deformations thereof.

The challenge of fusing SAR and optical data is greatest when data of very high spatial resolutions covering complex built-up areas are to be fused. One example for this is very-high-resolution (VHR) multi-sensor stereogrammetry as discussed by (Qiu et al., 2018). In this application sparse tie-point matching is combined with estimation of the corresponding 3D point coordinates. While the study demonstrated the general feasibility of sparse SAR-optical stereogrammetry of urban scenes, it also brought to light the difficulties involved with robust tie-point matching in the domain of VHR remote sensing imagery. These difficulties, which had also been discussed by (Zhang, 2010, Dalla Mura et al., 2015, Schmitt and Zhu, 2016) before, are caused by the vastly different imaging geometries of SAR and optical images. This difference hinders any straight-forward alignment by exploiting the image geo-coding or classical image-to-image registration methods, and makes prior information about the acquisition and 3D scene geometry a necessity. Even with the use of prior 3D scene knowledge, SAR and optical image tie-point matching still relies on image based multi-modal matching methods. However, these methods are not robust to artefacts caused by the fundamental nature of the imaging geometries (Dalla Mura et al., 2015). For example, multi-path signals, speckle and layover in SAR images can create visual features which have no valid correspondence in the optical image. Nevertheless, image similarity metrics might still detect structurally similar areas in the optical image which then leads to incorrectly matched tie-points. Similarly, points visible in the SAR image might be occluded in the optical image and thus could end up incorrectly matched. These incorrectly matched pixels will lead to a degraded, and sometimes meaningless, fusion product.

In order to be able to develop more sophisticated fusion techniques, it is imperative that the causal effects between scene geometry, imaging modality and acquisition parameters are fully understood, such that an intuition can be built up as to what scene parts are jointly visible between SAR and optical images of complex urban scenes.

In this paper we make use of a remote sensing simulation framework in order to get a feeling for the smallest common denominator, i.e. to produce joint visibility maps for VHR SAR and optical images. Using these maps we aim to provide a better understanding of the causal relationships between the various imaging factors and their effects on the upper bound of possible fusion products. For this task, we first extend the SimGeoI simulation framework (Auer et al., 2017) to allow for dense, pixel-wise simulation of SAR and optical images. Using this extended framework, we develop a processing chain to create easily interpretable joint visibility maps of VHR SAR-optical images. Finally, we produce such joint visibility maps for a test dataset consisting of a TerraSAR-X staring spotlight and a Worldview-2 image acquired over the city of Munich, Germany, to provide the first educated estimation regarding the limitation of SAR-optical data fusion for urban scenes.

The remainder of this paper is structured as follows: Section 2 describes our adaptions to the SimGeoI simulation framework, while Section 3 explains how the adapted framework can be

used to generate joint visibility maps. Section 4 shows the results achieved on real experimental SAR and optical very-high-resolution imagery. Finally, we discuss our findings in Section 5 and provide a conclusion in Section 6.

2. EXTENSION OF SIMGEOI FOR JOINT VISIBILITY MAPPING

2.1 The SimGeoI Simulation Framework

SimGeoI (Auer et al., 2017) is an object-level simulation framework which enables automated alignment and interpretation of SAR and optical remote sensing images. The SimGeoI framework makes use of prior scene knowledge, remote sensing image metadata and a ray-tracing procedure in order to simulate the remote sensing images, and derive object level interpretation layers of the scene from these images. The SimGeoI work-flow is summarized in the flowchart shown in Fig. 1.

The prior scene knowledge is defined by a digital surface model (DSM) provided in UTM coordinates. The DSM is represented by a raster file with pixel values describing the height of each point in the scene. The second input, the image metadata, is extracted directly from the original remote sensing images, which also have to be geo-coded to a UTM coordinate system. The image metadata and geometric prior knowledge in the same coordinate system allow for automated alignment of remote sensing images based on simulation techniques.

The first stage of the process consists of filtering and decomposing the raw DSM in order to create a digital terrain model (DTM) and a normalized DSM (nDSM) (Ilehag, 2016). DTM and nDSM are then triangulated in order to form a closed 3D scene model from the 2.5D DSM data. The next stage is to extract sensor parameters from the image metadata. These parameters include sensor perspective, image properties and average scene height and are used to define signal source, sensor perspective, and image parameters for the ray tracing procedure. Surface parameters are defined appropriately in order to separate object (white) from background (black) in generated images. The image simulation then takes place using a sensor specific ray-tracing engine, GeoRaySAR (Tao et al., 2011) for SAR and GeoRayOpt (Auer et al., 2017) for optical images, and the defined scene model and sensor. This ray tracing step is repeated for the DSM, nDSM and DTM, respectively. Finally the simulated images are geo-coded by rotating the images to a north-east orientation, and then correcting for the constant shift caused by different imaging planes between the original image and the simulated images. With this the simulated images are geo-coded into the UTM coordinate system and aligned with both the DSM and the original image data.

Using the simulated images from the DSM, DTM and nDSM, SimGeoI is able to create various object-level interpretation layers of the scene (Auer et al., 2017). These layers include: ground and vegetation extent; as well as shadow and layover in the case of SAR images; and sun shadow and building extent in the optical case. As the simulated images have been aligned to the DSM and are geo-coded in the same coordinate frame as the original images, these interpretation layers can be used to extract and compare object-level features between remote sensing images of the same scene, from different view points or imaging modalities.

While SimGeoI provides accurate image alignment, and various interpretation layers to aid in understanding SAR and optical images, these insights are only applicable to the object-level of a scene. However, to fully understand all the factors involved in joint visibility of image parts and features across multi-modal remote sensing data, and to build up an intuition of the upper bound of fusion products we require a more fine-grained interpretation of the scene.

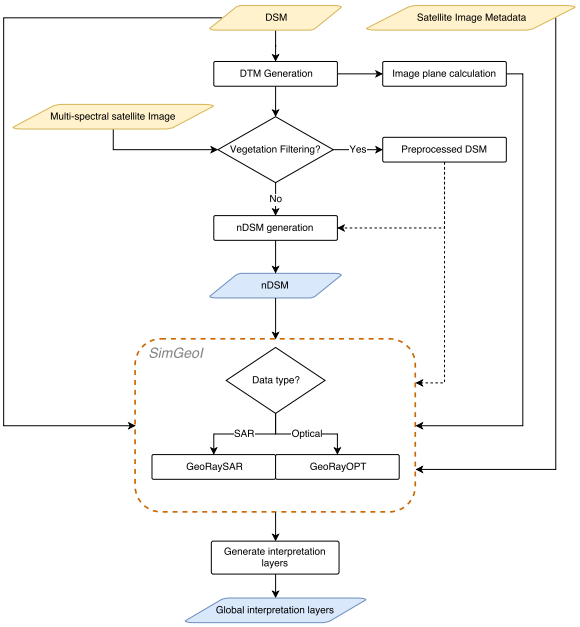


Figure 1. Automated simulation and alignment of remote sensing images with SimGeoI. The red framed section represents the core of SimGeoI which is responsible for the ray-tracing of the DTM, DSM, and nDSM and geo-coding of the resulting images. Yellow: user provided inputs, blue: output products which are used in later processes.

2.2 Extension of SimGeoI for the Simulation of Individual Pixels

In order to perform a detailed analysis of the scene in terms of joint visibility, and uncertainty with respect to artefacts and imaging modality, we extend the SimGeoI framework to enable pixel-level alignment and simulation of the scene. To achieve this pixel-level simulation we add an iterative pixel modelling and ray-tracing procedure as an additional stage to the original SimGeoI pipeline. These additions are depicted in Fig. 2.

Our pixel-level simulation starts by segmenting the preprocessed, non-triangulated nDSM into sub-DSMs using a grid based system. This is done in order to ensure large scenes can be processed in a parallel manner, as each sub-DSM is independent in the ray-tracing phase. Each sub-DSM is then processed in a pixel-wise manner, where each DSM pixel is modelled as a small sphere with its original X, Y coordinates, and a height corresponding to the DSM height at that point. It should be noted that each pixel is used to create a separate 3D model, such that only a single sphere exists in each model. These pixel-wise models are then fed into the ray-tracing engine, along with the camera definition which was created as per the standard SimGeoI simulation procedure. The simulated image, which contains only a single activated pixel, for each pixel-wise model is then geo-coded and aligned with the original remote sensing image. The location of the activated pixel, in UTM coordinates, is then extracted and used to sample the various interpretation layers generated during the object-level simulation. By doing so we are able to not only

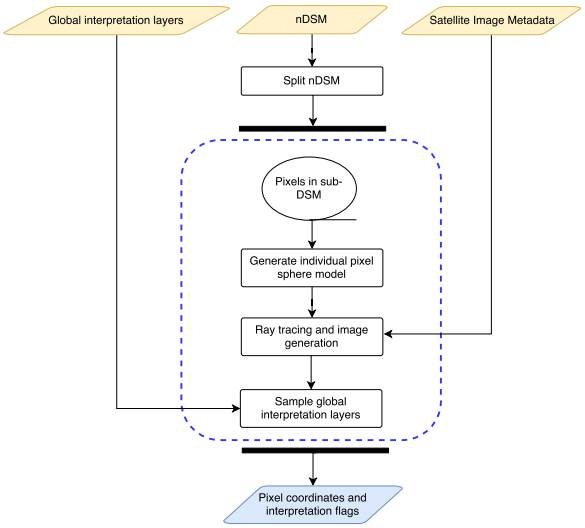


Figure 2. Our extension of the SimGeoI framework to pixel-wise simulation and interpretation. The process encapsulated in the dashed blue area is run in parallel and is independent for each sub-DSM. The results are collated at the end, into a single results file for the specified image. Yellow: inputs which are obtained from the original SimGeoI pipeline in Fig. 1. Blue: Collated output file containing pixel-wise results for a single satellite image.

obtain a pixel-wise correspondence between the multi-modal remote sensing images, as well as image pixel to DSM correspondence, but also a pixel-level interpretation of the scene. The DSM pixel coordinate, simulated image pixel coordinates, and pixel interpretation flags for each pixel are then collated and stored in a tabular format.

It should be noted that due to the DSM being a 2.5D raster representation of the scene, vertical regions in the DSM appear as discontinuities when converted to a 3D point cloud representation. Thus our simulation process is unable to obtain pixel correspondences, and interpretation of the facade regions of buildings. These vertical discontinuities can be seen more clearly in Fig. 3.

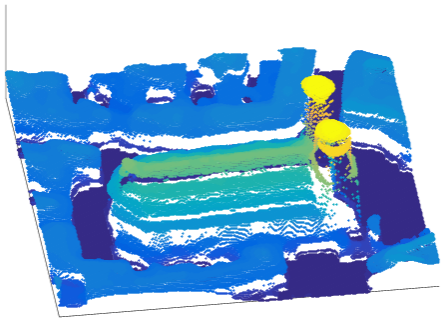


Figure 3. An exemplary point cloud which was extracted from a DSM. The vertical discontinuities are clearly visible as white patches in the point cloud.

Furthermore, as we simulate the DSM pixels individually, imaging effects such as occlusion and radar shadow are not accounted for during simulation. Thus we are able to obtain the theoretical image pixel coordinates for every DSM pixel, irrespective of its true visibility in the original remote sensing image.

3. GENERATING JOINT VISIBILITY MAPS

Using the outputs of the extended SimGeoI framework described in Section 2 for both the SAR and optical images, we are able to derive joint visibility maps for the scene. However, as the DSM pixels are simulated independently, we first need to apply a sensor specific post-processing stage to the results in order to generate additional interpretation layers. These layers are used to impose the original scene geometry constraints on the simulation results. The results from post-processing can then be fused into a final dataset which is used to generate the joint visibility maps. The post-processing and merging process is depicted in Fig. 4 and described in detail below.

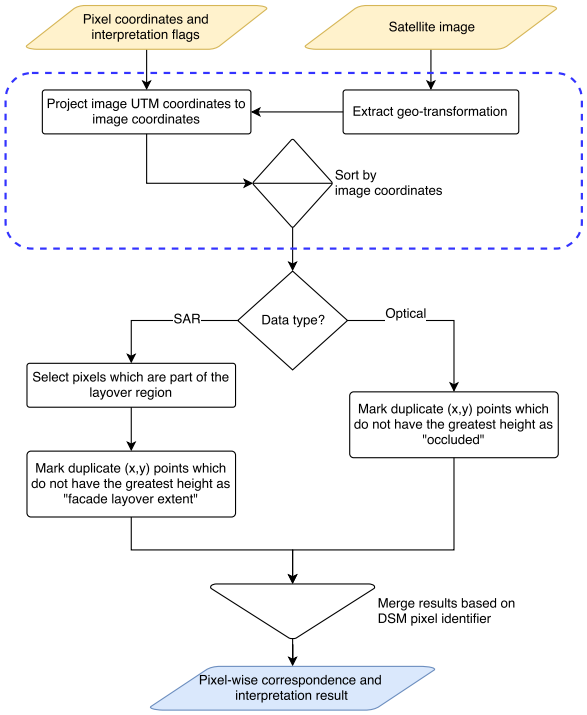


Figure 4. Our post-processing and merging stage. The process highlighted in blue is run separately for both the SAR and optical pixel-wise results. The projected and sorted image coordinates for both the SAR and optical simulation are then processed to enforce geometric constraints and finally merged into a single output result. Yellow: inputs from previous stages of the pipeline, Blue: the final merged and post-processed pixel-wise interpretation and correspondence dataset which is used to create our joint visibility maps.

3.1 Post-Processing

As the simulation results do not account for the geometric constraints of the scene, we use a post-processing step to add additional interpretation flags to each pixel. These flags specify whether the pixel is subject to any geometric constraints. As these constraints are different between SAR and optical images we require a sensor specific approach to post-processing.

In the case of the optical image simulation, as all the DSM pixels are simulated independently it is possible that many co-linear points exist. Co-linear points are points in the 3D scene which line along the same line of projection, and thus are not truly visible as only the point closest to the camera will be seen. The

other points along this line of projection will be occluded. For this reason we add an additional interpretation flag to the optical simulation results specifying whether a simulated pixel is occluded or not. In order to determine co-linear points we make use of a simple strategy which does not require storing intermediate ray-tracing products. Firstly the geo-coded image pixel coordinates are converted to image (x, y)-coordinates such that co-linear points have the same (x, y)-coordinates in the image space. We then select the image pixel which has the greatest corresponding DSM pixel height as the visible pixel, and define all other pixels as being occluded. This strategy holds due to the fact that the remote sensing images we are using are guaranteed to be taken from an aerial vantage point within a relatively small range of image incidence angles. A visual description of why this assumption and technique works can be seen in Fig. 5.

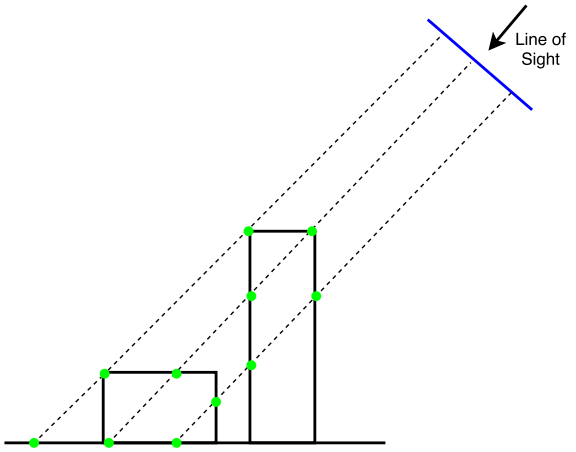


Figure 5. Simplified imaging geometry of an optical satellite sensor. It can be seen that colinear spheres (green), will project to the same image plane (blue) coordinates. However, only the sphere with the greatest height will truly be visible on the image plane.

For SAR images, post-processing is used to determine the extent of facade layover in the image. While SimGeoI provides a layover interpretation layer, this layer masks all scene object-pixels which are subject to layover. However, as the roof structure remains the same and is not often heavily distorted by layover we wish to exclude it from this mask. The reasons for excluding the roof region of buildings is that this region is often jointly visible and may contain important features. Layover pixels are additive in nature and contain, for instance, signal components from both the ground and a building. For this reason we wish to only mask the layover regions which contain ground signal and signal from the facade of the building, not the roof. this is achieved by converting the geo-coded image coordinates to (x, y)-pixel coordinates, and then extracting the pixel with the greatest height to be the building roof. The duplicate pixels are then defined as the layover extent of the building facade. This strategy holds as only a direct signal response occurs on the surface of the modelled DSM pixel sphere. A visual argument for this post-processing stage is depicted in Fig. 6.

3.2 Merging SAR and Optical Simulations

As the SAR and optical images are simulated independently of each other, it is required that we merge their simulation files in order to be able to assess joint visibility between the original images. When we split the nDSM into sub-DSMs we make use of

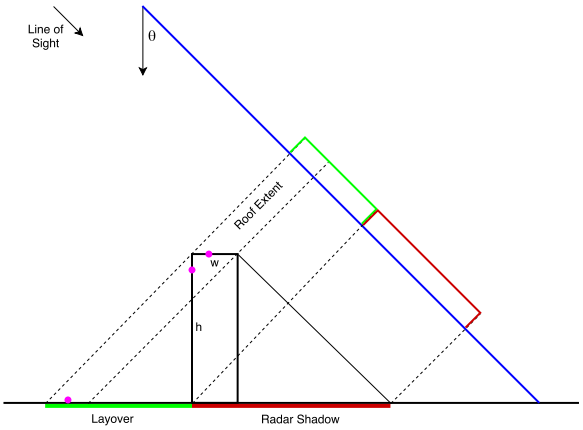


Figure 6. Simplified model of a SAR sensor, and the formation of layover (green) and shadow (red) in a simple scene. The magenta spheres will map to the same image coordinates in the layover region. However, we can ignore the point on the building facade as it is not modelled due to the DSM being 2.5D. Thus by selecting the point with the greatest height we are able to extract the roof extent of the layover, and thereby can obtain the extent of the facade.

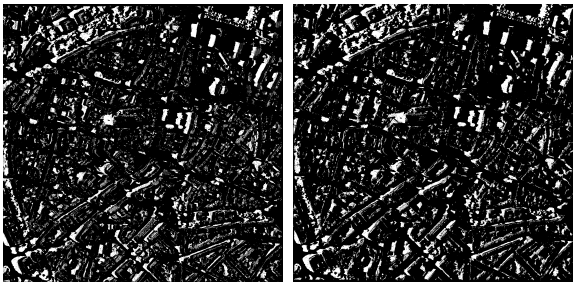
a grid based strategy, such that each grid block can be assigned a unique identifier. Furthermore, when processing the individual pixels in each sub-DSM, the pixels are labelled and processed in a left to right, top to bottom manner. This ensures that each DSM pixel has a unique identifier. Additionally, the SAR and optical simulations make use of the same DSM, thus the DSM identifiers in the SAR and optical image simulation results are equivalent and can be matched by a simple inner join on the data. This enables us to easily determine corresponding pixels between the original SAR and optical images as well as the joint visibility of pixels based on filtering the merged result set by features described in the various interpretation layers and marking the appropriate pixels in the original images. For exemplary demonstration, a small subsection of a final merged simulation result set is presented in Tab. 1.

Table 1. An example of a merged simulation output. Note: the UTM coordinates have been reduced in precision for formatting reasons.

block_id	B2674	B2593	B2594
point_id	P186	P889	P341
UTMx_sar	691489.874	691481.371	691477.364
UTMy_sar	5334883.531	5334887.031	5334881.032
height_sar	655.292	655.290	655.282
shadow_sar	0	0	0
layover_sar	1	1	1
ground_sar	0	0	0
facade_sar	True	False	False
UTMx_opt	691414.419	691405.919	691401.919
UTMy_opt	5334878.698	5334882.199	5334876.201
height_opt	655.292	655.290	655.282
shadow1_opt	0	0	0
layover1_opt	274	0	498
ground1_opt	0	0	0
layover2_opt	274	0	498
shadow2_opt	0	0	0
ground2_opt	0	0	0
occluded_opt	False	True	False

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-IV-2-129-2018 | © Authors 2018. CC BY 4.0 License.

In order to generate the joint visibility maps we use the merged pixel-wise simulation product, as well as the original remote sensing images. Using these data, generating joint visibility maps for both the SAR and optical images becomes a trivial task. By filtering the dataset to only include the points which make up a specific interpretation layer in either the SAR or optical image, we are able to exploit the list of corresponding SAR and optical image coordinates and plot the extent of this interpretation layer in both images. An example of a joint interpretation layer generated in this manner is depicted in Fig. 7.



(a) (b)

Figure 7. An example of an interpretation layer mask. The extent of radar shadow in the SAR image (a), as well as the extent of shadowed pixels in the optical image (b), is shown in white. Black pixels are unaffected by radar shadow.

4. EXPERIMENT AND RESULTS

4.1 Test Data

For our experiments we make use of a dataset consisting of VHR optical and SAR images, as well as a DSM of the city of Munich, Germany. The DSM of the Munich scene was derived from a Worldview-2 stereo image pair and has a horizontal resolution of $0.5m$ and vertical resolution of $1m$. The details of the remote sensing images are summarized in Tab. 2.

Table 2. Parameters of the test images over Munich, Germany

Data	**WorldView-2**	**TerraSAR-X**
Acquisition Date	12/07/2010	07/06/2008
Imaging Mode	panchromatic	staring spotlight
Off-nadir angle (at scene center)	$14.5°$	$49.9°$
Orbit	770km	515km
Heading angle	$189.0°$	$188.3°$ descending
Pixel spacing (east, north)	0.5m	0.5m

4.2 Joint Visibility Map Results

In order to understand which pixels are visible in both the SAR and optical images, we propose the concept of cross-modal and joint visibility maps. These maps describe which pixels can be seen in both images, and thus which pixels are appropriate for matching and fusion applications such as stereogrammetry or tie point detection for image registration.

By masking the facade layover extent and optical occlusion interpretation layers in the SAR image, and the radar shadow and facade extent layers in the optical image, cross-modal joint visibility maps are generated for the scene described in Section 4.1. These cross-modal joint visibility maps can be seen in Figs. 8 and 9. A cropped area around the Frauenkirche (church) is depicted in Fig. 10. For easier reference, the extent of this area is marked by a white frame in Figs. 8 and 9.

In addition to these cross-modal joint visibility maps, we create a joint visibility map which is the projection of both cross-modal joint visibility maps onto an ortho-image, in our case an OpenStreetMap layer. This joint visibility map, seen in Fig. 11, represents the full extent of visible, non-visible and uncertain regions of the scene with respect to both sensors.

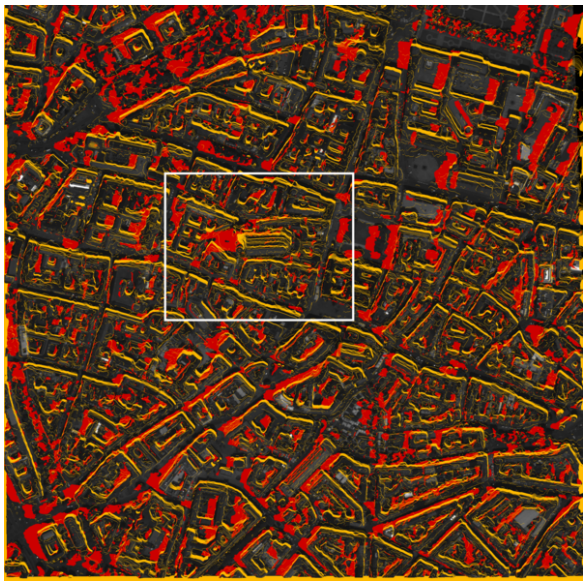


Figure 8. Cross-modal joint visibility map of Munich projected onto the WorldView-2 image. Red: radar shadow extent; yellow: building facades.

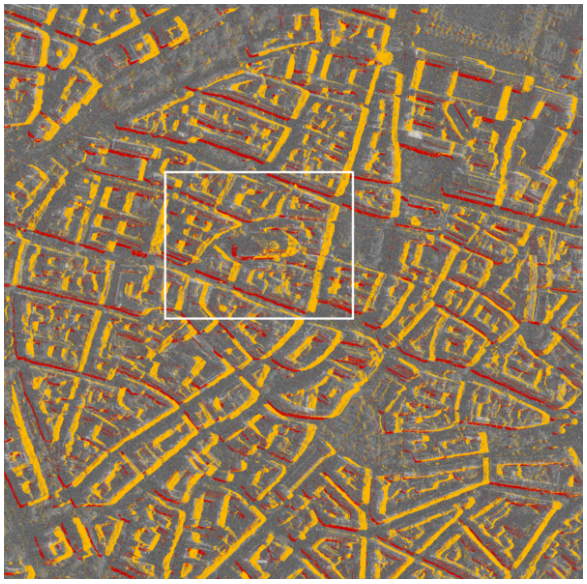
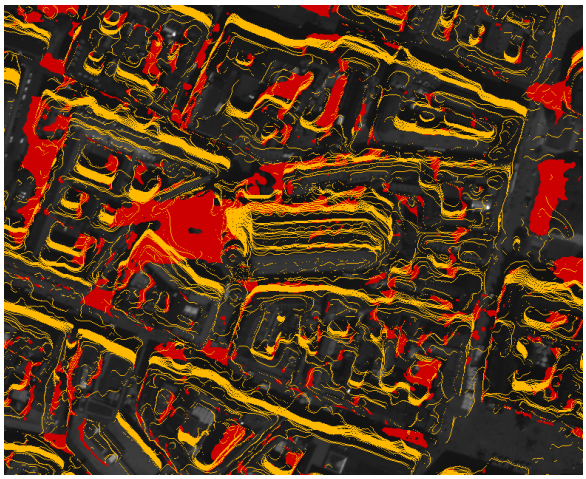
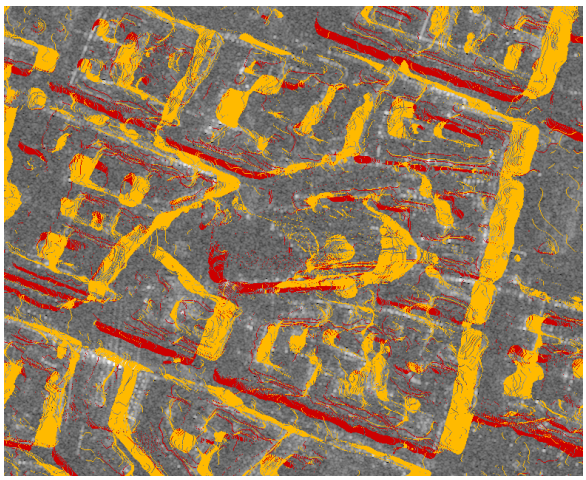


Figure 9. Cross-modal joint visibility map of Munich projected onto the TerraSAR-X image. Red: optical occlusions; yellow: facades layover extent.

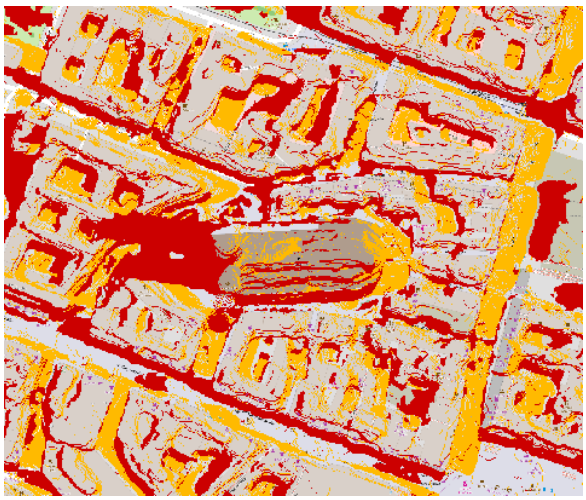
The red pixels in these figures represent regions of each image which are not visible in the other modality. For example, in the



(a)



(b)



(c)

Figure 10. Cross-modal joint visibility maps for optical (a) and SAR (b) images; and joint visibility map (c), of Frauenkirche (church) Munich. (a) Red: radar shadow extent; yellow: building facades. (b) Red: optical occlusions; yellow: layover facade extent. (c) Red: Not jointly visible points; yellow: uncertain vertical points (i.e. facades).

case of the optical joint visibility map shown in Fig. 8, the red pixels represent areas of the optical image which cannot be seen in the SAR image due to radar shadow. The yellow pixels in the joint visibility maps describe regions in the image which have high uncertainty with respect to matching, or whose visibility is dependent on the spatial relationship between the sensors and the geometric distortion effects which occur during imaging. For instance, in the SAR visibility map (Fig. 9), the yellow regions represent the extent of building facade in the layover region, while the yellow in the optical visibility map describes the extent of the facade in the optical image. In the case of the joint visibility map, Fig. 11, the red and yellow pixels are formed by combining the results of the cross-modal joint visibility maps described above.

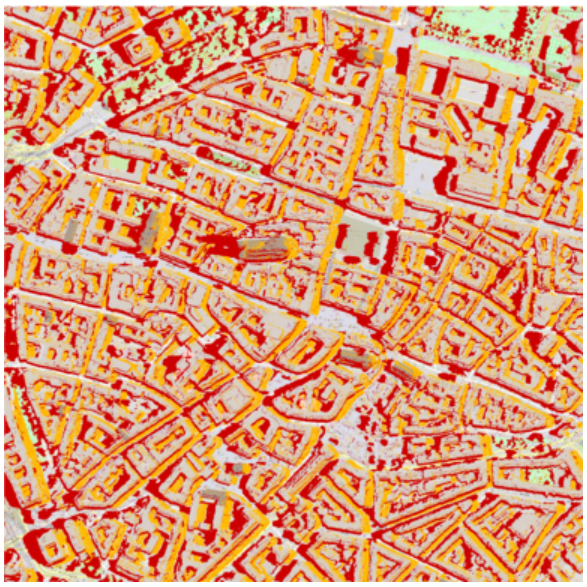


Figure 11. Joint visibility map of Munich projected onto an OpenStreetMap layer. Red: image parts that are not jointly visible due to radar shadow or optical occlusion; yellow: uncertain vertical areas (e.g. facades).

The regions in red cannot be matched and thus do not contribute to the fusion product as they are only visible in one of the images. In contrast, the areas in yellow can still provide useful data, and high quality matching results, if the imaging parameters and scene structure are such that:

- the SAR and optical sensors image the same facade,
- the layover of the facade does not overlay another area with prominent signal response,
- the image matching technique does not rely purely on image geo-coding for defining search areas,
- the scene structure is such that the building facades produce matchable features in both the SAR and optical domain.

5. DISCUSSION

The results presented in Section 4 show that even when accounting for imaging effects such as radar shadow, optical occlusions and facade uncertainty, a significant portion of the scene remains jointly visible, even in complex urban scenes. However, many effects such as sensor baselines, scene geometry, and sensor viewing angels affect the extent of non-visible and uncertain pixels. In this section the effects of these factors on the joint visibility of the scene will be discussed.

5.1 Effect of Sensor Baseline

The baseline between the SAR and optical sensors determines the extent of the scene which is imaged. From our test scene we can see how a relatively wide baseline, coupled with different viewing directions, leads to the SAR and optical sensors capturing different building facades. Furthermore, this non-zero baseline also introduces larger regions of non-jointly visible points as the radar shadow and optical occlusions do not overlap, as is clear when comparing the cross-modal joint visibility maps (Figs. 8 and 9) to the final joint visibility map (Fig. 11).

As it was shown by (Qiu et al., 2018) in order to have favourable conditions for stereogrammetry, the baseline between the sensors should be as small as possible. This small baseline is also favourable for joint visibility. It ensures that the radar shadow (red pixels in Fig. 8) overlaps with the points which are occluded in the optical images (red pixels in Fig. 9), thus decreasing the non-visible regions.

However, a small baseline is not favourable for SAR-optical image matching as the layover of the building falls towards the sensors on the SAR image plane, while the building extent in the optical image falls away from the sensor. Thus it increases the number of uncertain (yellow) pixels in our joint visibility map. Furthermore, building facade images appear mirrored with respect to each other, while the roof structure remains in the same orientation, thus making purely image-based matching approaches more difficult. While prior information about the scene can assist in determining search regions to find corresponding features, and can provide information as to flips and rotations required for patch comparison, the matching of these features remains a difficult task.

5.2 Effect of Viewpoint and Scene Geometry

The viewing angle of the sensors on the scene combined with the scene geometry have the largest part to play in the joint visibility of scene parts. From the results presented in Fig. 10a we are able to see how the high Frauenkirche building causes a large number of pixels to be lost in the optical image due to the extensive radar shadow experienced at a viewing angle of $\theta = 49.9^\circ$. The extent of the radar shadow can be reduced by decreasing the viewing angle. However, this is at the cost on increasing the extent of the layover. In order to ensure that the layover does not fall on nearby buildings, and thereby cause interference with other feature rich areas, it is beneficial to ensure that the extent of the radar shadow is larger than the extent of the layover. From our test scene and resulting joint visibility map, Fig. 11, we can observe that the layover region is smaller than the shadow region, as there is little overlap between red and yellow pixels. This favorable condition is always true for incidence angles greater than $\theta = 45^\circ$.

In the optical case we see that it is preferable to have a viewing angle as close to nadir as possible. In doing so the number of ground points which are occluded by building structures (red pixels in Fig. 9) is minimized. Furthermore, a small viewing angle also reduces the extent of the building facade seen (yellow pixels in Fig. 8) and thus the uncertainty in matching facade regions. Unlike the SAR imaging case, there is no trade-off between a large and small viewing angle in the optical case. For our optical test data, a small viewing angle of $\theta = 14.5^\circ$ was used, and the resulting cross-modal joint visibility map depicts this in the small extent of the facade and occluded regions.

In order to decrease the number of not jointly visible pixels, the smallest viewing angle obtainable by the SAR sensors should be utilized (20° for TerraSAR-X). However, while this provides the greatest joint visibility the extent of the uncertain regions will be large, and thus could degrade matching accuracy and fusion products. For this reason we argue that the optimal viewing angle needs to be considered with reference to the application and scene structure at hand, in order to ensure accurate feature matching can occur but also that a large enough number of pixels remain available to produce a meaningful fusion product.

5.3 Upper Bound of Data Fusion

Apart from developing an intuition as to how scene geometry, viewing angles and sensor baseline play a role in joint scene visibility, we can also extract a theoretic upper bound for data fusion from our joint visibility maps. In order to do this we obtain quantitative results as to the coverage of the scene, when viewed from a nadir angle. These results are presented in Table 3, both from the point of view of the individual images as well as regarding the full scene extent.

Table 3. Breakdown of the scene coverage of various layers in the cross-modal and joint visibility maps.

Image Type	Not Jointly Visible	Uncertain	Jointly Visible
SAR Image	9.50%	17.77%	72.73%
Optical Image	14.53%	14.73%	70.74%
Scene	25.89%	18.89%	55.22%

From the breakdown in Table 3 it is clear that in our test scene only slightly more than half of its extent is jointly visible in both the SAR and the optical satellite image, while the rest is either missing because of optical occlusion or radar shadowing, or uncertain because of belonging to vertical surfaces (i.e. facades). As the imaging geometries of our test scene are typical and are not extreme in viewing angle, scene geometry nor sensor baseline, it can be inferred that this upper bound is likely achievable for scenes of a similar nature.

5.4 Simulation Limitations

As our simulation process makes use of a 2.5D DSM, several limitations exist in our output data. The main limitation is that we cannot obtain pixel-level correspondences on building facades, even when both sensors image the same facade. This leads to building facade pixels being missing from the final merged simulation results, and thus we cannot draw precise conclusions as to the level of joint visibility present in the facade regions. However, we can infer the possibility of joint visibility based on our joint visibility maps, and the sensors viewing angles of the scene.

Furthermore, due to not modelling facades, it is possible that incorrect correspondences can occur when the visible co-linear point lies on a building facade and occluded points on a building rooftop or on the ground. We can see this situation by observing the scene in Fig. 5 and noting how the ray passing through all the facades of the tall building may land upon the roof of the lower building and thus provide an incorrect response. However, due to optical remote sensing data having a look angle of less than 45°, and more commonly less than 25°, as well as the optical scene being modelled using an nDSM, such a situation will only occur

with closely spaced buildings of significantly different heights. Furthermore, in order for such an error to have a negative influence on the accuracy of the joint visibility maps, the incorrectly labelled point needs to occur with the same incorrect label in both the SAR and optical cross-modal visibility maps.

In the case of the SAR simulation, the effects of not modelling the building facades are not as apparent. This is due to the fact that the rooftop and facade points layover onto the ground, and while the facade pixels are not simulated these ground and rooftop pixels are, thus encapsulating the full extent of the layover.

6. CONCLUSION AND OUTLOOK

Through our experiments for the first time a strong intuition on the bounds of joint visibility in multi-modal remote sensing was gained – backed by quantitative results. To achieve this, we developed a framework which allows for pixel-wise correspondence to be determined between multi-modal remote sensing images. This framework can provide the basis for many other applications involving the investigation of joint-visibility as well as for data acquisition in applications where high quality labelled data and correspondence information is required, such as training deep matching algorithms.

We further developed an intuition as to the appearance and effect of the various factors involved in the imaging of the scene. We were able to show why a small baseline between the sensors is favourable for stereogrammetry applications. We further described the trade-off between non-visible regions and uncertain regions and present an argument for why the selection of the scene viewing angle is mainly dependent on factors influencing the SAR image. Our results further describe the joint visibility for our test scene is around 55%, even without any optimizing of viewing angle or sensors baselines. This number can serve as an approximate upper bound for matching and image fusion endeavours. Since our test scene was fairly typical, it can be expected that this upper bound approximately extends to scenes with a similar structure and imaging geometry.

In future work the simulation of the building facades will be included in order to gain a more accurate understanding of the nature of uncertain areas in the image, and to what degree these areas remain uncertain and difficult to match. An investigation into the visibility of strong feature points, and their transferability between the SAR and optical domain will be discussed, with the aim of assisting in the selection of high quality feature points and regions to aid matching in SAR-optical stereogrammetry. We will further present a mathematical framework to allow for easier selection of an optimal viewing angle and baseline for use in matching and SAR-optical stereogrammetry data acquisition.

ACKNOWLEDGEMENTS

This work is supported by the German Research Foundation (DFG) under grant SCHM 3322/1-1.

REFERENCES

Auer, S., Hornig, I., Schmitt, M. and Reinartz, P., 2017. Simulation-based interpretation and alignment of high-resolution optical and SAR images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 10(11), pp. 4779–4793.

Dalla Mura, M., Prasad, S., Pacifici, F., Gamba, P., Chanussot, J. and Benediktsson, J. A., 2015. Challenges and opportunities of multimodality and data fusion in remote sensing. *Proceedings of the IEEE* 103(9), pp. 1585–1601.

Ilehag, R., 2016. Exploitation of digital surface models from optical satellites for the identification of buildings in high resolution SAR imagery. Master’s thesis, KTH, Sweden.

Qiu, C., Schmitt, M. and Zhu, X. X., 2018. Towards automatic SAR-optical stereogrammetry over urban areas using very high resolution images. *ISPRS Journal of Photogrammetry and Remote Sensing* 138, pp. 218–231.

Schmitt, M. and Zhu, X. X., 2016. On the challenges in stereogrammetric fusion of SAR and optical imagery for urban areas. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 41(B7), pp. 719–722.

Schmitt, M., Tupin, F. and Zhu, X. X., 2017. Fusion of SAR and optical remote sensing data – challenges and recent trends. In: *Proceedings of IEEE International Geoscience and Remote Sensing Symposium*, Fort Worth, TX, USA, pp. 5458–5461.

Tao, J., Palubinskas, G., Reinartz, P. and Auer, S., 2011. Interpretation of SAR images in urban areas using simulated optical and radar images. In: *Proceedings of Joint Urban Remote Sensing Event*, pp. 41–44.

Tupin, F., 2010. Fusion of optical and SAR images. In: *Radar Remote Sensing of Urban Areas*, Springer, pp. 133–159.

Zhang, J., 2010. Multi-source remote sensing data fusion: status and trends. *International Journal of Image and Data Fusion* 1(1), pp. 5–24.