EVALUATION OF SKYBOX VIDEO AND STILL IMAGE PRODUCTS

Pablo d'Angelo,* Georg Kuschk, Peter Reinartz

German Aerospace Center (DLR), Remote Sensing Technology Institute D-82234 Wessling, Germany
email: Pablo.Angelo@dlr.de

Commission I, WG I/4

ABSTRACT:

The SkySat-1 satellite lauched by Skybox Imaging on November 21 in 2013 opens a new chapter in civilian earth observation as it is the first civilian satellite to image a target in high definition panchromatic video for up to 90 seconds. The small satellite with a mass of 100 kg carries a telescope with 3 frame sensors. 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 8 km in the panchromatic band, and multispectral images with 4 bands. Using super-resolution techniques, sub-meter accuracy is reached for the still imagery. The paper provides an overview of the satellite design and imaging products. The still imagery product consists of 3 stripes of frame images with a footprint of approximately 2.6 x 1.1 km. Using bundle block adjustment, the frames are registered, and their accuracy is evaluated. Image quality of the panchromatic, multispectral and pansharpened products are evaluated. The video product used in this evaluation consists of a 60 second gazing acquisition of Las Vegas. A DSM is generated by dense stereo matching. Multiple techniques such as pairwise matching or multi image matching are used and compared. As no ground truth height reference model is availble to the authors, comparisons on flat surface and compare differently matched DSMs are performed. Additionally, visual inspection of DSM and DSM profiles show a detailed reconstruction of small features and large skyscrapers.

1. INTRODUCTION

In the last years an increasing number of small optical remote sensing satellites are being deployed (Skybox Imaging, 2014, Planet Labs, 2014). The Skysat satellites of Skybox Imaging are a very interesting platform, as they provide new data with a resolution of 1 meter or better, and can acquire both mapping products and Full HD video sequences to 90 seconds in length. The space segment is simplied as much as possible, and tasks usually performed on board of the satellites are performed by the ground station software. This reduces complexity of the space segment and allows construction of smaller and less expensive satellites. Skybox Imaging plans a constellation of 24 satellites, orbiting in multiple sun synchronous orbits at various times of the day, providing multiple revisits per day at different times. In August 2014, Skybox has been acquired by Google for 500 Million USD. The next 13 satellites are currently being build by Space Systems/Loreal and will be launched in 2015 and 2016.

2. SKYBOX

The Skysat-1 satellite was launched on 21.11.2013 from Yasniy, Russia on a Dnepr rocket into a sun synchronous orbit with a height of 578 km. The similar Skysat-2 was launched on board of Soyuz-2/Fregat on 08.07.2014 from Baikonur. It reached a sun synchronous orbit with a height of 637 km.

The satellites have a size of 60x60x95 cm and weight of approximately 100 kg. They are equipped with a Ritchey-Chretien Cassegrain telescope with a focal length of 3.6 m, and a focal plane consisting of three 5.5 megapixel CMOS imaging detectors. Images are compressed with JPEG 2000 and then stored or downlinked to the ground station. 768 GB of on board storage are available and the data downlink rate is 450 MBit/s.

The spacecraft is three axis controlled though reaction wheels and TQ 15 torque rods, and uses 2 ST 15 star trackers (Dzamba, 2014). Skysat-1 & 2 have no active propulsion system, but further satellites of the constellation will include propulsion.
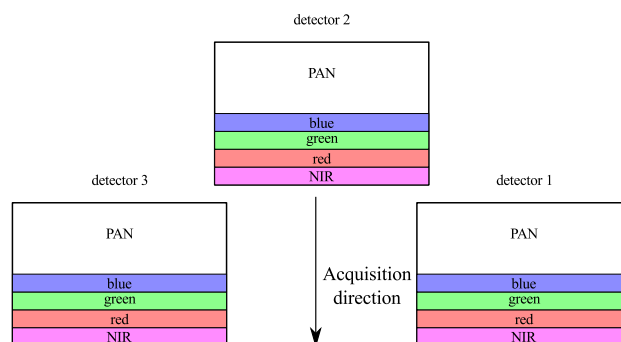
2.1 Focal plane layout

Figure 1: Skysat-1 focal plane, as projected to the ground.

Skysat-1&2 use 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. A schematic of the focal plane layout is shown in Fig. 1. The native resolution at nadir of the SkySat-1 and Skysat-2 is around 1.1 m. Further satellites will be placed in lower orbits, leading to increased image resolution.

The Raw Video and Frame products contains both a physical camera model and a RPC for each individual frame. The interior orientation is given by the location (X,Y,Z) and tilts the CMOS detector planes with respect to the projection center of the telescope. The unconventional interior orientation with 3D rotation of the focal plane with respect to the telescope requires extension

---
*Corresponding author.

of the ordinary frame camera geometry routines. Due to the time constraints while preparing the article, evaluation of the physical model was not performed.

2.2 Video product

For the video product, the panchromatic part of a single detector records a video with 30 frames per second while the spacecraft pointing follows the target. Video sequences up to 90 seconds in length can be recorded.

The video product can be delivered in different formats, a stabilized Full HD video in MP4 format, where all video frames have been co-registered, and an unstablized video without coregistration. The video size of both products is 1920x1080 pixels. A raw video product with individual TIFF files with 11 bit of radiometric resolution and per frame orbit and attitude parameters and RPCs is also available. The raw video frames are available at the full panchromatic detector area size of 2560x1080 pixels.

2.3 Frame product

In addition to the video product, larger areas can be covered by strips with a swath width of 8 km. These are acquired in a "push-frame" mode, where all three detectors acquire a highly overlapping video sequence, for example at 40 Hz (Smiley et al., 2014). All pan and multi-spectral images overlapping with a single panchromatic "master" frames are co-registered and fused using a super-resolution algorithm (Robinson et al., 2014). During the fusion, a superresolution process is used to increase the resolution from 1.1 m to 90 cm. Panchromatic, multispectral and several variants of pansharpened images are delivered. Fig. 2 shows the effect of the multi image fusion algorithm on a plane.
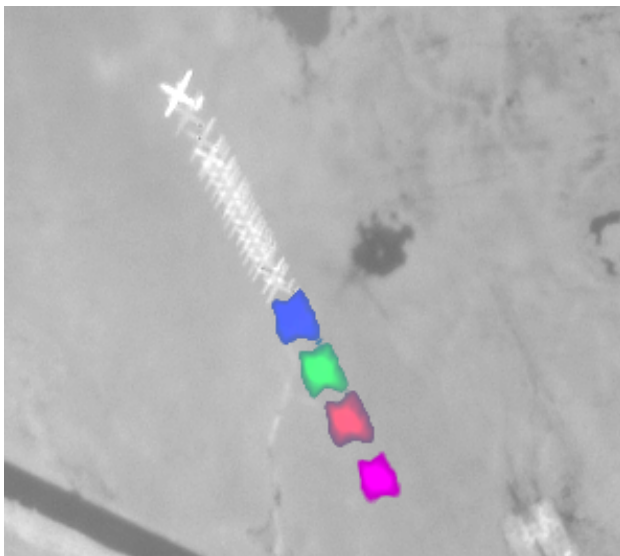

Figure 2: Effect of image fusion on a moving object. The purple/near infra-red, red, green and blue blobs show the plane as imaged by a handful of multi spectral frames, and the approximately 18 gray planes illustrate how many panchromatic images were fused.

The master images are chosen to have some overlap in the along track direction, and there is a small across-track overlap between detector 2 and detectors 1 and 3, see Fig. 3 for an overview.

As handling and mosaicing of the individual frames is not a straightforward operation for most imagery customers, Skybox will offer an mosaiced Geo product in the future. However, it was not available at the time of writing.
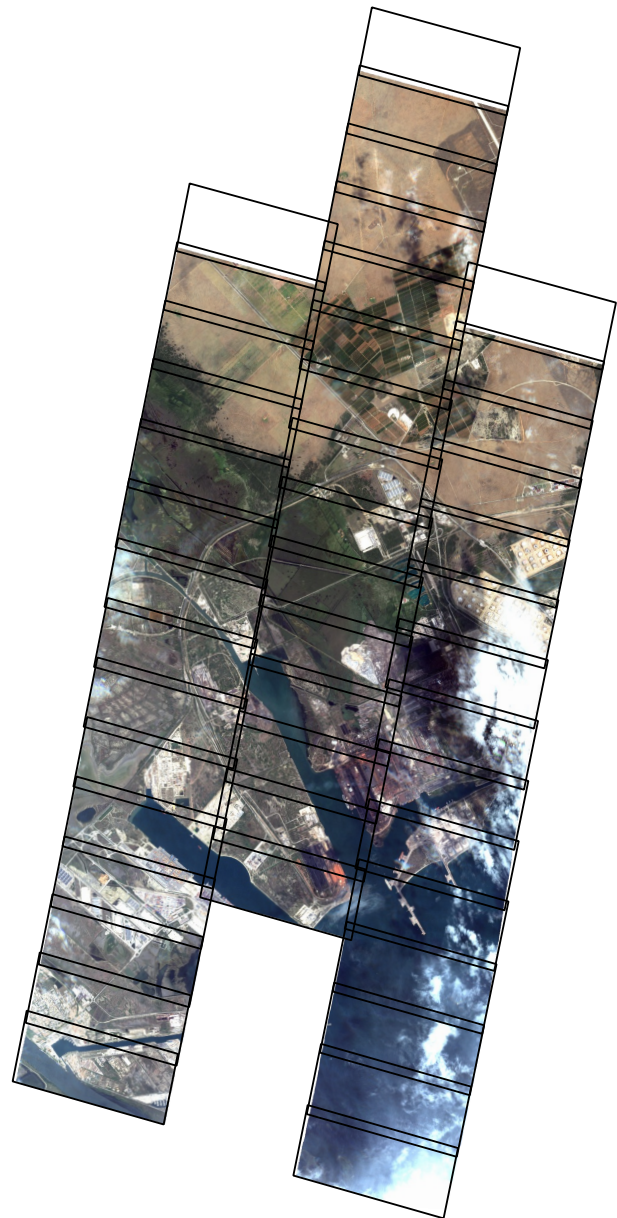

Figure 3: Fos sur Mer industrial zone as seen by Skysat-1. The bounding boxes show the individual frames after coregistration and multi-image fusion.

3. EVALUATION

A frame product covering the harbor and industrial zone of Fos sur Mer in France, cf. Fig 3 and a video product of Las Vegas[1] and the Mirny diamond mine in Russia were available for evaluation. The Fos sur Mer image was acquired on 8th August 2014.

3.1 Image orientation

To keep costs and weights down, Skysat was not designed to offer the best direct georeferencing performance. It is nevertheless interesting to evaluate the direct georeferencing performance. For the Fos sur Mer, GCPs were measured using OpenStreetMap (OpenStreetMap, 2014) vector data and SRTM as reference data. Unfortunately, only a few gcps could be identified, leading to a relatively weak coverage of the area.

---

[1] see the stabilized video product at http://vimeo.com/92072374

| Image | No. CPs | mean (m) | | std (m) | | RMSE |
		X	Y	X	Y	(m)
d1_0006	5	-149.1	-59.3	1.1	3.4	160.5
d1_0009	4	-215.8	-58.8	1.0	2.6	223.7
d2_0008	4	-202.9	-72.5	0.7	1.1	215.4
d2_0015	5	-101.7	-42.6	1.3	2.2	110.3
d3_0014	1	-93.6	-20.3	0.0	0.0	95.7
d3_0015	8	-88.2	-56.3	1.9	2.8	104.7

Table 1: Planimetric checkpoint errors against Open-StreetMap/SRTM based checkpoints for 6 images before adjustment.

Table 1 shows that each image has a significant, but different bias, but the standard deviations are low, indicating systematic offsets in the orbit and attitude parameters.

Thus, in addition to direct georeferencing of single frames, an RPC bundle adjustment was performed, with and without GCPs. The initial relative bundle adjustment used tiepoints matched using SIFT and further refined using local least squares matching. As the baseline between adjacent images is very small, rays between the tie points are almost co-linear. We softly constrain the tie point Z coordinates to the SRTM heights (d'Angelo, 2013) to avoid singularities during adjustment. Note that this is this potentially already adds a weak height reference to the adjustment. Tie point RMSE after adjustment were 0.15 pixels when using simple image space row/column shift correction. Checkpoint errors are reported in table 2

| Image | No. CPs | mean (m) | | std (m) | | RMSE |
		X	Y	X	Y	(m)
d1, 6	5	-109.0	-12.5	1.1	3.4	109.7
d1, 9	4	-102.9	-40.0	1.0	2.6	110.5
d2, 8	4	-113.3	-14.7	0.8	1.1	114.3
d2, 15	5	-94.1	-21.3	1.3	2.2	96.4
d3, 14	1	-90.6	-43.5	0.0	0.0	100.6
d3, 15	8	-85.1	-45.0	1.9	2.8	96.3

Table 2: Planimetric GCP errors after relative RPC bundle block adjustment.

After relative adjustment, the absolute mean X,Y error is reduced down to around 100 meters for the evaluated images. The X,Y standard deviations did not change significantly after the adjustment, and are still around 2..4 meters.

Finally, GCPs were included in the adjustment, due to the low number, unfortunately, no independent checkpoints could be used, cf. table 3. The mean and RMSE errors are still quite high, but could be caused by the with the accuracy of the OpenStreetMap data, which in this areas relies on consumer quality GPS tracks and BING maps imagery.

| Image | No. GCPs | mean (m) | | std (m) | | RMSE |
		X	Y	X	Y	(m)
1, 6	5	-6.8	9.8	1.1	3.4	12.4
1, 9	4	-3.8	-16.4	1.0	2.6	17.0
2, 8	4	-10.9	6.7	0.8	1.1	12.9
2, 15	5	5.6	13.4	1.3	2.2	14.8
3, 14	1	2.9	-8.1	0.0	0.0	8.4
3, 15	8	7.4	-9.1	1.9	2.8	12.2

Table 3: Planimetric GCP errors after adjustment.

3.2 Radiometry

Except for the video product, Skybox imagery has been processed with image fusion and super-resolution algorithms. In the Fos Sur Mer product, multispectral channel data effectively uses 10 bit, where as the panchromatic imagery uses 11 bits, cf. Fig. 4. Very few pixels (less than 0.01 %) reach higher DN, but these are likely outliers created during the image fusion process.
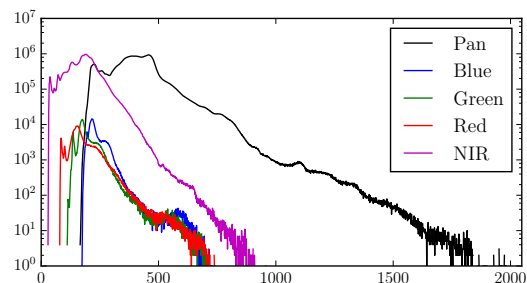

Figure 4: Histograms of panchromatic and multispectral channels of all images in the Fos sur Mer scene.

Image noise has been estimated using a method similar to (Crespi and De Vendictis, 2009). The standard deviation of all possible 3x3 windows for all Fos sur Mer images have been computed, and the 2 % quantile is used as an estimate for of the standard deviation σ, based on the assumption that 2 % of the scene do contain locally uniform DN values. The overall signal to noise ratio is computed by $SNR = (DN_{max} - DN_{min})/\sigma$. DN_{min} and DN_{max} are the 0.5% and 99.5% quantile of the image digital numbers. The results are shown in table 4. It is obvious that the video is noisier than the still images, both in terms of σ and SNR. This is expected, as the Still image is a fusion of more than 15 frames. Multi-spectral and Panchromatic images show similar noise, when considering the different bit depth of the images.

Image	DN_{min}	DN_{max}	σ	SNR
Video Pan	361	853	4.3	114
Still Pan	205	802	1.8	338
Blue	189	351	0.5	324
Green	126	281	0.5	310
Red	82	241	0.4	360
NIR	36	392	0.4	807

Table 4: Overall image noise of Skysat-1 video, panchromatic still and multispectral still images.

In order to check if the noise is dependent on the DN values, σ was estimated for individual bins. Due to the non-uniform histogram, the bins where choosen with a size of 64 DN, except for the lowest bin, and values above 1024, where a bin size of 256 DN was used. Figure 5 shows a linear dependency of the the noise and DN value, for the pansharpened still image product provided by Skybox.

Due to time constraints, no MTF analysis was performed during this study. Visually, the products show a high amount of detail and the noise level is reasonably low. The dynamic range of the video product is lower than the still imagery product, for example when looking at areas shadowed by skyscrapers. No information can be extracted for those from the video product, but the still image products show a bit more information in shadow areas, although with very low contrast. Areas with a direct reflection of the sun into the sensor, for examply by highly reflecting roofs result in a small, overexposed circle in the images, this is an im-
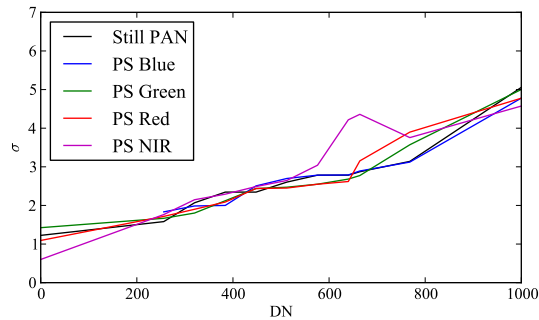
Figure 5: Noise level versus DN of panchromatic and pansharpened image channels. The plot shows a linear relationship between noise and DN values.

provment over the extended "spill" areas seen in images from some other sensors.

3.3 DSM generation

One possible application for the video product is DSM generation. With traditional satellite imaging, typically, only a stereo pair or triplet is available. We have processed the raw video data of Las Vegas and the Mirny mine sequences into DSMs, each consisting of 1800 images. For DSM generation, we have temporally subsampled the sequence at 1 Hz, leaving 60 images, as adjacent frames have a very small baseline and cannot be used for matching. In the future, the high redundancy could be used for additional noise reduction and super-resolution matching. A relative block adjustment has been performed on the images, resulting in a tie point RMSE of 0.1 pixels.

Different image matching strategies are possible when a multi-view sequence is available. The most straightforward way would be to match images pairwise and then later fuse the results. This works very well for triples acquired with other VHR satellite sensors. Another strategy is to perform multi-image matching by computing an average data term/correlation score of multiple slave images to a single master image, followed by regularisation with SGM (d'Angelo and Reinartz, 2011) or total variation (Kuschk, 2013). For sequences with a limited number of frames, for example with south, nadir and north looking images, this did not work very well in urban areas as occlusions cannot be considered during computation of the the average correlation, leading to a bad data terms for occluded regions. With the high number of frames in the skybox data, many frames from similar perspectives are available, thus reducing the likelyhood for averaging of occlusions during multi-view correlation computation.

Different image matching strategies were used for the images. ADCensus with a small adaptive support windows was used as the data term for all methods. Except for method 2, total variation was used to compute a regularized solution. Fig. 6 shows DSMs generated with the following methods:

1. Pairwise matching of three images and averaging of the resulting DSMs. similar to triplets acquired by other satellite sensors, such as Pleiades or WorldView-2. This DSM is the baseline and can show the advantage of using a larger number of frames.

2. Matching of one master image against the 20 closest images. Here the data term is computed as average of the AD-Census score of the master with respect to every slave image. To

evaluate whether the higher redundancy allows the derivation of DSMs without a complex regularisation algorithm, only a simple winner takes it all approach was used.

3. Method 2, but with total variation regularization.

4. Applying method 3 to 20 master images and averaging the resulting DSMs.

Unfortunately, no reference DSM of Las Vegas was available for comparison. By visual evaluation of the results shown in Fig. 6, it can be seen that the general structure and the buildings have been recovered with all methods. The winner takes it all method (2) is still much noisier than the triplet (1), so regularization should still be used, even when multiple images are available, and data redundancy alone is not enough for good reconstruction, particularly in shadow areas. A profile though the MGM Grand hotel shows the difference in noise, cf. Fig. 7.
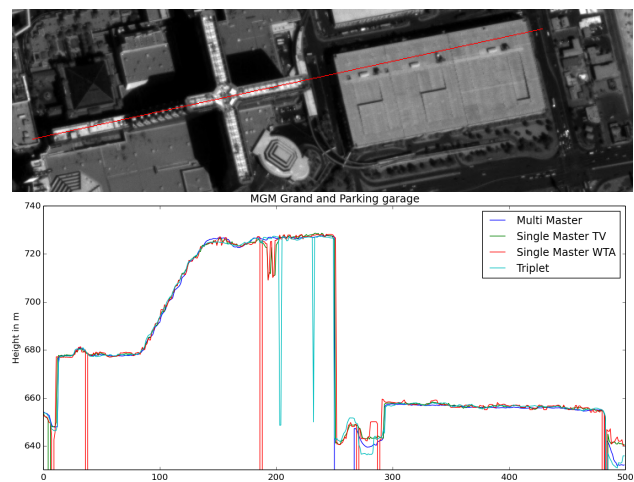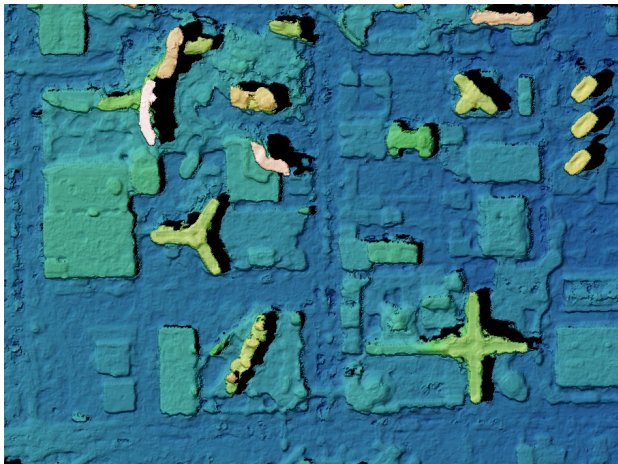

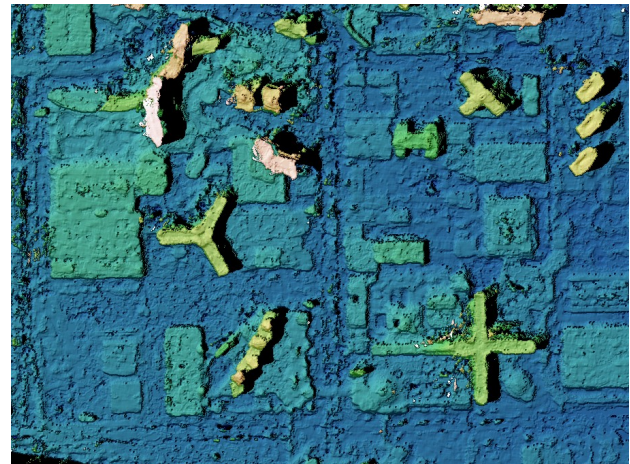Figure 7: Profile though the MGM Grand Hotel

3.4 Conclusion

With the first civil VHR video products, the Skysat satellites offer very interesting possibilities for future applications. The "push-frame" architecture and the super-resolution approach reduce the complexity of the Skysat satellites and will allow launch of a constellation with multiple daily visits. A drawback of the constellation is the comparably small footprint of the still and video products, Skybox is thus primarily suited for monitoring applications and not for the mapping of large areas.

Considering the small and relatively low cost satellites, images quality is good, with some slightly higher noise level in the videos. Direct georeferencing accuracy is low and it was around 100 m after relative adjustment of still image collection. Reference images or GCPs are thus required for accurate orthorectification. Due to missing reference data of higher quality the absolute georeferencing performance when using GCPs could not be evaluated properly in this paper. Future studies with better reference data will shed more light onto this.
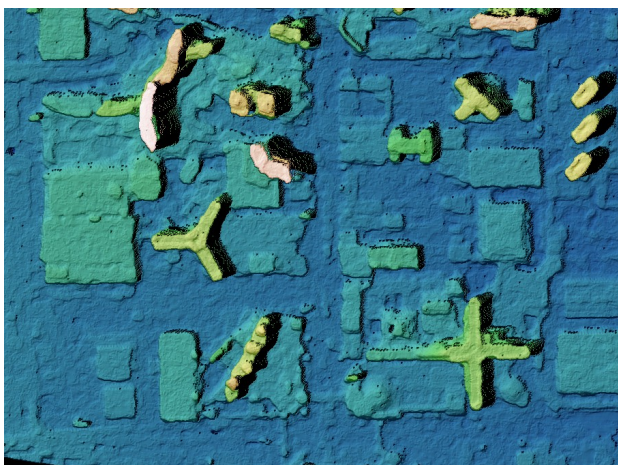
DSMs were generated from Skybox video sequences. Compared to triplets, multi image matching increases the height accuracy. Further research needs to be done to effectively utilize the highly redundant information, and exploit the video frame rate for image denoising and super-resolution matching. The video obviously has many other use cases, such as analysis of dynamic processes, object detection and tracking.
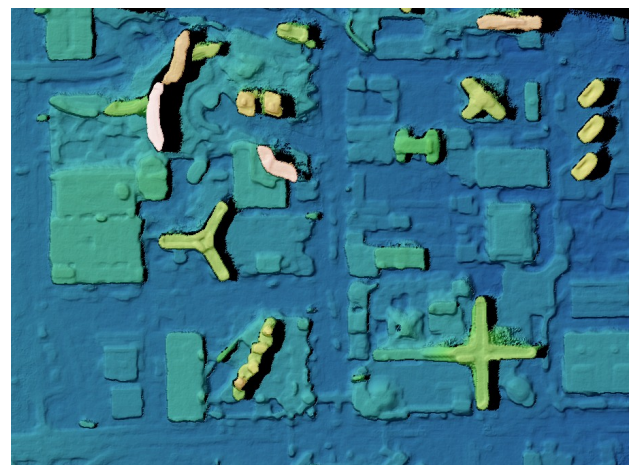
(a) Pairwise matching of three images

(b) 20 images against one master image, without regularization

(c) 20 images against one master image, with regularization

(d) 15 master images, each matched against 20 slave images.

Figure 6: Cutout of the Las Vegas DSMs computed with different matching methods.

REFERENCES

Crespi, M. and De Vendictis, L., 2009. A procedure for high resolution satellite imagery quality assessment. Sensors 9(5), pp. 3289–3313.

d'Angelo, P., 2013. Automatic orientation of large multitemporal satellite image blocks. In: Proceedings of International Symposium on Satellite Mapping Technology and Application.

d'Angelo, P. and Reinartz, P., 2011. Semiglobal matching results on the isprs stereo matching benchmark. Proceedings IS-PRS Hannover Workshop 2011: High Resolution Earth Imaging for Geospatial Information.

Dzamba, T. e. a., 2014. Success by 1000 improvements: Flight qualification of the st-16 star tracker. In: 28th Annual AIAA/USU Conference on Small Satellites.

Kuschk, G., 2013. Model-free dense stereo reconstruction creating realistic 3d city models. In: Proceedings JURSE, pp. 202–205.

OpenStreetMap, 2014. http://www.openstreetmap.org.

Planet Labs, 2014. http://www.planet.com.

Robinson, D., Berkenstock, D., Mann, J., Dyer, J. and Trela, M., 2014. High-resolution imagery and video from skysat-1. In: JA-CIE 2014.

Skybox Imaging, 2014. http://www.skybox.com.

Smiley, B., Levine, J. and Chau, A., 2014. On-orbit calibration activities and image quality of skysat-1. In: JACIE 2014.