DIGITAL ELEVATION MODELS FROM STEREO, VIDEO AND MULTI-VIEW IMAGERY CAPTURED BY SMALL SATELLITES

Pablo d'Angelo, Peter Reinartz

German Aerospace Center (DLR), Remote Sensing Technology Institute, 82234 Wessling, Germany
(Pablo.Angelo, Peter.Reinartz)@dlr.de

Commission II, WG 2

KEY WORDS: digital surface models, dense matching, small satellite, stereo, multi-view

ABSTRACT:

Small satellites play an increasing role in earth observation. This article evaluates different possibilities of utilizing data from Planet's SkySat and PlanetScope satellites constellations for derivation of digital elevation models. While SkySat provides high resolution image data with a ground sampling distance of up to 50 cm, the PlanetScope constellation consisting of Dove 3U cubesats provide images with a resolution of around 4 m. The PlanetScope acquisition strategy was not designed for stereo acquisitions, but for daily acquisition of nadir viewing imagery. Multiple different products can be acquired by the SkySat satellites: Collects covering an area of usually 12 by 6 km, tri-stereo collects and video products with a framerate of 30 Hz. This study evaluates DSM generation using a Semi-Global Matching from multi date stereo pairs for SkySat and PlanetScope, and the dedicated Video and tri-stereo SkySat acquisitions. DSMs obtained by merging many PlanetScope across track stereo pairs show an normalized median deviation against LiDaR first pulse data of 5.2 meter over diverse landcover at the test sites around the city of Terrassa in Catalonia, Spain. SkySat tri-stereo products with 80 cm resolution reach an NMAD of 1.3 m over Terrassa.

1. INTRODUCTION

In the recent years small satellites have become an important tool in remote sensing. For example Planet operates a the PlanetScope constellation which can monitor the whole earth daily in 4 m resolution, as well larger SkySat satellites, allowing sub meter VHR coverage. These are mostly utilized for derivation of 2D information, such as semantic segmentation, object and change detection. Only a few studies have analyzed the possibilities of the extraction of digital surface models from the PlanetScope and SkySat data and almost all have focused on mountains, glaciers or mining areas. In contrast this paper evaluated data over typical urban and hilly terrain.

2. METHODOLOGY

We first describe the general properties of the satellites utilized in this area, and then detail the data processing workflow.

2.1 SkySat

The SkySat satellites with a mass of 100 kg carries a telescope with 3 frame sensors. Each SkySat satellite uses 3 CMOS frame detectors with a size of 2560x2160 pixels and a pixel size of 6.5 μm. The upper half of the detector is used for panchromatic capture, the lower half is divided into 4 stripes covered with blue, green, red and near infra-red color filters. The native resolution at nadir of the SkySat-1 and SkySat-2 is around 1.1 m. The later satellites are placed in lower orbits, leading to increased image resolution of 0.72 to 0.81 m, and up to 0.5 m after super-resolution. (Planet Imagery Product Specifications, 2021)

Two products are available: Panchromatic video with a resolution of around 1 meter and a frame size of 2560x1080 pixels at 30 frames per second. Additionally, the satellite can collect still imagery with a swath of 6 km, name Skysat Collect. The SkySat Collect product contains panchromatic and 4 multispectral channels and consists consists of 3 stripes of frame images with a footprint of approximately 2.5 x 1 km. The Video and Collect products contains both a physical camera model and a RPC for each individual frame.

Additionally, a tri-stereo Collect product consisting of one central collection at the closest approach with maximum satellite elevation and two collections with 27.5 degrees convergence angle to the center product. While smaller convergence angles would be beneficial for built-up areas, the satellite agility is not high enough for smaller convergence angles. A tri-stereo acquisition typically consists of around 100 individual images.

2.2 PlanetScope

The Doves 3U cubesats allow daily acquisition of the earths landmass at a resolution of 4 m with 4 to 8 spectral bands and a swath with of 24 km. Images are typically acquired with viewing angles within +/- 6 degrees of nadir. This results in the highest image resolution and constant perspective, which is helpful for 2D image analysis, but not ideal for stereo matching, as only image pairs with very small baselines are available. While the lower resolution and small baseline of the current acquisitions are limiting the accuracy of 3D information, the daily world-wide coverage is an important factor, which potentially allows generation of a global DSM, and dense reconstruction even for challenging areas with high cloud cover and difficult terrain.

The PlanetScope imagery is available in image frames covering approximately 24 x 6 km, in both orthorectified and non-orthorectified versions. Planet uses a reference image and DEM to refine the geolocation of the imagery and provides RPC coefficients for the non-orthorectified imagery, enabling stereo processing or custom orthorectification.

2.3 Related work

SkySat still and video products were evaluated in (d'Angelo et al., 2016) in terms of radiometry and geometry. The video dataset has been used for DSM generation over Las Vegas. However, tri-stereo products were not available at that time. DSM generation from PlanetScope imagery (Ghuffar, 2018) was proposed and showed that PlanetScope Imagery could be used for DSM generation for extreme mountainous terrain with reasonable accuracy. Later works by (Aati, Avouac, 2020) proposes a multi-step RPC based bundle adjustment procedure for SkySat Triplet acquisition and PlanetScope imagery and applies these two mining and glacier monitoring applications. A more recent work (Bhushan et al., 2021) adds SkyBox support to the AMES Stereo Pipeline (Beyer et al., 2018) by utilizing a frame camera model instead of the RPC model during image orientation and dense matching of SkySat tri-stereo and video products. Their work proposes a free network adjustment and rigid co-registration of the resulting DSM to a reference DSM and has been released as open source software. Evaluation is performed over mountainous and glaciers.

In contrast to most prior works, this paper evaluates the performance of PlanetScope and SkySat imagery in less extreme conditions such as urban or hilly terrain and without specially tailored bundle adjustment procedures.

2.4 DSM derivation

Processing of all imagery starts with image orientation using the provided RPC coefficients. Planet provided refined RPCs, based on a world-wide reference image and height model. However, as dense matching requires sub-pixel accurate relative orientation between all images, bundle block adjustment is used to estimate a RPC bias correction for each image, based on tie points and the reference height model, to ensure a relative orientation with sub-pixel accuracy as well as good co-registration between the images and the reference height model (d'Angelo, Reinartz, 2012). Optionally, matching against a reference ortho image can be used to derive GCPs for improving the horizontal geolocation. This is especially important for SkySat Collects with many sub-images, as bundle-adjustment with tie points only results in a weak geometry due to the high number of small images and the need for 6 parameter affine corrections for each image.

DSM generation is performed using SGM (Hirschmüller, 2008, d'Angelo, Reinartz, 2011) using the Census transform (Zabih, Woodfill, 1994) as similarity measure. Images are matched pairwise and DSMs are generated for each stereo pair. To increase the subpixel disparity accuracy, the a subsampling of 0.5 pixel in disparity direction is performed for SkySat products, and an even higher subsampling of 0.33 pixels is used for PlanetScope stereo pairs, due to their limited disparity range and the high sensitivity to sub-pixel matching errors.

Merging of pairwise DSMs is performed by selecting the median height of all valid matches. For PlanetScope and SkySat Across-Track Imagery, pair selection is based on stereo convergence angle and acquisition time difference. All 3 possible pairs are matched SkySat Tri-Stereo acquisitions. From the SkySat video sequences, every 30th image is used and matched against all images with a time difference of more than 6 seconds.

3. EVALUATION

Four test regions are used in this work, three in in Catalonia, Spain near Barcelona, and one in Berlin, Germany. Multiple satellite datasets as well as LiDaR based ground truth are available for both areas. Images of the areas in Catalonia are shown in Fig. 1).
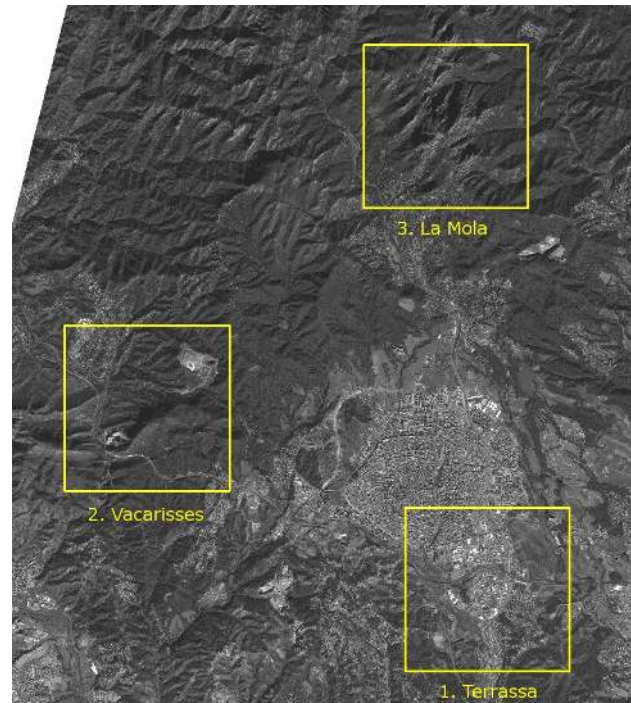

Figure 1. The three test areas near Barcelona, Spain

The areas in Catalonia covers a large variety of landscapes, including dense city, industrial areas, forested hills and steep mountains. It is thus very well suited for a comparison of performance in diverse conditions. A total of 1323 PlanetScope images captured between 2017-09-10 and 2020-06-15 were used in this study. As PlanetScope images are not collected for stereo reconstruction, but for 2D mapping, they are acquired close to nadir, strongly limiting the possible stereo convergence angles, cf. Fig. 2.

SkySat Tri-Stereo acquisitions for the three sub-areas were captured between 2019-06-10 and 2019-06-19. Unfortunately, SkySat data availability over this area limited. Berlin was used as additional site for DSM evaluation as SkySat Video, Tri-Stereo Collect and a time series of mono SkySat Collect products were available.

3.1 Image orientation

Before DSM generation, bundle adjustment was used to estimate image space RPC corrections. For PlanetScope Imagery, a 2 parameter row and column shift in image space was sufficient. The SkySat imagery required an affine 6 parameter correction. During adjustment of the PlanetScope imagery, SRTM was used as weakly weighted reference height model (d'Angelo, Reinartz, 2012), mainly to avoid issues due to the weak intersection geometry. As the SkySat Imagery requires larger corrections, the a LiDaR DSM as well as 10117 ground control points (GCPs) automatically extracted from airborne ortho images of the ICGC (Institut Cartogràfic i Geològic de Catalunya,
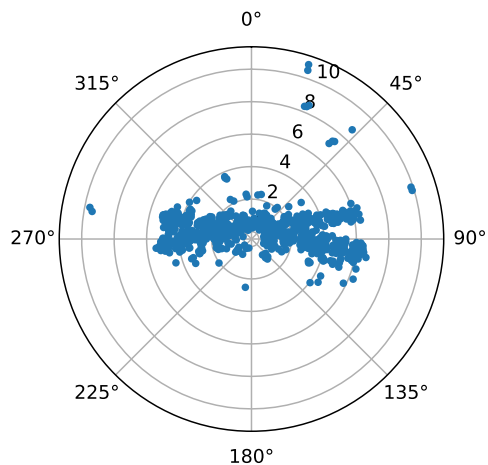
Figure 2. Skyplot showing the incidence and azimuth angles of the PlanetScope imagery. Only few images with incidence angles larger than $6°$ are available.

2020) were used as reference during the adjustment. The large number and small overlap of the individual images in SkySat Collect product lead to an unstable block configuration during RPC bundle adjustment, requiring GCPs for accurate alignment to the reference, cf. (d'Angelo et al., 2016).
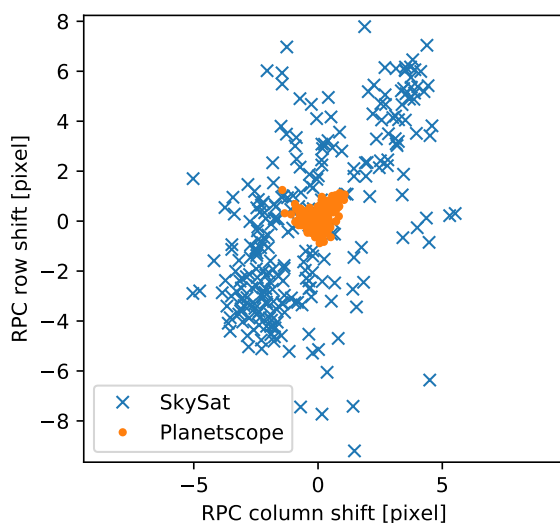

Figure 3. Image space RPC bias corrections for SkySat Triplet and PlanetScope imagery.

The RPCs of the PlanetScope imagery were already very consistent, the average magnitude of the computed corrections was 0.2 pixel, cf Fig. 3. The SkySat imagery required larger RPC corrections: The average RPC bias correction of the 263 image frames of the Catalonia Tri-Stereo acquisitions was 3.8 pixels.

For the Berlin data, we observed stronger non-linear, systematic errors in some images of the used still imagery products, leading to local height offsets in DSMs. We thus only performed an qualitative comparison of the different DSMs derived for the Berlin data. Future investigation into this issue are required, for example by additionally performing bundle adjustment using a

physical sensor model (Bhushan et al., 2021), or by using the SkySat L1A RPCs based on ephemeris data only.

3.2 DSM evaluation

For the PlanetScope imagery, only the area in Catalonia is evaluated, as the lower resolution and low height sensitivity do not allow detailed reconstruction of densely built up areas. Due to the limited viewing angle variation, cf. Fig. 2, only a few stereo pairs with large angle can be found. Fig. 4 shows the height errors of 100 stereo pairs. Additionally, the images of the pairs with convergence angles $> 10°$ where acquired multiple months apart, reducing the accuracy of dense matching due to temporal changes of the landscape and sun illumination. Stereo pairs covering a convergence angle range of 1 to 14 degrees with minimum time difference were selected. The larger convergence angles pairs ¿ $10°$ are few and have been acquired with a large time differences of many months, increasing the difficulty for dense image matching due to vegetation changes and different appearance. The robust error statistics reported in Fig. 4 indicate that pairs with convergence angles larger than $6°$ show similar error characteristics and are selected for DSM merging. A fused DSM was computed by taking the median height of the selected pairs, cf. Fig. 5.
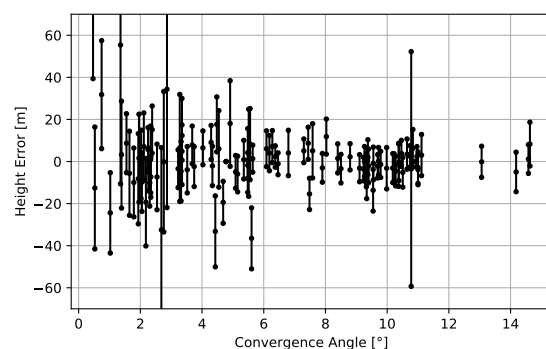

Figure 4. Height error spread vs stereo convergence angle for the 100 selected PlanetScope stereo pairs. The median is given by the central point on the lines, the spread is indicated by the normalized median deviation (NMAD)

Visual inspection of the Berlin DSMs, cf. Fig 6 show that video acquisition produces the smoothest and most complete results, the tri-stereo acquisition performs well in open areas, but is incomplete in densely built up neighbourhoods and forest. This is caused by the relatively large convergence angle between the individual datatakes leading to many occluded areas. The across track DSM generated from 23 mono Collects acquired over 13 months provides a sharper and more complete reconstruction due to the smaller convergence angles. However some outliers are present due to scene or appearance changes between the stereo pairs.

A comparison of PlanetScope and SkySat Tri-Stereo DSMs over the city of Terrassa is shown in Fig. 7. Large buildings, for example in the industrial area in the center of the scene are reconstructed well. Larger roads and infrastructure such as bridges present. The denser inner city blocks in the northern-western part of the scene are less well defined and contours are blurry and some buildings are reconstructed incompletely. The smaller residential buildings in the south-eastern part are mostly missing, as well as single trees or small patches of high vegetation. The PlanetScope DSM shown in Fig. 7 (c) does not
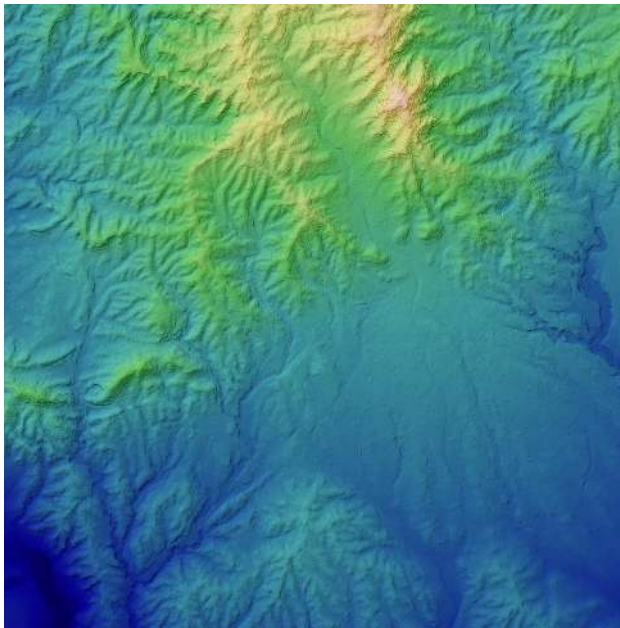
(a) PlanetScope DSM

Figure 5. DSM generated from 100 stereo pairs of PlanetScope imagery over the test areas in Catalonia.

(a) SkySat Video

(b) SkySat across track

(c) SkySat Tri-stereo

Figure 6. DSMs of Berlin, generated from different SkySat acquisitions

contain buildings, but shows the topography much clearer than lower resolution DSMs, such as the 30m SRTM-C Band data

displayed in Fig. 7 (d).

Statistics on the height differences of the generated DSMs and the LiDaR reference point cloud for the test areas over Catalonia are shown in Table 1. Given the small stereo convergence angles of the available PlanetScope image pairs, where a one pixel error in image orientation or stereo matching results in height errors of 10 to 30 m, the low offset of 0.2 m and RMSE of 5.5 m show that image orientation and sub-pixel accurate dense matching work reasonably well for stacks of typical PlanetScope imagery. As the normalized median deviation (NMAD) is similar to the RMSE, the height errors follow a normal distribution with very few outliers. The statistics for SkySat Tri-Stereo are best in the hilly and city area, and are worst in the steep mountainous area. In general RMSE is still quite high, and a significant height offset is present for all three areas. Contrary to the expected behaviour, relative accuracy (NMAD) is best in the complex city area, and worse in the hilly area. The LiDaR data was acquired in 2016, three years before the SkySat images. Additionally the first-pulse heights measured by LiDaR over the mostly forested areas seem to be significantly higher than the SkySat DSM heights, leading to a negative height difference. As thus, the height accuracy at flat, clear areas is higher than the overall SkyBox DSM error statistics reported in Table 1.

Satellite	RMSE	Median	NMAD
PlanetScope Across Track	5.5	0.2	5.2
tri-stereo 1. City	3.1	-0.8	1.3
tri-stereo 2. Hills	3.2	-0.8	2.0
tri-stereo 3. Mountain	5.8	-1.7	2.5

Table 1. Height difference statistics in m between LiDAR first return points and DSM for different AOIs in Catalonia. Note that differences are influenced by additional systematic differences between LiDaR and satellite DSMs in forest areas and due to a temporal difference of over 3 years.

4. CONCLUSIONS

This paper evaluated DSM generation from images of PlanetScope and different SkySat acquisition modes. For SkySat, the DSMs from video provide the highest quality, but cover only a small area. When many mono acquisitions are available, across-track matching performs well, but contains more outliers. The dedicated Tri-stereo acquisitions work well for areas with little occlusions, but some unmatched areas remain in dense city areas due to occlusions caused by the large convergence angle. Compared to "traditional" VHR Satellites, SkySat image orientation is more demanding, and dense ground control is required for highly accurate results.

Future work could include the development of improved bundle adjustment procedures utilizing the special SkySat focal plane layout with its three frame sensors to reduce the number of exterior adjustment parameters. This could reduce the need for dense ground control.

As PlanetScope imagery is acquired very frequently, dense matching becomes feasible, but is limited by the very short baselines and the lower image resolution. Nevertheless, even with small baselines of around 6-10 degrees, fusion of multiple stereo pairs leads to a height accuracy (NMAD) of 5.2 meters, not far from the GSD of the imagery. However, typical man made structures such as buildings and bridges are not recovered.

ACKNOWLEDGEMENTS

REFERENCES

Aati, S., Avouac, J.-P., 2020. Optimization of Optical Image Geometric Modeling, Application to Topography Extraction and Topographic Change Measurements Using PlanetScope and SkySat Imagery. *Remote Sensing*, 12(20). `https://www.mdpi.com/2072-4292/12/20/3418`.

Beyer, R. A., Alexandrov, O., McMichael, S., 2018. The Ames Stereo Pipeline: NASA's Open Source Software for Deriving and Processing Terrain Data. *Earth and Space Science*, 5(9), 537-548. `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018EA000409`.

Bhushan, S., Shean, D., Alexandrov, O., Henderson, S., 2021. Automated digital elevation model (DEM) generation from very-high-resolution Planet SkySat triplet stereo and video imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 173, 151-165. `https://www.sciencedirect.com/science/article/pii/S0924271620303506`.

d'Angelo, P., Reinartz, P., 2011. Semiglobal matching results on the ISPRS stereo matching benchmark. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XXXVIII-4/W19, 79–84. https://www.int-arch-photogramm-remote-sens-spatial-inf-sci.net/XXXVIII-4-W19/79/2011/.

d'Angelo, P., Reinartz, P., 2012. DSM based Orientation of large stereo Satellite Image Blocks. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XXXIX-B1, 209–214. `https://www.int-arch-photogramm-remote-sens-spatial-inf-sci.net/XXXIX-B1/209/2012/`.

d'Angelo, P., Máttyus, G., Reinartz, P., 2016. Skybox image and video product evaluation. *International Journal of Image and Data Fusion*, 7(1), 3-18.

Ghuffar, S., 2018. DEM Generation from Multi Satellite PlanetScope Imagery. *Remote Sensing*, 10(9).

Hirschmüller, H., 2008. Stereo Processing by Semiglobal Matching and Mutual Information. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(2), 328-341.

Institut Cartogràfic i Geològic de Catalunya, 2020. `https://www.icgc.cat/en/Downloads`.

Planet Imagery Product Specifications, 2021. `https://assets.planet.com/docs/Planet_Combined_Imagery_Product_Specs_letter_screen.pdf`.

Zabih, R., Woodfill, J., 1994. Non-parametric local transforms for computing visual correspondence. *European conference on computer vision*, Springer, 151–158.

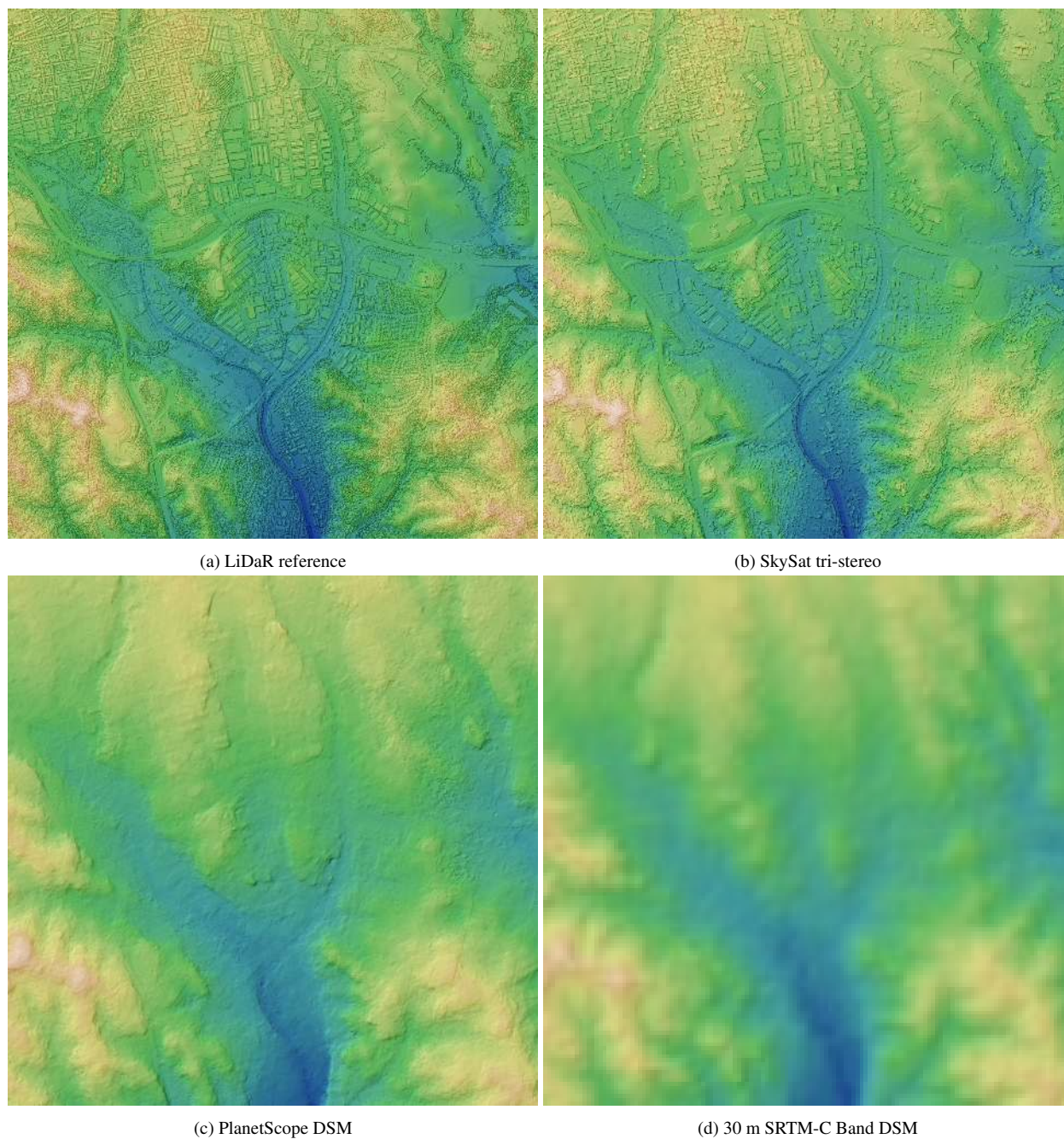(a) LiDaR reference

(b) SkySat tri-stereo

(c) PlanetScope DSM

(d) 30 m SRTM-C Band DSM

Figure 7. Different DSMs over the Terrassa area.