The TUM-DLR Multimodal Earth Observation Evaluation Benchmark

Tobias Koch[1], Pablo d'Angelo[2], Franz Kurz[2], Friedrich Fraundorfer[2,3], Peter Reinartz[2], Marco Körner[1,*]

[1]Remote Sensing Technology
Technical University of Munich
Munich, Germany
{tobias.koch,marco.koerner}@tum.de

[2]Remote Sensing Technology Institute
German Aerospace Center
Oberpfaffenhofen, Germany
{pablo.angelo,franz.kurz,peter.reinartz}@dlr.de

[3]Institute for Computer Graphics and Vision
Graz University of Technology
Graz, Austria
fraundorfer@icg.tug.at

Abstract

We present a new dataset for development, benchmarking, and evaluation of remote sensing and earth observation approaches with special focus on converging perspectives. In order to provide data with different modalities, we observed the same scene using satellites, airplanes, unmanned aerial vehicles (UAV)*, and smartphones. The dataset is further complemented by ground-truth information and baseline results for different application scenarios.*

The provided data can be freely used by anybody interested in remote sensing and earth observation and will be continuously augmented and updated.

1. Introduction

The overall availability of free and open large-scale datasets for the purpose of developing, testing, and benchmarking novel methods is one of the key reasons for the enormous progress achieved in the computer vision and machine learning research community during the last years. Most prominently, methods like *deep learning* benefit massively from annotated data and show state-of-the-art results which can today be regarded as the gold standard.

In order to further promote such an evolution in the remote sensing and earth observation community, we present a new extensive dataset covering different image modalities. For this purpose, we collected image data from satellites, airplanes, drones, and smartphones capturing the same area (*cf.* Fig. 1). Furthermore, we provide interior and exterior images of one specific building located in the target area, as well as baseline results for selected application scenarios, such as, for instance, 3d reconstruction, ego-motion estimation, or next-best-view planning. We are further planning to enrich this dataset continuously by adding new data modalities and ground-truth annotations.

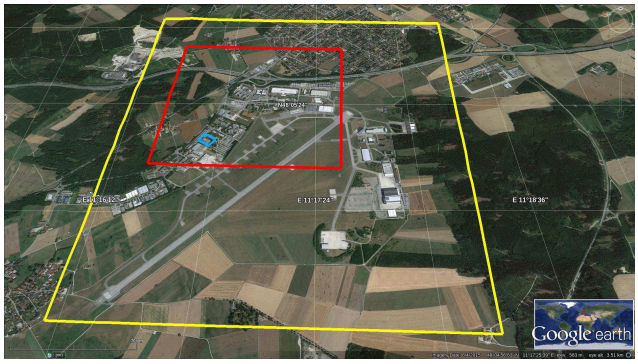


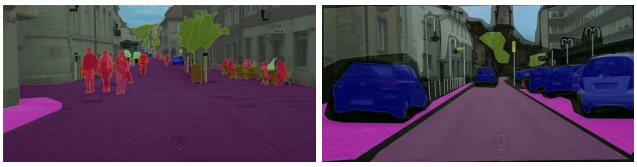
Figure 1: The area covered by the dataset is located in the southwest of Munich, Germany, and was observed by WORLDVIEW-2 satellite (yellow), airborne DSLR cameras (red), as well as UAV-borne ILCE and GoPro cameras (blue).

This dataset can be freely used by everybody interested in optical remote sensing and earth observation. We believe that this will help to compare and boost competing approaches.

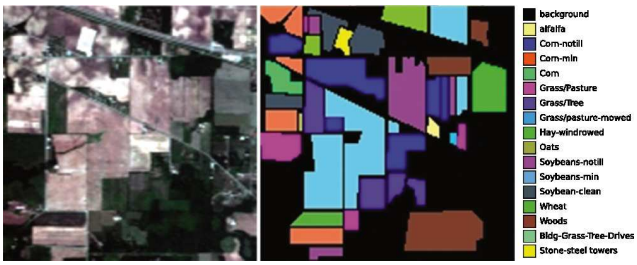
The remainder of this paper is dedicated to briefly review existing benchmark suites and to describe our proposed dataset in detail in Sections 2 and 3, respectively. Section 4 will outline prospective application scenarios.

*Corresponding author

(a) *ImageNet* [8] semantic hierarchy



(b) *Cityscapes* [7] street-view dataset



(c) *indian pine* land cover dataset [6]

(d) *ISPRS Benchmark* with images from terrestrial, UAV, and airborne cameras [14]

Figure 2: Exemplary datasets from the research domains of computer vision and machine learning, as well as of remote sensing and earth observation.

2. Related Work

The free and unrestricted availability of data enriched by ground-truth annotations became ubiquitous in the field of computer vision and machine learning within the recent past and these research communities benefited a lot from it. For instance, the *ImageNet* [8] database semantically organizes more than 14,000,000 images into a hierarchy of more than 20,000 classes, as exemplified in Fig. 2a. As another example, the recently published *Cityscapes* [7] dataset provides 25,000 street view stereo images acquired from 50 German cities. These come with manual pixel-level as well as instance-level class annotations (*cf.* Fig. 2b) and various meta-data, such as GPS coordinates, ego-motion data, and further vehicle sensor data.

These datasets, among many others, enabled for a massive evolution of computer vision and machine learning methods, such as, for instance, object detection, classification, segmentation, or tracking.

In contrast to that, researchers from the field of remote sensing and earth observation to date only have access to a rather limited amount of free data for scientific purposes. For instance, the 224-band multi-spectral *indian pine* [6] dataset, as shown in Fig. 2c, is extensively used for evaluation of land cover classification approaches, despite its rather small extent of 145×145 px, as well as its poor coverage and completeness. If ground-truth information is dispensable, data from public access satellite missions—*e.g.*, LANDSAT [15] or COPERNICUS SENTINEL [3, 5]—can be obtained at very low effort.

Nevertheless, this issue came more and more into focus of the remote sensing and earth observation research community in the recent past. For instance, the *ISPRS Benchmark for Multi-Platform Photogrammetry* [14] provides airborne, UAV, and terrestrial images of urban scenes in Dortmund, Germany, as exemplary shown in Fig. 2d, as well as corresponding LIDAR point clouds.

With releasing our dataset to the public, we aim to complement these existing ones by augmenting it by satellite imagery and further annotations.

3. Dataset Description

All acquired images in our dataset show the region around the *German Aerospace Center (DLR)* campus in Oberpfaffenhofen near Munich, Germany, including the *Earth Observation Center (EOC)* building, as shown in Fig. 1. The dataset contains images of different modalities acquired by satellites, airplanes, UAVs, and smartphones, as will be described in the following and summarized in Table 1.

3.1. Satellite Images

The `DLR-Satellite`[1] dataset contains four panchromatic and multispectral images of a $2.6 \times 3.2\,\mathrm{km}^2$ large area around the DLR campus, as illustrated in Fig. 3. These were captured by the DIGITALGLOBE satellites WORLDVIEW-2 [1, 4] and GEOEYE-1 [2] operating in descending, sun-synchronous orbits. Both satellites simultaneously carry a pan-chromatic and a multi-spectral sensor with *ground sampling distances (GSD)* of around 50 cm and

[1] For the sake of discriminability, we set parts of our proposed dataset in `typewriter` font, while external datasets are set in *italics*.

(a) panchromatic image, 2010

(b) RGB image, 2010

(c) RGB image, 2015

Figure 3: The satellite data included in this dataset consist of panchromatic and multispectral images acquired in 2010 by WORLDVIEW-2 as well as in 2015 by GEOEYE-1.

2 m, respectively. The multi-spectral sensor of GEOEYE-1 delivers four color channels—*i.e.*, red, green, and blue—, as well as near infrared. WORLDVIEW-2 offers 4 additional multi-spectral channels supporting applications like vegetation classification or imaging of shallow water areas. The exact resolution depends on the imaging geometry, which varies for each scene. All images are provided in *Level 2 Ortho Ready* processing level. The images are radiometrically and geometrically corrected and resampled to their final resolution. Each image is accompanied by a set of *rational polynomial coefficients (RPC)* [9] describing the projection from world into image coordinates. These can be used for ortho-rectification, co-registration, and stereo processing.

In order to allow change detection applications to benefit from these data, images were collected with a temporal difference of almost five years: 2010-10-22 and 2015-10-15.

3.2. Airborne Image Sequence

The DLR-3K aerial image dataset was acquired on 2015-04-09 using the airborne *DLR 3K* sensor system [13] and covers about 25% of the target area, as indicated by the red polygon in Fig. 1. This setup contains cameras looking in nadir, forward, and backward direction (*cf.* Fig. 4a) and was operated in *along track mode*. Due to the flight height of about 400 m above ground, the images show a ground sampling distance of about 6 cm. We selected 105 backward looking, 105 forward looking, and 111 nadir looking images from the test area. Figure 4c shows the respective footprints of the whole image sequence.

Each image in the aerial image sequence comes with a individual meta-data file reporting the position of the projection center in UTM coordinates, the intrinsic and extrinsic camera parameters—such as camera orientation as a 3×3 Euclidean rotation matrix and focal length given in pixels—as well as other auxiliary information. In order to ease the usage of the data, all aerial images were undistorted

(a) the *DLR 3K* sensor setup uses 3 DSLR cameras with different viewing angles

(b) satellite image augmented by orthorectified and tessellated nadir images

(c) footprints of complete image sequence: nadir (red), forward (magenta), and backward (cyan) looking camera

Figure 4: The DLR-3K aerial image sequence was acquired by the air-borne *DLR 3K* camera setup and covers an area of about 1 km^2.

wrt. radial lens distortion and the principal points coincide with the image central points, *i.e.*, no further corrections have to be applied to the aerial images. Beside of that, the dataset also contains orthorectified aerial images obtained by projecting the individual images onto a global *digital elevation model (DEM)* using the extrinsic and intrinsic camera calibration parameters. All nadir images were further tessellated and radiometrically corrected.

3.3. UAV Image Sequences

The EOC-UAV dataset provides images of the EOC building located at the DLR campus, as indicated by the blue polygon in Fig. 1. It consists of two complementary image sequences showing the roof and façades of the structure. While one sequence was captured using a high-resolution ICLE camera, the images of the second sequence where acquired by a low-resolution and wide-angle GOPRO action camera. These cameras were mounted on a ASCTEC

Figure 5: Target building of the EOC-UAV aerial and indoor datasets.

FALCON 8 UAV. In order to ensure comparable recording and environmental conditions, both sequences were created at the same day (2014-11-12). The target building is characterized by a prismatic structure of approximately 60 m length, 25 m width, and 15 m height. As exemplary shown in Fig. 5, the building is dominated by windows, as well as poorly textured and highly reflective façade elements, impeding feature-based image matching.

3.3.1 ICLE Sequence

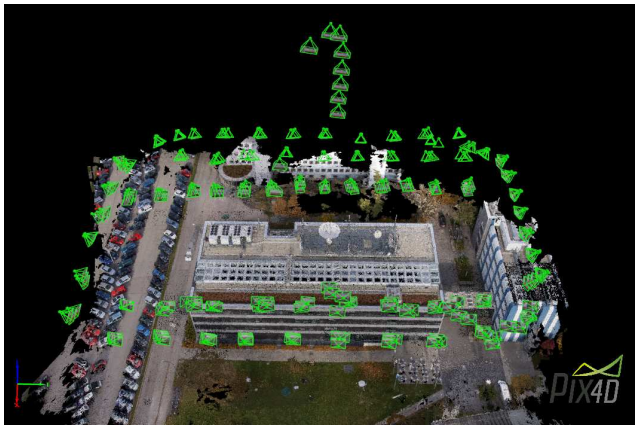
The `EOC-UAV-ICLE` sequence contains 376 24 MP images captured by a UAV-mounted SONY NEX-7 ICLE camera fitted with a zoom lens fixed at a focal length of 17 mm and an aperture stopped down to $f/9.0$. The sequence shows the entire building from oblique and nadir perspectives, as well as additional images facing the façades horizontally. In total, there are 49 nadir, 200 oblique, and 127 façade images available in the dataset. Sample images of the different views, as well as the path of all nadir and oblique image positions, are shown in Fig. 6a and Fig. 6c, respectively. While the façade images were acquired during a manual flight mode, the flight trajectories of the oblique and nadir images have been pre-planned, in order to ensure an overlap at ground level of roughly 80%. Besides the intrinsic calibration parameters of the camera, additional localization information is provided for each image, *e.g.*, UTM coordinates, barometer heights, and yaw orientation angles. The baseline between GPS antenna and the optical camera center is not known.

3.3.2 GoPro Sequence

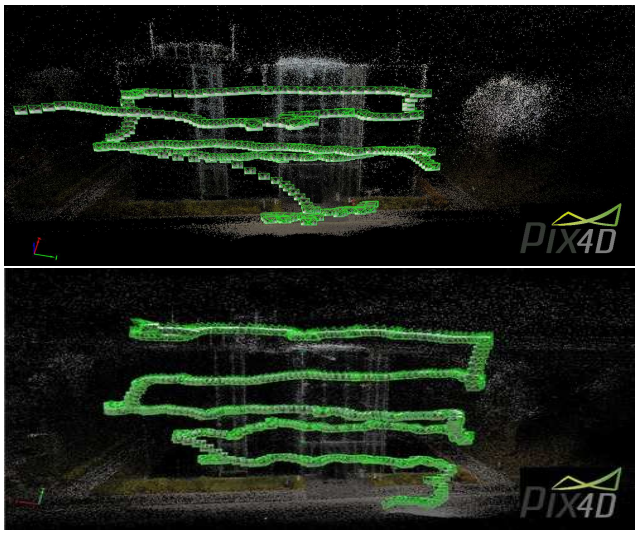
The `EOC-UAV-GoPro` dataset compiles 1018 12 MP frames extracted from video streams acquired by a GOPRO HERO4 SILVER during two flights along the front entrance

(a) oblique, nadir, and facade views from the `EOC-UAV-ICLE` sequence

(b) horizontal and oblique views from the `EOC-UAV-GoPro` sequence



(c) reconstructed `EOC-UAV-ICLE` flight trajectory



(d) reconstructed `EOC-UAV-GoPro` flight trajectories

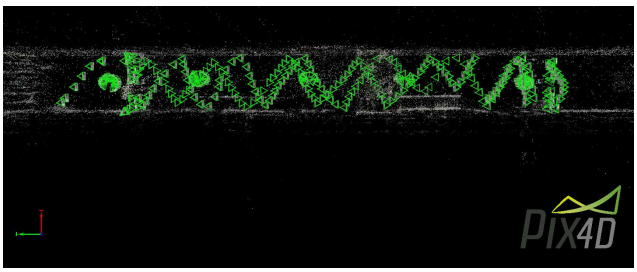
Figure 6: Sample images from the `EOC-UAV` datasets showing the exterior of the EOC building.

(a) front entrance

(b) back entrance

(c) corridor

(d) stairway



(e) camera trajectory reconstructed from the image sequence

Figure 7: Sample images from the EOC-DSLR dataset showing the interior of the EOC building.

of the EOC building. The separate flights only differ wrt. camera orientation, *i.e.*, horizontal and 45° oblique. Due to the very wide-angle lens with focal length 3.0 mm, the images contain very large overlap (*cf.* Fig. 6b) and can be used to complement the EOC-UAV-ICLE dataset. On the downside, the open aperture ($f/2.8$) determines a rather shallow depth of field and the images are further affected by rolling shutter effects and motion blur. Camera intrinsics are provided.

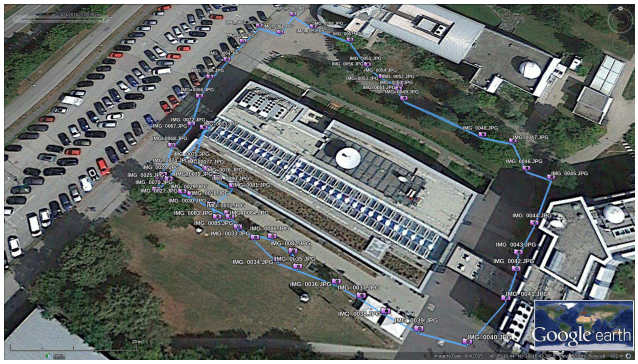
3.4. Handheld Image Sequences

The presented dataset is complemented by sequences of images acquired by humans walking on the ground plane around and inside the EOC building. In order to cover multiple image modalities and their specific properties, these sequences were recorded using both a DSLR camera, as well as a smartphone.

3.4.1 DSLR Indoor Sequence

The EOC-DSLR image sequence was acquired using a hand-held CANON EOS 500D DSLR camera and contains

(a) sample images showing the building exterior



(b) GPS track of the exterior sequence extracted from the image meta-data

Figure 8: Sample images from the EOC-smartphone dataset and the corresponding (noisy) GPS track.

340 high resolution images (15 MP) showing the interior of the EOC building. The building comprises three floors with open-spaced, elongated corridors and large panorama windows at both face sides. Similar to the outdoor façades, the indoor walls are barely textured. All images were acquired on 2015-10-09 following a a zig-zag trajectory through the ground floor in both directions, connected by three panorama views (*cf.* Fig. 7e). Additionally, the sequence contains images of both window façades captured from the first floor level. A 18–55 mm zoom lens was mounted to the DSLR and fixed to the minimal focal length of 18 mm. The aperture was stopped down to $f/9.1$, in order to obtain a larger depth of field. Exposure was controlled automatically with priority to lower ISO values. Camera intrinsics are provided.

Sample images from this dataset are compiled in Figs. 7a to 7d.

3.4.2 Smartphone Indoor and Outdoor Sequences

The EOC-smartphone image sequences contain 340 and 63 images of lower resolution (8 MP) showing the interior and exterior of the EOC building, respectively, and was acquired on 2016-03-24 using the rear camera of an APPLE IPHONE 6 smartphone in portrait orientation. According to the manufacturer specifications, this camera has an $f/2.2$

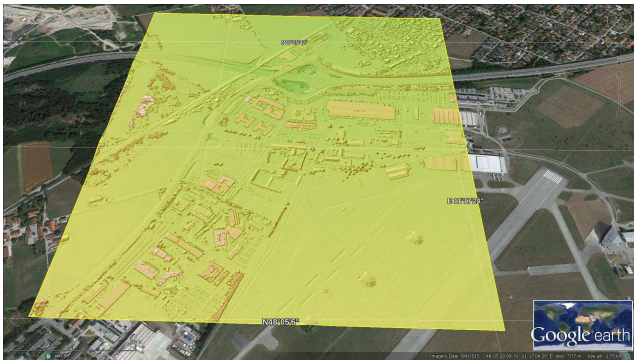


Figure 9: A *digital surface model (DSM)* generated from the DLR-3K aerial image sequence.

aperture and a focal length of 4.15 mm. Exposure was automatically controlled with priority on low ISO values. While this acquisition mode resulted in sharp and well-exposed images in the outside scenario, the images captured inside the EOC building are more commonly affected by motion blur.

Figure 8 shows exemplary images aside a camera path extracted from GPS tags encoded in the image meta-data.

4. Applications

We believe that the data provided by our dataset is of relevance for the remote sensing and earth observation research community. In order to substantiate this claim, we want to briefly outline a selection of possible applications and baseline results.

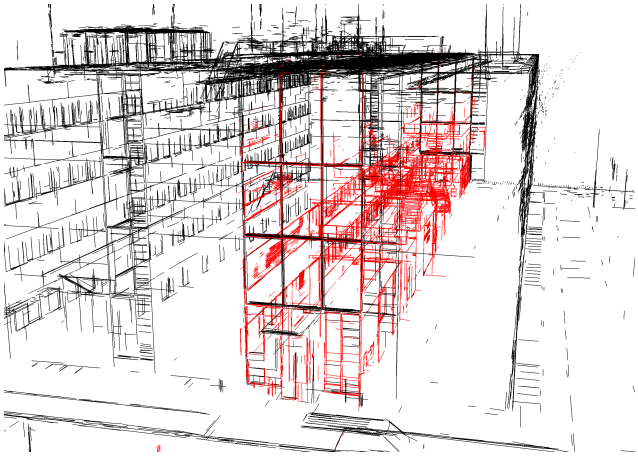
4.1. Surface Modeling from 3K Image Data

The high degree of coverage and large overlap of images included in the DLR-3K aerial sequence allows for computation of a dense *digital surface model (DSM)*. For this purpose, we used the extension of the *semi-global matching (SGM)* algorithm [10] proposed by Kraus *et al*. [12]. The final DSM included into the dataset was resampled to 20 cm ground pixel size and is shown in Fig. 9.

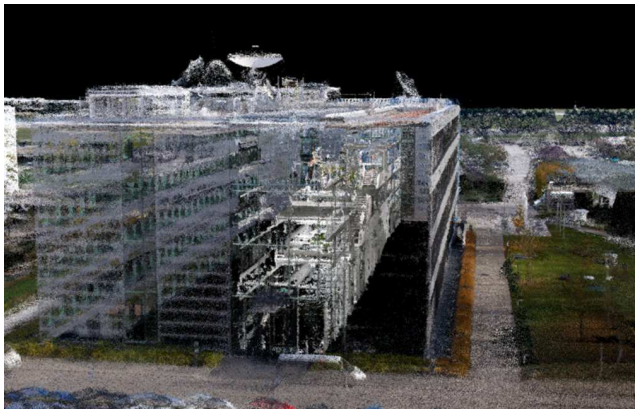
4.2. Building Model Fusion

The availability of interior and exterior images of the EOC building offers the opportunity to combine both datasets for generating a holistic topologically and geometrically correct building model. Due to the presence of large façade windows, which can be simultaneously identified in both the outside an inside views, such a model can be generated by fusion based on geometric as well as semantic cues.

Figure 10 exemplary shows the result of aligning both particular models by the approach proposed by Koch *et al*. [11] relying on straight 3d line segments extracted directly from the input images.



(a) aligned indoor and outdoor 3d lines



(b) aligned indoor and outdoor point clouds

Figure 10: Building model fusion of the EOC building using the EOC-UAV-ICLE and EOC-DSLR images based on 3d line segments [11].

4.3. Fusion of Aerial and UAV images

The pleasing properties of the EOC-UAV-ICLE aerial image dataset—*i.e*., the high spatial resolution and low operating altitude —combined with with the accompanied GPS information can be exploited to generate a high-quality orthophoto and DSM with ground sampling distance of approximately 1 cm, as exemplary shown in Fig. 11. Fusing the high-resolution and accurately geolocalized aerial images with the lower-quality geolocalized UAV images can help to partially increase the resolution of the DLR-3K images.

4.4. Change Detection

One of the central tasks in remote sensing and earth observation is the monitoring of land areas and the automatic detection of changes. For this purpose, the large temporal baseline represented in the DLR-WorldView-2 and

Table 1: Overview over the characteristics of the sequences included within the presented dataset.

Property	Dataset						
	DLR-Satellite		DLR-Aerial	EOC-UAV		EOC-Handheld	
	DLR-WorldView-2	DLR-GeoEye-1	DLR-3K	EOC-UAV-ICLE	EOC-UAV-GoPro	EOC-DSLR	EOC-Smartphone
Observed area	$2.6 \times 3.2\,\mathrm{km}^2$	$2.6 \times 3.2\,\mathrm{km}^2$	$1.4 \times 1.5\,\mathrm{m}^2$	$100 \times 60\,\mathrm{m}^2$	$100 \times 60\,\mathrm{m}^2$	$60 \times 25\,\mathrm{m}^2$	$60 \times 25\,\mathrm{m}^2$
Target	DLR campus	DLR campus	EOC building	EOC building	EOC building	EOC building (indoor)	EOC building (indoor, outdoor)
Acqusition date	2010-10-22	2015-10-15	2015-04-09	2014-11-12	2014-11-12	2015-10-09	2016-03-24
Sensor	DigitalGlobe WorldView-2	DigitalGlobe GeoEye-1	DLR-3k	Sony Nex-7	GoPro HERO4 Silver	Canon EOS 500D	Apple iPhone 6
Spectral resolution	panchromatic: 450–800 nm multi-spectral: 400–450 nm (coastal) 450–510 nm (blue) 510–580 nm (green) 585–625 nm (yellow) 630–690 nm (red) 705–745 nm (red edge) 770–895 nm (NIR 2) 860–1040 nm (NIR 2)	panchromatic: 450–800 nm multi-spectral: 450–510 nm (blue) 510–580 nm (green) 655–690 nm (red) 780–920 nm (NIR)	RGB	RGB	RGB	RGB	RGB
Geometric resolution	50 cm (GSD, pan.), 200 cm (GSD, mult.)		6 cm (GSD)	N/A	N/A	N/A	N/A
Sensor resolution	5292×6410 px (pan.), 1323×1602 px (mult.)		5623×3712 px	6000×4000 px	4000×3000 px	4752×3168 px	3264×2448 px
Image format	TIF (lossless), 16 $^{\mathrm{bit}}/_{\mathrm{px}}$	TIF (lossless), 16 $^{\mathrm{bit}}/_{\mathrm{px}}$	JPEG, 8 $^{\mathrm{bit}}/_{\mathrm{px}}$	JPEG, 8 $^{\mathrm{bit}}/_{\mathrm{px}}$	JPEG, 8 $^{\mathrm{bit}}/_{\mathrm{px}}$	JPEG, 8 $^{\mathrm{bit}}/_{\mathrm{px}}$	JPEG, 8 $^{\mathrm{bit}}/_{\mathrm{px}}$
Frames	1	1	105 backward, 105 forward, 111 nadir	49 nadir, 200 oblique, 127 façade	454 nadir, 564 oblique	340	340 indoor, 63 outdoor
Additional data	rational polynomial coefficients (RPC)		camera calibration & orientation, auxillary metadata, orthorectif. images	camera calibration & yaw orientation, GPS tags, barometer heights	camera calibration	camera calibration	camera calibration, GPS tags
Post-processing	radiometricall & geometrically corrected, re-sampled		geometrically corrected	N/A	N/A	N/A	N/A

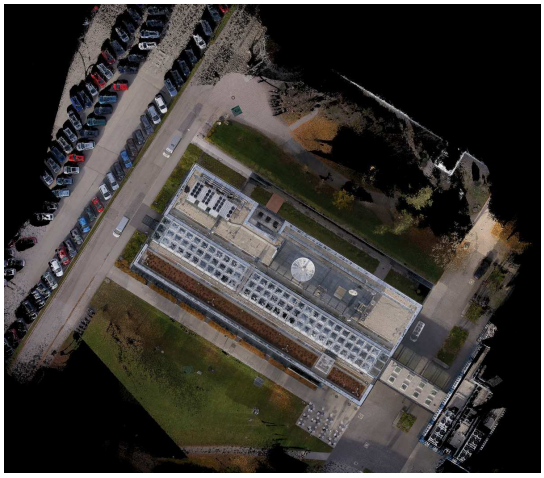


Figure 11: Georeferenced orthophoto created from the EOC-UAV-ICLE images with 1 cm GSD.

DLR-GeoEye-1 datasets, as well as the smaller temporal baselines between DLR-3K and EOC-UAV datasets, can be exploited to track structural changes in urban environments.

5. Summary and Outlook

We presented a new dataset of image sequences showing a concise land area intended to assist the development and evaluation of approaches raising from remote sensing and earth observation research questions. These sequences were acquired by different sensors—*e.g.*, multi-spectral, RGB, and panchromatic sensors with differing image resolution—and from converging perspectives—*e.g.*, satellite, air-born, UAV-born, as well as hand-held sensors—, literally *from satellite to street*.

The complete dataset including meta information and baseline results can be downloaded from http://www.lmf.bgu.tum.de/tum-dlr-multimodal. While the use of the WorldView-2 and GeoEye-1 satellite images is restricted to the enclosed license agreement, all other image sequences included into this dataset can be freely used by anyone interested for scientific purposes.

We intend to augment this dataset in the future by further sensor modalities—such as *synthetic aperture radar (SAR)* data—and ground-truth information, *e.g.*, pixel-wise segmentation, object annotations, cadastral information, or CAD building models. Baseline results for possible application scenarios will be added consecutively.

References

[1] The Benefits of the 8 Spectral Bands of WorldView-2. Technical report, DigitalGlobe, 2010. http://global.digitalglobe.com/sites/default/files/DG-8SPECTRAL-WP_0.pdf (retrieved: May 1, 2016).

[2] GeoEye-1. Technical report, DigitalGlobe, 2013. http://global.digitalglobe.com/sites/default/files/DG_GeoEye1.pdf (retrieved: May 1, 2016).

[3] Sentinel-2 User Handbook, 2013. https://earth.esa.int/documents/247904/685211/Sentinel-2_User_Handbook (retrieved: May 1, 2016).

[4] WorldView-2. Technical report, DigitalGlobe, 2013. http://global.digitalglobe.com/sites/default/files/DG_WorldView2_DS_PROD.pdf (retrieved: May 1, 2016).

[5] Copernicus Sentinel Data, 2016. https://scihub.copernicus.eu/dhus, (retrieved: May 1, 2016).

[6] M. F. Baumgardner, L. L. Biehl, and D. A. Landgrebe. 220 Band AVIRIS Hyperspectral Image Data Set: June 12, 1992 Indian Pine Test Site 3, 2015. https://purr.purdue.edu/publications/1947/1 (retrieved: May 1, 2016).

[7] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele. The Cityscapes Dataset for Semantic Urban Scene Understanding. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016. (to appear).

[8] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. ImageNet: A Large-Scale Hierarchical Image Database. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 248–255, 2009.

[9] G. Dial and J. Grodecki. RPC replacement camera models. In *Proceedings of the ASPRS Annual Conference*, pages 1–9, 2005.

[10] H. Hirschmüller. Stereo Processing by Semiglobal Matching and Mutual Information. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 30(2):328–341, 2008.

[11] T. Koch, M. Körner, and F. Fraundorfer. Automatic Alignment of Indoor and Outdoor Building Models using 3D line segments. In *Proceedings of the CVPR Workshop on Visual Analysis of Satellite to Street View Imagery (VASSI)*, 2016. (to appear).

[12] T. Krauß, P. d'Angelo, J. Tian, and P. Reinartz. Automatic DEM Generation and 3D Change Detection from Satellite Imagery. In *European Space Agency Living Planet Symposium*, pages 1–6, 2013.

[13] F. Kurz, O. Meynberg, D. Rosenbaum, S. Türmer, P. Reinartz, and M. Schroeder. Low-cost Optical Camera Systems for Disaster Monitoring. In *XXII ISPRS Congress*, volume XXXIX-B8 of *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, pages 33–37. ISPRS, 2012.

[14] F. Nex, M. Gerke, F. Remondino, H.-J. Przybilla, M. Bäumker, and A. Zurhorst. ISPRS Benchmark for Multi-Platform Photogrammetry. In *PIA15+HRIGI15 – Joint ISPRS conference*, volume II-3/W4 of *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, pages 135–142. ISPRS, 2015.

[15] C. E. Woodcock, R. Allen, M. Anderson, A. Belward, R. Bindschadler, W. Cohen, F. Gao, S. N. Goward, D. Helder, E. Helmer, R. Nemani, L. Oreopoulos, J. Schott, P. S. Thenkabail, E. F. Vermote, J. Vogelmann, M. A. Wulder, and R. Wynne. Free Access to Landsat Imagery. *Science*, 320(5879):1011–1011, 2008.