Skybox Image and Video Product Evaluation

Skybox Imaging launched the SkySat 1 & 2 satellites in 2013 and 2014. They are small and low cost satellites, capable of recording still images with sub-meter resolution and high definition video for up to 90 seconds of a single target, using a novel focal plane with three 5.5 mega pixel frame sensors. To improve resolution and signal-to-noise ratio of the still imagery, multiple frames are fused with superresolution image merging. This article evaluates the SkySat image quality and georeferencing performance. For comparison, Worldview-3 data acquired over the same AOI is evaluated using the same procedure and reference data. Results show that georeferencing of SkySat still imagery products requires more ground control than other VHR sensors. Results from the Las Vegas video show that speed of moving vehicles in videos can be detected to better than 0.8 mph and that automatic motion detection can be used to georeference videos to street vector databases. Generation of digital surface models from video show a height accuracy of 1.2 meters (NMAD).

Introduction

In the last years an increasing number of small optical remote sensing satellites are being deployed (Skybox Imaging, 2015; Planet Labs, 2015) 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 of up to 90 seconds in length. The space segment is simplified as much as possible, and tasks usually executed 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 has 2 satellites in orbit and 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 (Robinson, et al., 2014). In August 2014, Skybox has been acquired by Google for 500 Million USD. The next 13 satellites are currently being built by Space Systems/Loreal and will be launched in 2015 and 2016 (Skybox Imaging, 2015).

SkySat Satellites

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. 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 Complementary metal–oxide–semiconductor (CMOS) imaging detectors with a pixel size of of 6.5 µm (Robinson, et al., 2014). 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 torque rods, and uses 2 small star trackers (Dzamba, 2014). SkySat-1 & 2 are not equipped with active propulsion systems, but further satellites of the constellation will include propulsion and improved reaction wheels.

The upper half of each 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 Figure 1 a). By using a "push-frame" imaging mode, a ground pixel is imaged by up to twenty images. A kind of "virtual" time delayed integration (TDI) is performed by image registration and super resolution image fusion. While classical TDI requires high precision attitude control, to avoid "smear" across adjacent TDI columns, the push frame can tolerate larger attitude deviations, as the image processing will co-register the images on ground and compensate the attitude drift (Murthy, et al., 2014). The native resolution at nadir of the SkySat-1 and SkySat-2 is 1.1 m. Further satellites will be placed in lower orbits, leading to increased image resolution.

Sensor model

The SkySat interior model describes the location and orientation of each detector in the optical coordinate system, centered on the optical center of the telescope (Smiley, et al., 2014).

The SkySat satellites use an unconventional interior orientation with 3D rotation of the focal plane with respect to the telescope requires extension of the ordinary frame

camera geometry routines. The location x'_{mn} of a pixel in the optical coordinate system is given by Eq. 1

$$x'_{mn} = R_{\theta_{dz}} R_{\theta_{dy}} R_{\theta_{dx}} \begin{pmatrix} (m - m_{dc})p_{dx} \\ (n_{dc} - n)p_{dy} \\ 0 \end{pmatrix} + o_d \qquad (1)$$

Where d stands for the detector number, R for the rotation matrices around z,y,x axis respectively. m_{dc} and n_{dc} are the center pixel of detector d, p_{dx}, p_{dy} is the size of a pixel in x and y. $o_d = (o_x, o_y, o_z - f)$ is the position of the center pixel in the optical coordinate system. Additionally, radial and tangential coefficients are present in the metadata files, but are currently set to 0. The exterior orientation is given as latitude, longitude, height in WGS84 coordinates and omega, phi, kappa angles.

Additionally, rational polynomial coefficients (RPC) derived from the physical model are provided for every image.

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 unstabilized video without co-registration. 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.

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 super-resolution process is used to increase the resolution from 1.1 m to 90 cm. Panchromatic, multispectral and several

variants of pansharpened images are delivered. Figure 1 b) shows the effect of the multi-image fusion algorithm on a plane.

The master images are chosen to have some overlap in along-track direction, and there is a small across-track overlap between detector 2 and detectors 1 and 3, cf. Figure 2 a).

As handling of the individual frames is not a straightforward operation for most imagery customers, Skybox offers a Geo product, where multiple images of a single detector are projected onto a plane with constant height. Each Geo image is accompanied by a corresponding RPC file. The product level is thus similar to the DigitalGlobe Ortho Ready Product, except that a typical collection still consists of 6 images. As the fused images contains content capture from different viewing directions, a single frame or push-broom model is not sufficient to model the acquisition geometry, and there are discontinuities in the ground to image projection function on the image boundaries of the used frames. These discontinuities cannot be modelled well by the 3rd order polynomials used in the RPCs shipped with the product.

Methodology

Image Orientation

Image orientation and orthorectification is the basic prerequisite for many applications of satellite data. Compared to traditional push-broom satellites, the image frames captured by the SkySat satellites are relatively small, around 2.5 by 1 km. Thus many images are needed for covering a typical target. The precise co-registration and mosaicking of these images is thus a prerequisite for further use of the data. In this work, the RPC sensor model is used either directly or after block adjustment (Grodecki & Dial, 2003).

Radiometric Quality

Evaluation of the radiometric properties is best performed using specially prepared and well measured calibration sites. However, special calibration acquisitions are usually not available. Often, methods based on detecting homogeneous areas or edge features are used for this purpose (d'Angelo, et al., 2014; Aguilar, et al., 2014). However these require images containing natural, homogenous regions. For the different SkySat datasets evaluated in this study, these methods produced quite different

results, depending on the dataset, as some scenes such as the Chicago collection contain very few completely homogenous areas.

Thus a new method was developed for this study, exploiting the fact that the same area is acquired multiple times in both video and the still image datasets. Overlapping areas are co-registered using a local least squares matching and one image is resampled to fit the other using affine transformation and bilinear interpolation. Then a difference image can be computed to remove the systematic scene content.

Under the assumption of equally distributed Gaussian noise in both images, the standard deviation of the noise is given as: $\sigma = \sigma_d/\sqrt{2}$, where σ_d is the standard deviation of the difference image.

For the video product, this can be done using adjacent frames with a time difference of 1/30th of a second. The still image noise can be estimated using the overlapping areas between detector 1 and 2 or 2 and 3. Here a larger time difference of 3 seconds is present between the observations, where scene structure due to the change in viewing directions is visible in the difference images. Thus only flat areas should be evaluated with this method.

Motion analysis

The recorded video enables new ways of information extraction. The high spatial overlap, acquisition frequency and good resolution make also the small moving objects visible, e.g. the cars on the street. As the original stabilized SkySat videos are in sensor geometry, we bundle adjusted the video to improve the relative orientation and then orthorectified each frame to a plane of constant height. This co-registration process allows motion tracking in georeferenced coordinates. A car is around 4-5 meters long, resulting in a pixel size of ~4 pixels. This size is not enough to detect these objects from a single frame reliably, but the movement in the video makes the objects visible and detectable.

To extract the moving pixels we apply the Gaussian mixture model background subtraction described in (Zivkovic & van der Heijden, 2006) on the co-registered video frames. This gives a pixelwise foreground mask. The blobs in the foreground are extracted as objects in each frame. These objects are assigned along the consecutive frames by Kalman-filtering similar as in (Mattyus, et al., 2010). The match between the objects is calculated as the sum of the distance between the predicted and the current

(i.e. detected) location; and the correlation of the visual object appearance. This method is not robust enough to track a specific car for a longer time, but it can estimate the flow of traffic in certain locations.

Automatic vector based georeferencing of video

The acquired object tracks in the video can be considered as pieces of the street network, assuming that the vehicles drive on the road. Thus it can also be seen as an incomplete road detector.

Here we present a proof of concept how the pieces of the road network can be applied to match the image to a road database and thus improve the geolocalization of the video. We use the geometric hashing (Wolfson & Rigoutsos, 1997) presented in the paper (Mattyus & Friedrich, n.d.). This hashing method is invariant to rotation and translation, thus it can find the correct location of the image even over a larger search area (i.e. large error in the geolocalization), and thus correct even large errors due to attitude disturbances automatically, if enough road segments could be detected by the motion tracker.

Evaluation

Dataset description

The evaluation was performed using a scene of Frame products and a Geo product of Chicago, and Video products of Las Vegas and the Mirny mine.

The Chicago acquisition is shown in Figure 2. Both a Frame and a Geo product of the scene were available and were evaluated. This allows the direct comparison of both product types. The scene was acquired by SkySat 2 on 28th of January 2015, with an azimuth of 90 degrees and 60 degrees elevation. Solar elevation was only 22 degrees and the scene is heavily covered by snow. As reference for the geometrical elevation, National Agriculture Imagery Program (NAIP) ortho imagery with a resolution of 1 m, acquired on 13th of June 2014, and National Elevation Dataset (NED) digital terrain model (DTM) with a resolution of 1/9 arc seconds (around 3 m grid size) were used as reference for ground control points (GCP) and check point extraction. The reference data was downloaded from USGS (USGS 2015), and reprojected into UTM before it was used. Unfortunately we did not have access to SkySat data over sites where we have access to DGPS points.

Additionally, a WorldView-3 (WV-3) collection acquired on 5th of January 2015 was available for comparison. The SkySat AOI is covered by two adjacent WV-3 Level 2 Ortho Ready strips acquired in the same orbit with a resolution of 0.33 m for the panchromatic images and 1.32 m for the multispectral images. The sun elevation was 24° and the off nadir angles were 16.2° and 20.5°.

Due to the difference in snow cover and general appearance, automatic image matching between the winter SkySat images and the summer NAIP ortho failed. Thus, GCPs were determined by manual image matching. Crossing sidewalks were one of the few features that could be localized in both images, but their measurement is laborious and 97 points were measured. 25183 tie points were matched using SIFT (Lowe, 2004) and refined using local least squares matching (Ackermann, 1984), similar to the procedure described in (d'Angelo, 2013). See Figure 2 for an overview of the SkySat dataset and details of the input images.

Direct Georeferencing

To keep costs and weights down, SkySat satellites were not designed to offer the best direct georeferencing performance. It is nevertheless interesting to evaluate the direct georeferencing performance. Two different tests were used, first images were checkpoints could be measured were evaluated. We compute the planimetric error by projecting the GCPs image coordinate onto a plane defined by the GCPs height, and compute $RMSE_{XY}$. The overall RMSE using 97 checkpoints is 199.4 m, the best frame has an RMSE of 49.7 m and the RMSE of the worst frame is 285.6 m. When performing a relative bundle block adjustment, using affine RPC correction parameters, overall checkpoint RMSE decreased slightly to 186 m, and a good tie point RMSE of 0.09 pixels was achieved, indicating a good relative fit. When only bias RPC correction was used, overall RMSE increased to 211.4 m, slightly worse than the results without adjustment. Tie point residuals are 0.27 pixels, but systematic patterns, reaching up to 1 pixel in the corners of images captured by detector 2 are present. This is a first indication that a simple bias RPC correction is not sufficient to align SkySat imagery on the sub-pixel level.

Evaluation with GCPs

Evaluations with different spatial GCP distributions were performed, using bias and

affine RPC corrections in image space. The results are reported in Table 1, and Figure 3.

An interesting question is whether bias or affine RPC correction is required, as this determines the number and distribution of ground control point used during the processing. For practical applications, sensors that require only bias RPC correction are favourable, as only a single or very few GCPs are required even when adjusting a larger block, such as SkySat still imagery. The results for bias correction show a very high absolute error of over 40 meters, and high systematic tie point residuals; cf. Figure 3 a). One possible source for these residuals might be a bias in the satellite yaw angle measurement. Bias RPC correction is thus not applicable to still imagery blocks.

Affine RPC correction shows better results; with a low tie point RMSE of 0.1 pixels, systematic errors in the image corners reach up to 0.4 pixels. Absolute error is still quite high at 5 m, even when multiple GCPs in every fourth image are used; cf. Figure 3 b). This plot shows that the largest and systematic errors are located in frames far from GCPs. This either indicates the need for more control points, or an over parametrisation of the adjustment. As mentioned earlier, GCP collection from the reference imagery was not straightforward, and the accuracy of the NAIP data does allow evaluation of the data on the pixel or sub-pixel level. Despite this shortcoming of the available reference data, it can be concluded that with standard, RPC based processing, a large amount of ground control is needed to reach pixel or sub-pixel level.

Geo Image Evaluation

A similar evaluation was performed with the SkySat GEO image product, which consists of stitched versions of the frame product. This is an attempt for simplifying the handling of SkySat data, by providing larger images, projected onto a plane with constant elevation. RPCs are provided for each mosaic. The resulting mosaicked image is neither of frame nor pushbroom geometry, its colinearity equations are discontinuous at the seams between the frames, and are not modelled well by the third order RPC terms. According to Skybox, this product is meant for easier ingestion into GIS software, and not for precise orthorectification, or other applications requiring precise geometry. Nevertheless, it is an interesting product for many users, as it strongly reduces the number of image frames that need to be handled, and users will try to improve the georeferencing by orthorectification.

For the Chicago dataset, every 12 consecutive images of each detector were merged; the GEO product thus consists of 6 images, while the frame product consists of 63 images. The GCPs and checkpoints used in the single frame images were transferred to the geo images using local least squares matching and a similar geometric accuracy evaluation was performed. The results reported in Table 1 show that the SkySat Geo product is, at least currently, not suitable for precise orthorectification, with errors of up to 86 m. These are far larger than the accuracy of the used GCPs. For orthorectification, the image frames product, adjusted using many GCPs is thus the only option.

Comparison with WorldView-3 Data

To check if these errors are due to the GCP measurement and the reference data used, the WV-3 dataset was processed using the same reference data. Due to different snow cover conditions, GCPs could not be transferred from the SkySat to the WV-3 images. Thus 16 different points were manually measured the reference Image and the WV-3 MS images. As we evaluate only two images instead of 63 in the SkySat case, the reduced number of points still allows a valid comparison. A checkpoint RMSE of 1.9 m was archived when adjusting the multispectral WV-3 images using 2 GCPs, one for each image covering the AOI. A RMSE of 1.9 m is still above the 1.2 m resolution of the WV-3 multispectral data. Earlier experience shows that with better reference data, errors can be reduced to less than a pixel (Aguilar, et al., 2014). As no systematic pattern was visible in the WV-3 checkpoint residuals, we assume that the difficulties encountered in manual GCP selection and the accuracy of the NAIP imagery and NED DEM are the limiting factors.

The best result we got from the SkySat imagery was 5 m RMSE, which is still significantly lower than the 1.9 m accuracy of the WV-3 data. As the other limiting factors, such as difficulty in GCP measurement and reference data were the same for both evaluations, the errors can be attributed to the SkySat image geometry. Considering that reference data might not be available for the complete AOI, or automatic processing might not always be feasible due to strong changes in appearance between different seasons, the SkySat image frames product requires more, possibly manual work than comparable traditional satellite data, but still yields lower accuracy results. In this case, 2 GCPs were sufficient for WV-3, while 51 GCPs were needed for the SkySat processing. Ideally, the SkySat imagery should be processed with multiple

GCPs in every image, but that is not always possible due to difficulties in GCP acquisition. These results were obtained using standard processing with RPCs, but for SkySat frame imagery, the use of the physical model and adjustment of relatively instable satellite attitude angles might provide better results, but requires custom engineered adjustment software, due to the special interior orientation parameterisation with tilted focal planes.

Table 1: Bundle block adjustment results.

Product GCP Layout # GCP / # CP	Bias RPC correction			Affine RPC correction		
	GCP RMSE / MAX	CP RMSE / MAX	Tie RMSE / MAX	GCP RMSE / MAX	CP RMSE / MAX	Tie RMSE / MAX
No GCPs 0/97	-	211.4 m/ 318.3 m	0.27 px / 1.00 px	-	186.0 m / 294.2 m	0.09 px / 0.49 px
Corners 18/79	42.3 m / 71 m	40.2 m / 65.3 m	2.29 px / 5.88 px	1.4 m / 2.54 m	12.1 m / 21.9 m	0.10 px / 0.41 px
Corner/Center 23/74	44.9 m / 75.9 m	42.8 m / 65.8 m	2.13 px / 5.78 px	1.4 m / 2.6 m	8.4 m / 18.9 m	0.10 px / 0.44 px
Grid 51/46	49.6 m / 94.6 m	48.2 m / 77.6 m	1.76 px / 4.51 px	1.8 m / 4.4 m	5.0 m / 10.4 m	0.12 px/ 0.52 px
GCPs 97 / 0	49.8 m / 99.8 m	-	1.68 px / 4.67 px	2.0 m/ 4.8 m	-	0.12 px/ 0.54 px
SkySat Geo 38/59	91.0 m / 353.9 m	68.64 m / 213.6 m	2.49 px / 21.07 px	26.9 m / 74.5 m	25.2 m / 86.0 m	1.61 px / 15.11 px
WV-3 MS 2/14	1.4 m / 2.0 m	1.9 m / 3.5 m	0.14 / 0.36	-	-	-

Radiometry

Except for the video product, SkySat imagery has been processed with image fusion and super-resolution algorithms to improve resolution and radiometric quality.

In the Chicago product, multispectral channel data effectively uses 10 bit, and the panchromatic imagery uses 11 bits, cf. Figure 4. Less than 0.01 % pixels reach higher DN, but these are likely outliers created during the image fusion process.

Figure 5 shows the area and the region used for image noise analysis. The difference image mostly contains the noise signal. Systematic strips in flight direction are visible in the multi-spectral and pan-sharpened images, cf. Figure 5 c), but not in the panchromatic images. Additionally, two consecutive frames from the Mirny mine video acquisition were also analysed using the same process.

Table 2 reports the noise of video, panchromatic, pan-sharpened and multispectral SkySat images. When comparing the signal to noise ratio, computed as ratio of mean DN and standard deviation, the multi-spectral images shows the lowest noise, and the video acquisition image the highest, as it is a single frame, without image fusion. The numbers show that the image fusion technique improve signal to noise ratio from 100 for the single, raw video frame to 267 for the panchromatic fused still image. A second noise estimation on the video frame was performed, where the image noise was computed using the whole difference image, and evaluated for 20 DN value bins, cf. Figure 6. It can be seen that the noise increases with increasing DN values.

Table 2: Result of image noise evaluation.

	Mean (DN)	Σ (DN)	SNR
PAN (video)	657	6.5	101
PAN	1096	4.1	267
MS Blue	664	1.9	356
MS Red	509	1.4	364
MS Green	476	1.4	350
MS NIR	406	1.3	323
PS Blue	1458	6.7	216
PS Red	1118	5.2	214
PS Green	1046	4.9	211
PS NIR	892	4.4	202

A similar evaluation was attempted on the WV-3 images, but due to the large viewing angle differences of the two collections, systematic effects due to BRDF

behaviour and moving clouds dominated the difference images, preventing a reliable quantitative evaluation of the WV-3 data. Thus, only a visual comparison of the images is performed. Figure 7 shows downtown Chicago with many skyscrapers as seen on the SkySat 2 pansharpend images and WV-3 multispectral images. The low sun elevation leads to deep shadows, and shows the performance of both satellites in a challenging scenario. Both images have been processed using histogram equalization, to enhance the shadow detail. Structures in the shadows are visible in both images, but the WV-3 image shows more details and does not contain color shadows, as seen in the SkySat images.

Vehicle speed estimation from video

For the Las Vegas video sequence, vehicles on the Las Vegas Freeway were tracked manually every 2 seconds, and their speed was estimated by using the distance driven between two images. Figure 8 shows the speed of the tracked cars. For vehicles cruising with constant speed, a standard deviation of 0.8 miles per hour was archived, verifying that highly accurate speed measurements are possible from SkySat videos.

We then extracted moving points based on the algorithm described in Section Motion Analysis. Figure 9 shows the generated background image with the acquired moving object tracks over Las Vegas. Most tracks correspond to vehicles on the roads, but some tracks correspond to high buildings and the image border.

Automatic vector based georeferencing of video

Figure 10 shows the Las Vegas image localized onto OpenStreetMap vector data, using the automatic, geometric hashing registration method, based on the tracks shown in Figure 9. The OpenStreetMap road vectors align well with the image. A comparison with USGS 0.5 foot high resolution ortho images, shows that the co-registration improves the SkySat $RMSE_{XY}$ from 77.2 m to 21.7 meters.

3D reconstruction from video

One possible application for the video product is DSM generation. With traditional satellite imaging, typically, only a stereo pair or triplet is available. For DSM generation, we have temporally subsampled the Las Vegas video sequence at 1 Hz, leaving 60 images, since adjacent frames exhibit a very small baseline and cannot be

used for matching. An absolute block adjustment has been performed on the 60 images, resulting in a tie point RMSE of 0.1 pixels.

DSM generation was performed by using a total variation (Kuschk, 2013) based image matching. 14 master images were used and each was matched against the 20 temporally closest images. The resulting 14 DSMs were merged using a mean and median approach. The resulting DSM was evaluated against a LIDAR point cloud, containing both ground and non-ground objects, showing an RMSE of 10.8 m, and a normalized median deviation (NMAD) of 1.2 m. The high RMSE value is due to small remaining co-registration and image matching issues on the high building edges and in noisy shadow regions, leading to a small amount of large elevation differences, which violate the assumption of normally distributed errors. For such scenes, the NMAD, which is not influenced by such gross differences, is an alternative measure (Höhle & Höhle, 2009).

Conclusion

With the first civil very high resolution spaceborne video products, the SkySat satellites offer a very interesting data source for future applications. The "push-frame" satellite architecture, together with the ground based image fusion and super-resolution approach reduce the internal complexity and therefore the cost of SkySat satellites, potentially enabling the launch of a large constellation, with multiple daily revisits. A drawback of the SkySat satellites is the relatively small footprint of the video and still imagery products. Primary applications of SkySat data are thus local high frequency monitoring applications.

Considering the small and relatively low cost satellites, the image quality is good, with a higher noise level in the videos. According to the visual comparison, the radiometric quality of SkySat data is lower than a WV-3 data acquired in similar conditions, but still sufficient for most applications.

Direct georeferencing accuracy is low and was in the range of 100 to 200 m. For high quality orthorectification using the standard RPC based workflow, multiple GCPs should be available for each of the many images in the raw frame product. The WV-3 data required only 2 GCPs while the SkySat frames required more than 50 GCPs, but could not achieve the same accuracy. Future work should include using the physical camera model for image orientation, this will potentially reduce the number of required

ground control points. The absolute accuracy reached in the evaluation was limited by the available reference data, not the satellite datasets. The merged SkySat geo product is easier to handle than the single frames product, but cannot be adjusted or orthorectified well, and is thus not suitable when precise georeferencing is required.

The video offers use cases for dynamic processes, object detection and tracking. Vehicle speeds could be detected with an estimated precision better than 0.8 mph. Automatic motion detection and subsequent georeferencing to vector data was performed and can improve the geolocation accuracy of videos.

Acknowledgements

References

Ackermann, F., 1984. Digital Image Correlation: Performance and Potential Application in Photogrammetry. *The Photogrammetric Record,* 11(64), pp. 429-439.

Aguilar, M. A., Saldaña, M. d. M., Aguilar, F. J. & Lorca, A. G., 2014. Comparing geometric and radiometric information from GeoEye-1 and WorldView-2 multispectral imagery. *European Journal of Remote Sensing,* Issue 47, pp. 717-738.

d'Angelo, P., 2013. *Automatic Orientation of large multitemporal Satellite Image Blocks.* Nanjing, China, Proceedings of International Symposium on Satellite Mapping Technology and Application.

d'Angelo, P., Kuschk, G. & Reinartz, P., 2014. *Evaluation of Skybox Video and Still Image products.* Denver, Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XL-1.

Dzamba, T. e. a., 2014. *Success by 1000 Improvements: Flight Qualification of the ST-16 Star Tracker.* Logan, USA, 28th Annual AIAA/USU Conference on Small Satellites.

Grodecki, J. & Dial, G., 2003. Block adjustment of high-resolution satellite images described by rational polynomials. *Photogrammetric Engineering & Remote Sensing,* 69(1), pp. 59-68.

Höhle, J. & Höhle, M., 2009. Accuracy assessment of digital elevation models by means of robust statistical methods. *ISPRS Journal of Photogrammetry and Remote Sensing,* 4(64), pp. 398-406.

Kuschk, G., 2013. *Model-free Dense Stereo Reconstruction Creating Realistic 3D City Models.* Sao Paulo, Brazil, Proceedings JURSE, pp. 202-205.

Lowe, D. G., 2004. Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision,* 60(2), pp. 91-110.

Mattyus, G., Benedek, C. & Sziranyi, T., 2010. *Multi Target Tracking on Aerial Videos.* Istanbul, ISPRS Workshop on Modeling of Optical Airborne and Space Borne Sensors.

Mattyus, G. & Friedrich, F., n.d. Aerial image geolocalization with road traffic as invariant feature. *Under review in Image and Vision Computing.*

Murthy, K. et al., 2014. *SkySat-1: very high-resolution imagery from a small satellite.* Amsterdam, Proc. SPIE 9241, Sensors, Systems, and Next-Generation Satellites XVIII.

Planet Labs, 2015. *Planet Labs.* [Online]
Available at: https://www.planet.com
[Accessed 23 April 2015].

Robinson, D. et al., 2014. *High-Resolution Imagery and Video from SkySat-1.* Louisville, KY, Proc. Joint Agency Commercial Imagery Evaluation.

Skybox Imaging, 2015. *Skybox Imaging.* [Online]
Available at: http://www.skyboximaging.com/
[Accessed 23 April 2015].

Smiley, B. et al., 2014. *On-orbit Calibration and Validation of the Skybox Imaging Constellation.* Denver, Pecora 19.

Wolfson, H. & Rigoutsos, I., 1997. Geometric hashing: an overview. *Computational Science Engineering, IEEE,* 4(4), pp. 10-21.

Zivkovic, Z. & van der Heijden, F., 2006. Efficient Adaptive Density Estimapion per Image Pixel for the Task of Background Subtraction. *Pattern Recognition Letters,* 27(7), pp. 773-780.
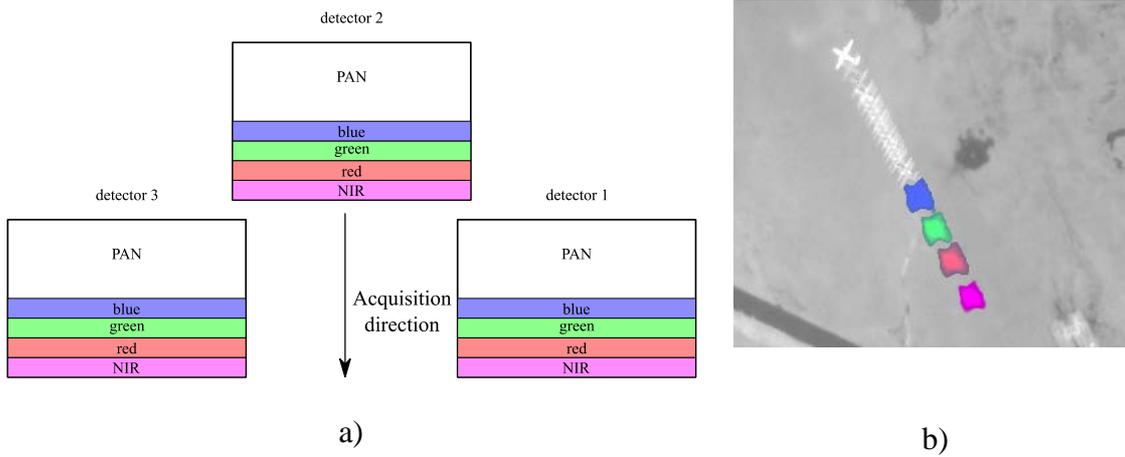
Figure 1: a) SkySat Focal Plane assembly, as projected to the ground. b) Effect of image fusion on a moving object. The purple/near infra-red, red, green and blue blobs show the plane as imaged by a handful of multi spectral frames, and the approximately 18 gray planes illustrate how many panchromatic images were fused.
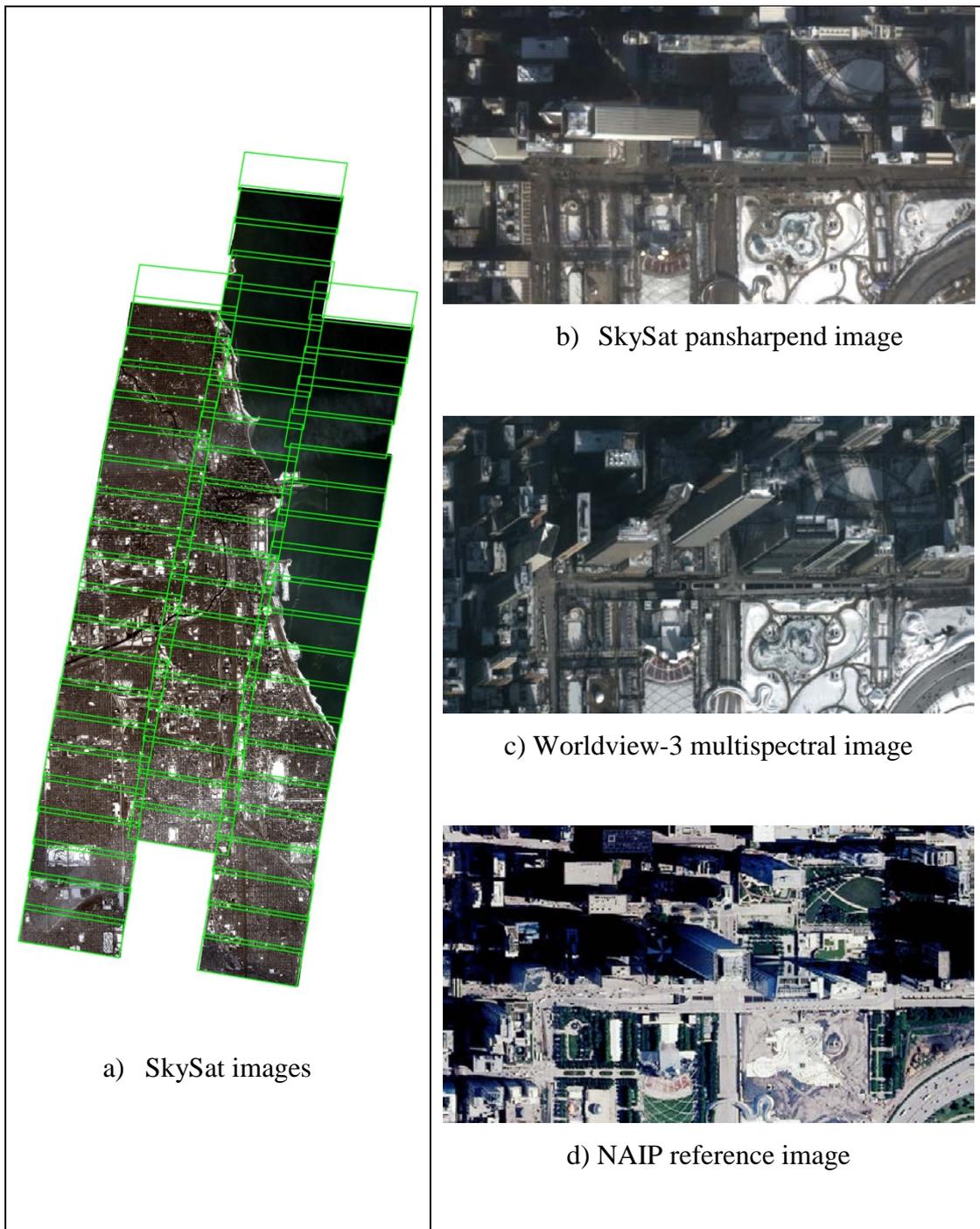
Figure 2: a) Chicago SkySat acquisition. The green rectangles show the footprint of the individual images. b-d) Image detail of the evaluated datasets.

100 meter

5 meter

a)

b)

100 meter

5 meter
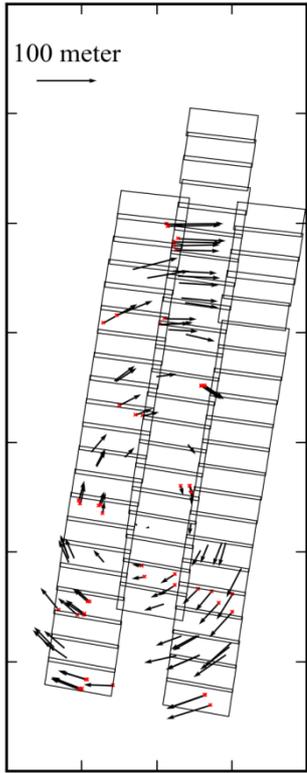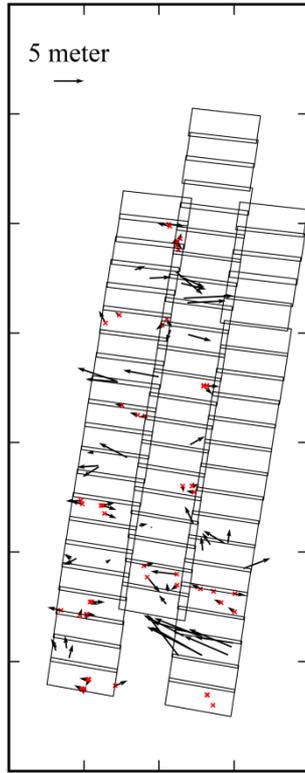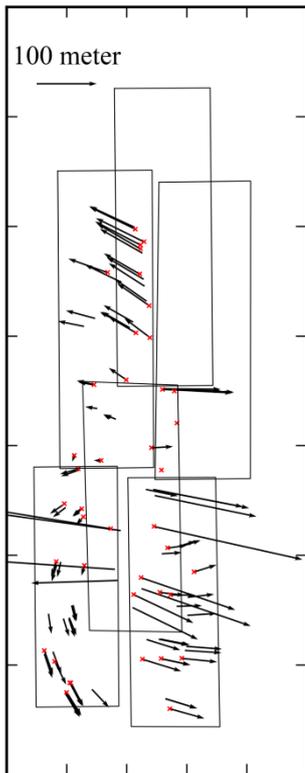
c)
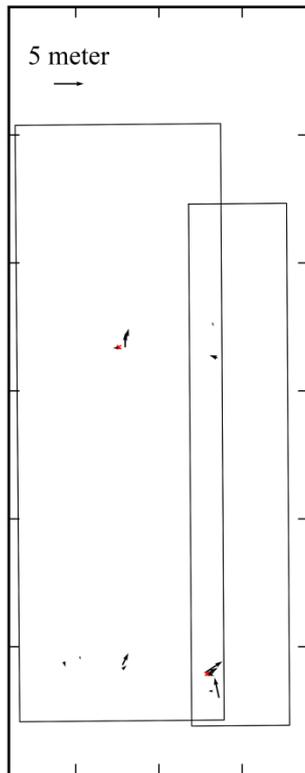
d)

Figure 3: GCP and checkpoint residuals after bundle block adjustment, with the Grid GCP distribution, using GCPs in every fourth frame. Arrows show the X/Y error after adjustment, GCPs used in the adjustment are marked with red crosses. a) SkySat frames adjusted using bias RPC correction. b) SkySat frames adjusted using affine RPC correction. c) SkySat geo product after adjustment with bias RPC corrections. Note the different scaling in a) and c). d) Results of adjustment of WV-3 multispectral scene.
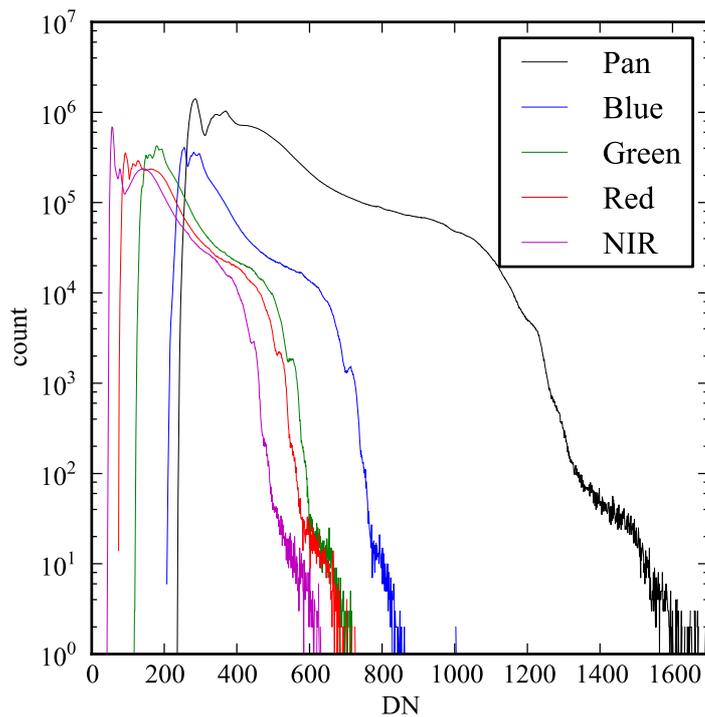
Figure 4: Histogram of the Panchromatic and multi spectral SkySat Chicago scene.
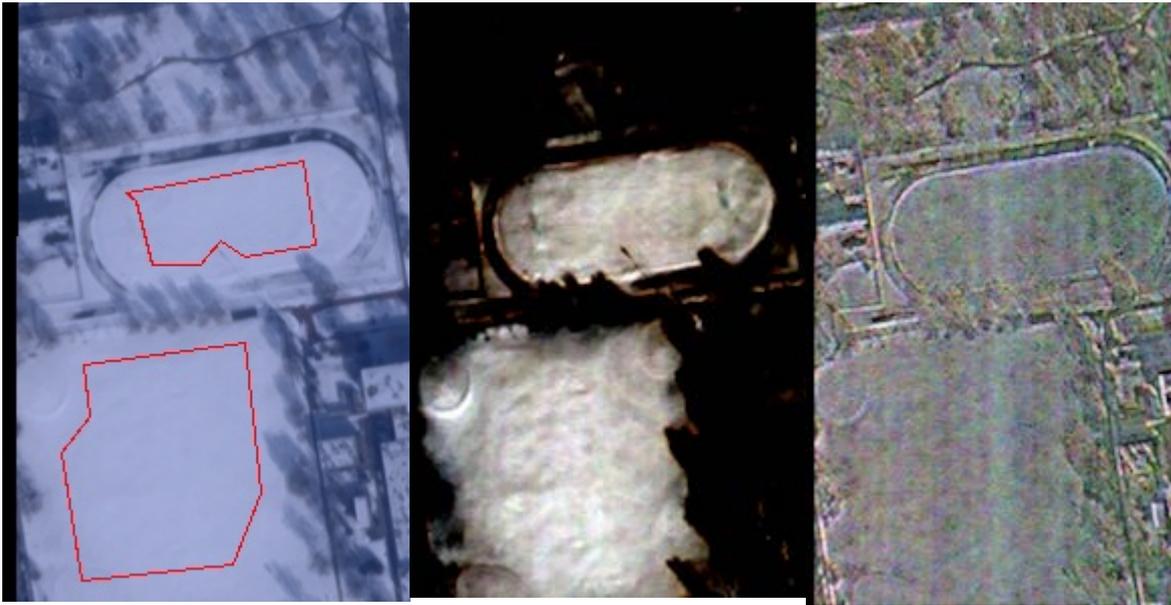
Figure 5: Image noise evaluation of SkySat data. a) Snow covered sports grounds used for noise evaluation. b) Contrast enhanced version of a), snow area is not completely uniform and cannot be used for direct noise estimation. c) Scene structure is removed by computing the difference of co-registered images of detector 1 and detector 2. Image noise is clearly visible and scene structures in flat areas are supressed.
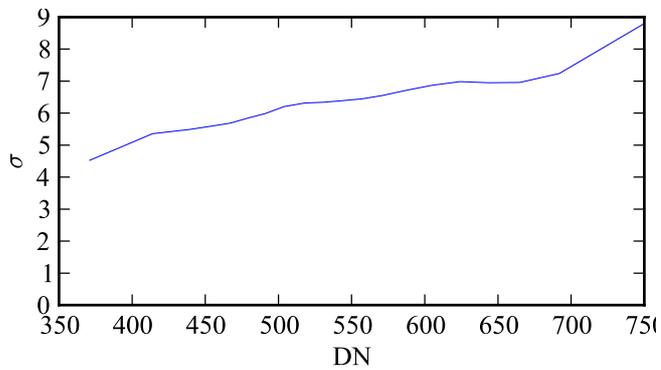

Figure 6: Image noise σ vs DN for the Mirny SkySat 1 raw video collection.
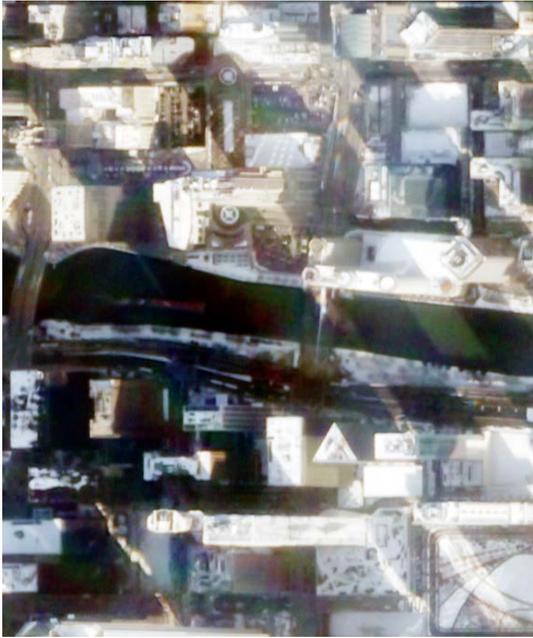
a) SkySat pansharpend   b) WV-3 multispectral

Figure 7: Visual comparison of a deep shadowed area in downtown Chicago. The WV-3 image is covered by light clouds in the upper part. Images were enhanced with histogram equalization.
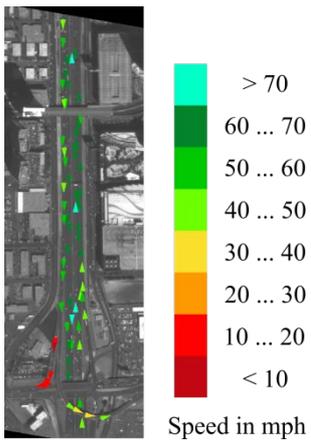
Figure 8: Speed measurement on Las Vegas Freeway
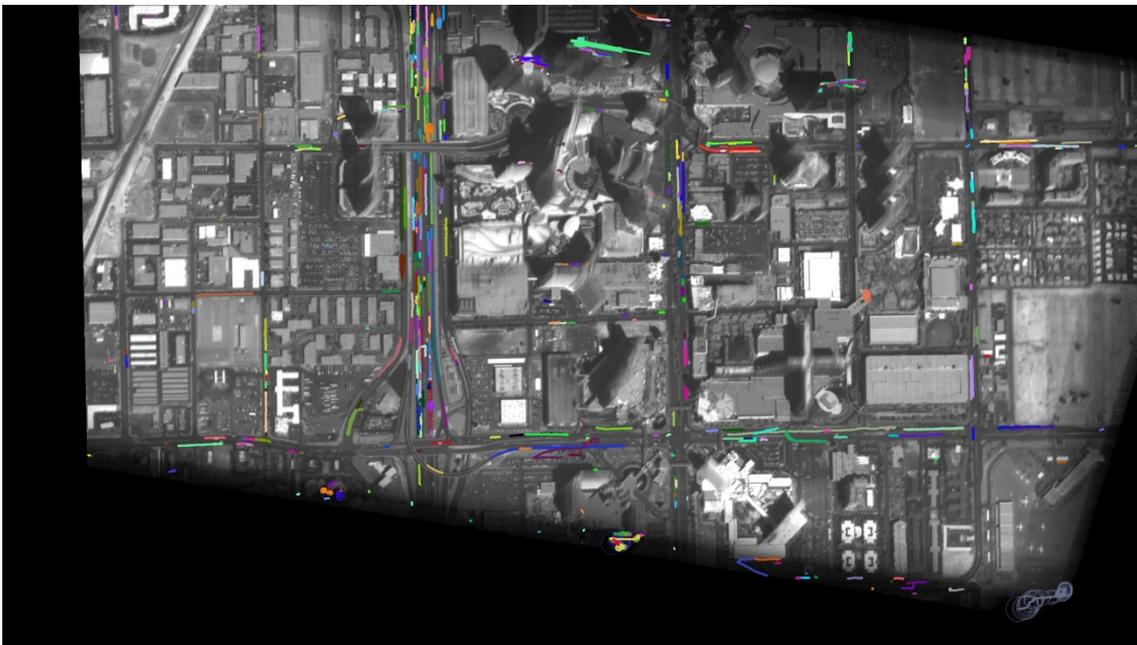
Figure 9: Las Vegas. The vehicle tracks on the background image generated from the video.
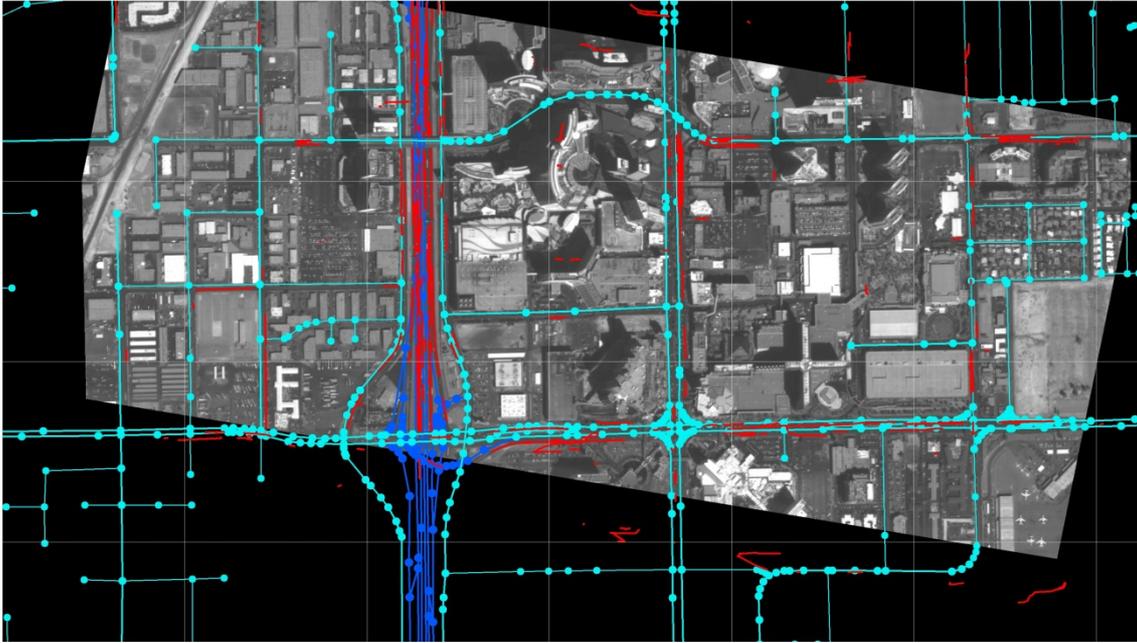
Figure 10: Las Vegas. The OpenStreetMap road network is projected into the image after localization. The object tracks (in red) are matched to the road network (in blue and cyan).
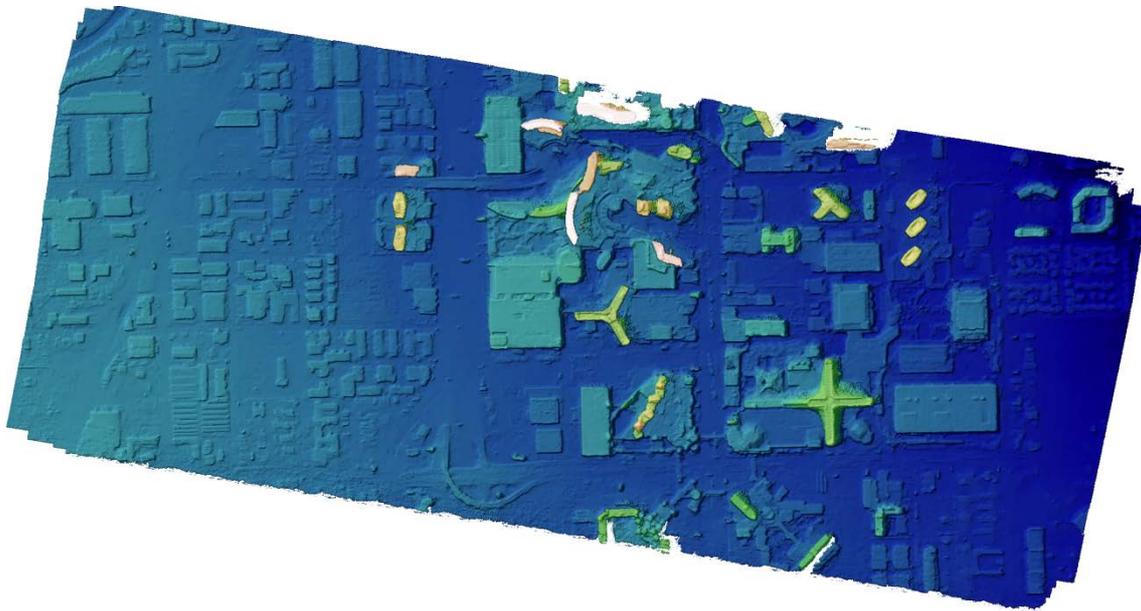

Figure 11: Las Vegas DSM computed from 60 images.