

Unmanned Aerial Systems for Building Footprint Extraction in Urban Area

Djenaliev, A.,^{1,2*} Chymyrov, A.,² Kada, M.,¹ Hellwich, O.,¹ Akmatov, T.,³ Golev, O.,³ and Chymyrova, S.²

¹Technical University of Berlin, Berlin, Germany

E-mail: djakjol@mailbox.tu-berlin.de, martin.kada@tu-berlin.de, olaf.hellwich@tu-berlin.de

²I.Razzakov Kyrgyz State Technical University, Bishkek, Kyrgyzstan

E-mail: chymyrov@kstu.kg, saltanat.chymyrova@kstu.kg

³State Agency “Cadaster”, Bishkek, Kyrgyzstan

E-mail: talant.akmatov01@gmail.com, golev.o@mail.ru

*Corresponding Author

DOI: <https://doi.org/10.52939/ijg.v20i2.3067>

Abstract

The use of Unmanned Aerial Systems (UAS) in remote sensing applications, specifically using the Trimble UX5 HP platform for aerial data collection over Karakol city in Kyrgyzstan. Photogrammetric technique used to identify and match common features in the overlapping aerial images to create a sparse point cloud, which were further processed to create a Digital Surface Model (DSM). A slope-based filtering algorithm was applied to the DSM data for generating a Digital Terrain Model (DTM). The normalized Digital Surface Model (nDSM) was derived from the DSM by subtracting the DTM. Object-based image analysis applied to UAS datasets for the extraction of building footprints in an urban area. The results indicate that extracted building footprints have been generated accurately with an overall completeness of 92.4% and correctness of 95.2%.

Keywords: Unmanned Aerial Systems, Orthoimage, Digital Surface Model, Building Footprint, Object-Based Image Segmentation, Classification, Evaluation

1. Introduction

Unmanned Aerial Systems (UAS) are known with different names and acronym such as Unmanned Aerial Vehicle (UAV), Remotely-Piloted Aerial Systems (RPAS) or simply called Drones. The widely used term UAS was adopted by the Department of Defense of the USA and the Civil Aviation Authority of the UK. Initially, UAS were born and raised in the military context [1]. The evolution of UAS development in last few years and its transition into civilian applications have increased the number of professional manufacturers in the world. Classifications of UAS are mostly based on the performance characteristics including weight, wing span, speed, flying altitude, operating range, production costs and other capabilities [2].

In recent years, the use of UAS has become increasingly attractive for a wide range of remote sensing applications and aerial surveys. UAS provide rapid deployment and efficient mapping capabilities for urban environments at user-defined spatial and temporal scales [2]. UAS can be equipped with a range of sensors and cameras [1]. Each type of sensor serves a specific purpose and the combination of

different sensors allows for the collection of detailed and accurate data. The integration of Real-Time Kinematic (RTK) technology and dual-frequency Global Navigation Satellite System (GNSS) receivers into UAS plays a significant role in enhancing precise positioning and navigation capabilities [3]. The GNSS receiver in the UAS determines its position using signals from satellites and receives a differential signal from a stationary base station. The RTK–GNSS technology helps for reducing errors caused by atmospheric delay and provides real-time positioning information with centimeter-level accuracy [4]. The use of specialized cameras on UAS significantly improves the capabilities of these systems, allowing for detailed understanding of the environment and to capture different types of data [1].

UAS are well suited for urban applications and the extraction of building footprints. Flying at lower altitudes allows UAS to capture images at very high spatial resolutions up to 0.01 m [5], allowing for the detailed mapping of urban areas and analysis of building structures.



The high level of detail is important for accurate extraction of building footprints and other building information. It provides the precise location and shape of individual buildings within a specific geographic area. Building footprints are fundamental component in the creation of building inventory databases. These databases are valuable for various purposes, including assessing vulnerability to natural disasters such as earthquakes [6]. They help in estimating potential casualties, damage to infrastructure, and economic losses.

High-resolution UAS image allows for the accurate delineation of building footprints and extraction of detailed building dimensions (parameters), including length, width, and area [7]. This data assists in updating building attributes in databases. Integrating building footprints with cadastral data is essential for creating comprehensive databases that include important attributes associated with each building. It helps for the classification of individual buildings based on their usage type, such as residential, commercial and other relevant information. Regularly updated building information are valuable for assessing the vulnerability of buildings to natural disasters [7].

2. Trimble UX5 HP

The application of the fixed-wing UAS platforms have significantly increased in last decades with the advent technologies such as the Trimble UX5 HP aerial imaging system. Trimble Company is a leader in aerial survey innovation and imaging solution [8]. The Trimble's unmanned aerial systems is a new standard in aerial surveying and mapping by combining a robust design and user-friendly system. The main components of this system are the Trimble UX5 HP Aerial Imaging Rover, Digital camera, Ground control station and Launcher [9]. The Trimble UX5 HP Aerial Imaging Rover contains of the several devices. The wing body of the rover is made on a carbon frame structure of foam that has exceptional pressure resistance. The exterior foam reduces the physical damage and protects the internal electronics of an incident [8]. The main removable boxes are the eBox and the gBox devices. The eBox contains the autopilot that controls the Trimble UX5 HP, which is connected to a GPS antenna for navigation and it has connection to the digital camera for sending commands and recording feedback events. The gBox contains a GNSS receiver, which is connected to a GNSS antenna for recording high level positions and it also has connection to the camera for capturing images with precise geodetic coordinates [9]. The Trimble UX5 HP Rover delivers very accurate data by integrating the GNSS receiver

and a superior camera. The camera is capable to capture low altitude images with spatial resolutions down to 1 cm during the flight of the rover. The Sony A7R camera of 36 Megapixel is mounted on board of the UX5 HP Rover that provides sharp, very detailed images and focal length can be fixed with a 15 mm, 25 mm or 35 mm lens. The selection of lens size is related to flying altitude that produces different image resolutions and the area coverages. In case of the same flight altitude at 100 m, where the image resolution is 3.3 cm when selecting the 15 mm lens that covers large area, the resolution is 1.9 cm when using the 25 mm lens and high resolution is 1.4 cm when using the 35 mm lens that offers lower coverage [10].

The desired flight height can be selected for the Trimble UX5 HP Rover, where is capable of flying altitude ranges from 75 m to 750 m above ground level. The UX5 HP Rover is an easy to use, fully autonomous flight and safe landing. It is operated remotely that follows a pre-programmed the flight path, where uses a multi-frequency radio antenna for communication with the receiver on the ground. The ground control station is aimed to monitor, control and command the Trimble UX5 HP Rover from the ground. It applies the Trimble Access™ Aerial Imaging software which is mostly designed for planning aerial missions, automatic performing pre-flight checks, controlling and monitoring of the flight paths, specifying take-off and landing locations [9]. The mechanical launcher provides a safe way to catapult launch the Trimble UX5 Rover in the direction of takeoff. It requires an open space area of approximately 25 m in lengths and 3 m in width for the catapult launching. The recommended landing space area of approximately 50 m lengths and 30 m width is necessary [9].

3. Methodology

3.1 Study Area

Our study focuses on Karakol city, which was formerly known as Przhevalsk and is the administrative capital of Issyk-Kul region in Kyrgyzstan according to Figure 1(a). The city is located about 1750 m above the mean sea level and near the eastern tip of Issyk-Kul Lake, about 150 kilometers from the Kyrgyzstan-China border and 380 kilometers from the national capital Bishkek city. It is the administrative capital of Issyk-Kul Province with 44 square km area [11]. The region of research interest is not included the small town of Pristan-Przhevalsk, which is located adjacent to Issyk-Kul Lake. The city and town are located separately and the distance between them is approximately 12 km.

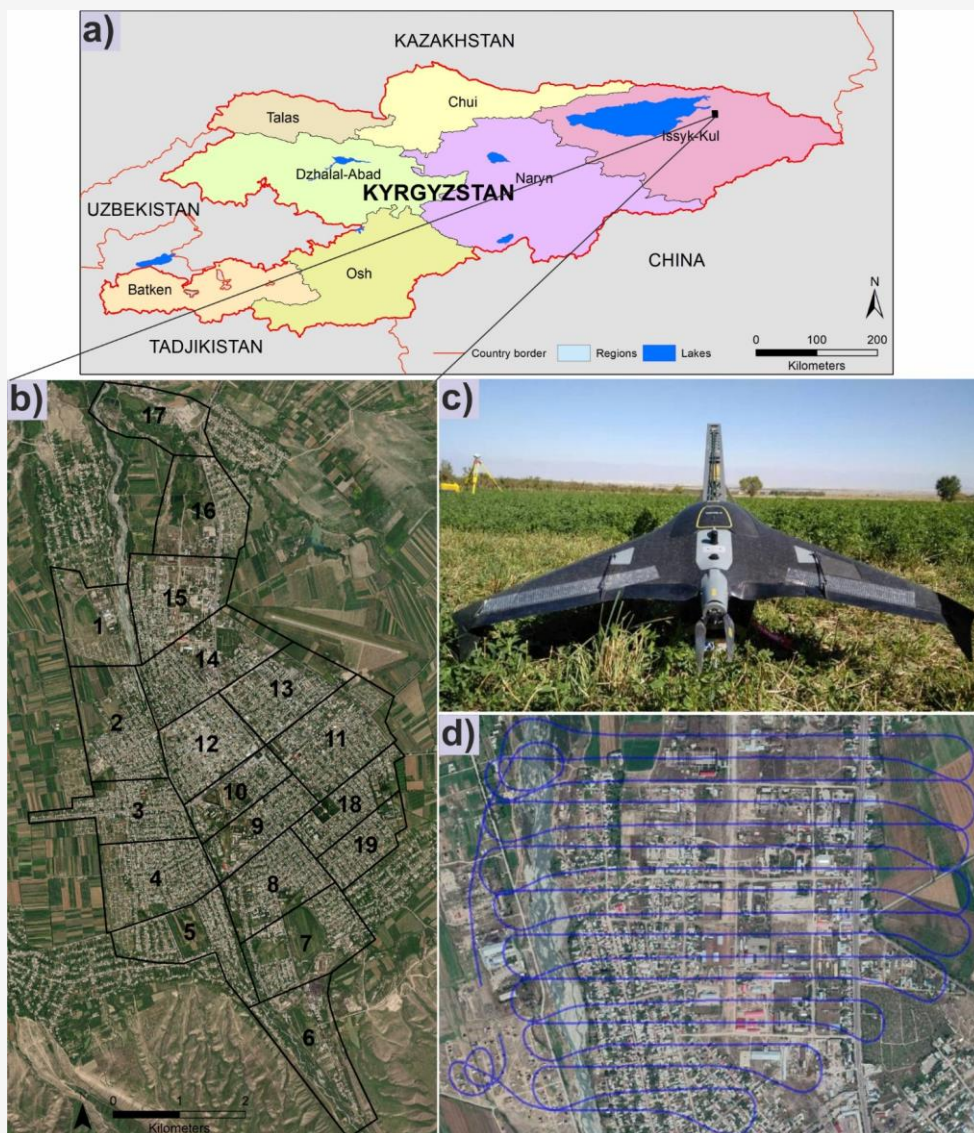


Figure 1: a) Country border, b) UAS flight blocks, c) Pre-flight view and d) the flight trajectory view

Karakol city covered with a mix of residential, commercial, industrial, and agricultural areas, along with efficient transportation infrastructure. Artificially planted trees are deciduous and mostly apple trees. The last two decades farmland and bare lands have been converted to urban uses. Karakol is one of Kyrgyzstan's major tourist destinations, serving as a good starting point for the excellent hiking, trekking, skiing and mountaineering in the high central Tian-Shan to the south and east. Agricultural areas including croplands, apple orchard and pastures are located in the south and west portions of the region. The river of Karakol runs southwest to northwest through the city.

3.2 Aerial Image Acquisition

The State Agency “Cadaster” is the national land property registration service and cadastral mapping agency in the Kyrgyz Republic. The Cadaster Agency has acquired an Unmanned Aerial Systems (UAS), a Trimble UX5 HP for the aerial image data acquisition and precisely delineation of the cadastral borders. The Trimble UX5 HP platform used for aerial surveying for the Karakol city as case study area. For safety purposes danger obstacles such as trees, buildings and engineering structures, electrical pylon and cables in urban areas have to be identified. Therefore, the inspection of the flight area is necessary before starting to use the Trimble UX5 HP aerial imaging rover.

It is very important to select a suitable open flat space area for the catapult launch and safety landing to the ground surface. While the Trimble UX5 HP platform is capable of landing on grass areas without any obstacles, so that the platform body and a digital camera will not be damaged. The Trimble UX5 HP Rover delivers high precision data by integrating 2-frequency Global Navigation Satellite System (GNSS) receiver with Post-Processed Kinematic (PPK) technology that minimizes the need for the ground control points (GCP) [9].

We together with GIS specialists of the Cadaster Agency used Trimble UX5 HP for the low altitude aerial image acquisition over the Karakol city. The aerial survey was performed on 10-20 May 2019. The weather conditions were suitable for the flight, that was without wind and cloudless sky during the time of the aerial surveying. Preliminary, the study area was split up in 19 blocks to be covered the most settlement areas in 19 separate flights. ArcGIS software was used to plan the flight areas and to delineate these blocks as shown in Figure 1(b). The area of each block ranged in size from 90 to 205 ha, where the total aerial surveyed area has 3025 ha. The coordinates for each flight areas were collected with the Trimble R6-4 GNSS receiver and transferred to the ground station operator to generate flight path.

The Trimble Access™ Aerial Imaging software running on the Trimble Tablet rugged PC was used for mission planning and flight monitoring. On each day at most two separate flights occurred between 10:00 and 16:00 local time. After the Trimble UX5 HP rover landed, the battery was replaced, data sets were downloaded, the acquired aerial images and information from the onboard GNSS receiver were checked. The maximum duration of each flight was 35 minutes, with about 800 images being acquired per flight. Ground control point (GCP) markers were designed with a black-yellow cross on the metal. During each flight about 4-6 markers were placed at different locations within the block area at the terrain level and measured using the GNSS receiver. The Trimble R6-4 GNSS receiver was initialized for each flight area as shown in the Figure 1(c), where the geographic coordinates were determined from the Karakol base station within "KYRPOS" network. The distance between the GNSS receiver and the base station were not exceeded 6 kilometers. Each flight lines were completed in the appropriate direction to deliver a frontal and a side overlap of images at 80 %. The flight overview for the Block 15 is presented in the Figure 1(d), where is showing the Trimble UX5 HP rover's detailed trajectory in the West-East direction including all turns from take-off to landing place.

During the aerial survey in total of 7180 images were taken at the altitude of 200 m with a ground sample distance (GSD) equal to 0.06 m covering all flight areas. The aerial images were acquired by the Sony A7R camera of 36 Megapixel resolution with 35 mm focal length, which is mounted onboard of the Trimble UX5 HP. The internal parameters of the Sony A7R camera was manually calibrated, where the ISO sensitivity selected as Auto and the aperture setting was fixed to f/5.6, the lens distortion as polynomial type, the focal length and the position of the principal point were selected as self-adjusted. The acquired digital images are true color RGB bands that saved in JPEG format with radiometric resolution of 8 bits and has an image size of 2048 x 2048 pixels.

4. Data Processing and Results

4.1 Adjusting Orthophotos

The acquired high-resolution images using the Trimble UX5 HP, aerial photo station data, GNSS information, coordinates of the reference station and ground control points (GCP), including an information of the camera sensor calibrations were imported into the Trimble Business Center Photogrammetry (TBC) software for processing data to produce an orthophoto images and point clouds. After importing data sets, all flight missions were merged to a single project data covering the whole study area. It contains of an interconnected series of the aerial images. Each time aerial photo images were captured, where an aerial photo station was created along the flight direction with records of the GNSS positions. It is necessary to adjust the aerial photo stations for delivering final products with the highest precision.

First, the relative adjustment process for the aerial photo stations was performed using the Photo Stations command within the TBC, in which automatically extracted tie points from the aerial images. A tie point is a measurement that represents the same position in the adjacent images. It automatically connects all images to exactly orient the aerial photo stations to each other and georeference them to the ground based on GNSS recorded information when the aerial images were captured [12]. To increase the reliability assessment, where the tie points are located on at least two or more adjacent images can be included for the auto adjustment procedure [13]. However, after the application of the relative photo station adjustment, there may be misalignment in the georeferenced data products. In case of availability of the ground survey data, it is recommended to continue with performing of the absolute adjustment before generating an orthophoto and point cloud data products [12].

The absolute adjustment process was performed immediately after the relative adjustment results in order to reach the highest precision. In the absolute adjustment process, the aerial photo stations were adjusted using the ground control points (GCPs). The coordinates of the GCPs such as the markers on the ground, the corner of foundations, sidewalks and road marking were measured with the Trimble R6-4 GNSS receiver. These ground objects were selected evenly throughout the city and their positions are clearly visible in the aerial images. The aerial photo stations covering the Karakol city were adjusted precisely with ground control points (GCPs), where the Adjust Photo Stations with GCPs command applied within the TBC. The complete flight overview of the Trimble UX5 HP Rover over the city of Karakol is shown in the Figure 2(a).

Aerial survey data were used for adjustment processing of the aerial photos. There are 19 separate flight blocks, where the direction of each flight paths in these blocks can be visually analyzed using spatially distributed green points. These green points indicate the camera locations for 6075 successfully adjusted photos in the project. A total of 19 flight missions were conducted. There are 291 flight strips, and each strip connects photo stations along the trajectory. Grey points in the background represent extracted tie points, which are based on the Structure from Motion algorithm [14]. Tie points have a connection with flight strips and provide stable block connectivity. The study area has a planimetric extent of about 7204 x 12762 m and the height range in the area is approximately 1640-1997 m.

The report statistics show a total of 313,006 tie points, which are classified into four different colors in Figure 2(b). Tie points are reference points identified in aerial images that are used to georeference of the orthophotos. The tie points are

classified into four different categories based on the number of aerial images in which they are found. Each category is represented by a different color: Red points found in less than 2 images; Orange points found in 3-4 images; Green points found in 5-10 images; and blue points found in more than 10 images. Red category points are mainly located at the edges of the blocks and are not considered for the adjustment. Orange, green, and blue points, occurring in more than three images, were included in the adjustment procedure. These points connect multiple strips and provide about 80% strip overlap. The tie point distribution map serves as an indication of block connectivity and strip overlap.

Accuracy of the final adjustment process is estimated by the Sigma naught σ_0 value. The values of standard deviations were computed for all the tie points which were included in the block adjustment process of aerial images. It measures the quality of each single tie point relative to pixel size of digital camera. Table 1 presents the results of adjusting the aerial images in 19 blocks over the study area. Based on the above-mentioned table, where the accuracy value of sigma naught measured of 5.9813 micron. The next accuracy is the mean standard deviation of translations that estimates of the photo positions calculated in the air, where the X, Y, Z coordinates lies within 0.019 - 0.0349 m. Mean standard deviation of rotations estimates the accuracy of the photo rotations calculated in the air, where the angular parameters Omega, Phi, Kappa values are ranged from 4.0034 to 7.4772 degrees. Mean standard deviation of terrain points estimates the accuracy of the tie points calculated on the ground, where the X, Y, Z coordinates estimation are located within 0.0754 - 0.1388 m. Based on the adjustment results in Table 2, it can be concluded that the adjusted aerial images gave a good result.

Table 1: Block adjustment results

1	Accuracy		
	Sigma naught σ_0 (micron)		5.9813
2	Mean standard deviation of translations		
	X (m)	Y (m)	Z (m)
	0.0190	0.0196	0.0349
3	Mean standard deviation of rotations		
	Omega (deg/1000)	Phi (deg/1000)	Kappa (deg/1000)
	7.4772	6.0488	4.0034
4	Mean standard deviation of terrain points		
	X (m)	Y (m)	Z (m)
	0.0754	0.0781	0.1388

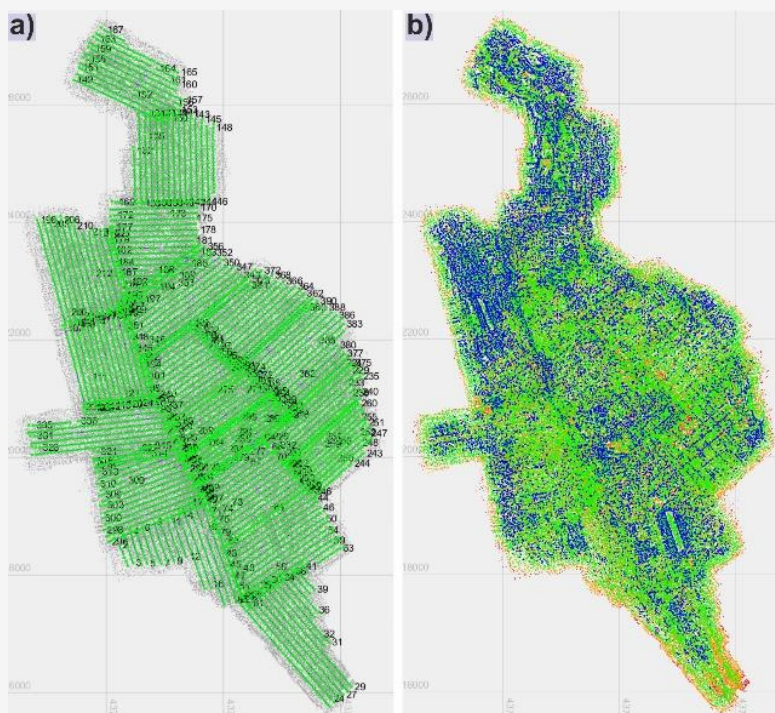


Figure 2: a) The complete flight, b) Tie points

4.2 Creation of an Orthomosaic Image

After completion of the final adjustments of the Trimble UX5 images, the orthomosaic and the point cloud datasets covering the whole study area have been generated in TBC software. The orthorectification process in TBC software is fully automated for the creation of the orthomosaic image. Orthorectification and orthomosaicking are crucial processes in aerial imaging, which are applied to create accurate, georeferenced and high resolution orthomosaic image from multiple aerial photos. The main goal of both orthorectification and orthomosaicking processes are to preserve the same level of quality and spatial accuracy from the original aerial photos [15].

Orthorectification is necessary because aerial photos might come with distortions due to several factors such as ground relief, camera lens and rapid change of lighting conditions [13]. Correcting these distortions is critical for accurate measurements and analysis. It involves geometrically correcting aerial photos to remove distortions caused by ground terrain variations and camera characteristics. The improved images are georeferenced orthophotos, where all objects are positioned accurately on the ground surface. The orthophotos are overlapped with a smaller coverage and are needed to stack for producing a final orthomosaic image [15].

Orthomosaicking process is based on stitching of the individual orthorectified photos by sewing on the overlapped areas of adjacent photos seamlessly and combining them into a single orthomosaic image. During the orthomosaicking process, the software aligns and merges the collection of the orthorectified aerial photos based on their geometric and radiometric properties. The final orthomosaic image was exported as a TIFF file format ensures that the image remains visually continuous and free of geometric distortions [13], making it suitable for detailed spatial analysis and integration with other geospatial datasets.

The created orthomosaic image has 103420 x 194138 pixels with a spatial resolution of 0.06 m by covering 33,4 km² area that represented in Figure 3. As an example, the extended view for small areas represented by two blue boxes, where the distribution of the multi-store buildings in subpart (a) and the single-store buildings in subpart (b) are clearly demonstrated. The coordinate system WGS84 UTM Zone 44N (EPSG:32644) was selected for study area mapping and analysis. The orthomosaic image is a natural true color RGB composite with the combination of visible red, green and blue channels. It ensures that the natural appearance of the landscape is highly interpretable in the orthomosaic image.



Figure 3: Orthomosaic map and the extended views in subpart a) and in subpart b)

4.3 Creation of Point Clouds

A point cloud is a huge collection of data points plotted in a three-dimensional space, where each point is defined by its X, Y, and Z coordinates. These points are a valuable resource for representing and analyzing geographical location [15]. Once individual points come together, it forms a 3D representation of the terrain, natural objects and engineering structures. In addition to its spatial coordinates, each record for point cloud data can store several attribute information such as the point source, GPS time, intensity, spectral information, etc. This information is very useful for distinguishing features, analyzing the point data, and visualizing the landscape in 3D space [16].

The Semi-Global Matching (SGM) algorithm developed by Hirschmuller in 2005 [17] is a powerful method for creating point clouds by automatically matching multiple images to each pixel in the aerial photos. It involves analyzing pairs of aerial photos to identify corresponding points in each image. SGM is known for its accuracy and precision in generating point clouds, which has a wide range of applications including a digital elevation model (DEM) creation and an orthoimage production [18]. The generated point clouds in TBC software were exported in LAS file format, which is associated with Light Detection and Ranging (LIDAR) sensors. It is a common file format used for storing and managing point cloud dataset [16]. Various Geographic Information System (GIS) software tools, including ArcGIS, QGIS, SAGA, and others, support the import and

visualization of LAS files. In ArcGIS software, the LAS Dataset toolset is used for exploring the properties, displaying and visually analyzing of the point clouds as a point and as a surface. The triangulated surface use elevation attributes to provide a continuous representation of the 3D objects in the area.

The LAS file statistics shows that the city area covers of 938,242,899 data points giving the average point spacing of 0,269 m and the point density of 27.4 points per m^2 . Additionally, the point data set has shown a minimum elevation of 1635,43 m and a maximum elevation of 1987,930 m. The density of point cloud is an important indicator, where higher density provides more information while lower density provides less information [15]. It might impact the quality of further data products that depends on point clouds such as using information about elevation. The point clouds are unclassified data sets that contain ground and non-ground objects. Because of the point clouds derived from the UAS orthophotos do not have multiple returns as LIDAR dataset [19].

Due to the large size of LAS file dataset, the smaller areas were extracted for efficient processing, where the LAS dataset 3D view tool used to create a 3D view as given in the Figure 4. Here, point clouds provide a detailed representation of the area, where visual colorations based on elevation values that makes easier to analyze objects. The actual height information can be obtained for each objects by selecting the point clouds.

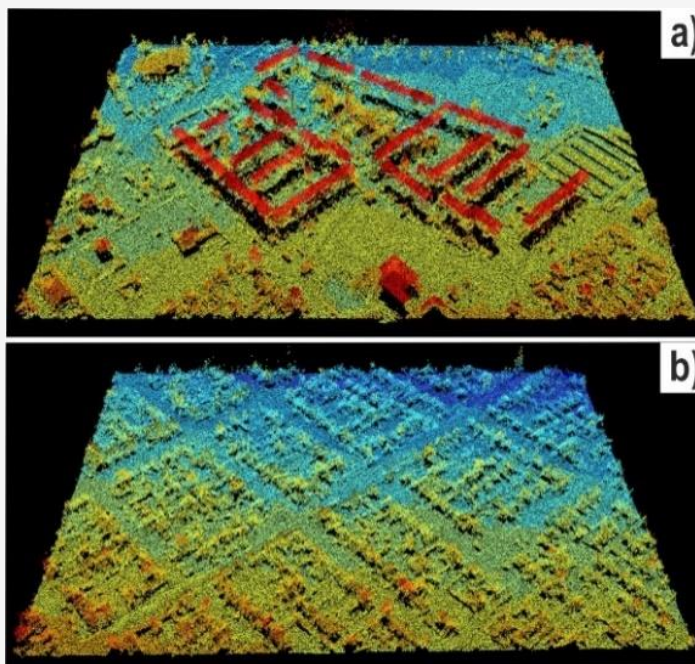


Figure 4: 3D view of the point cloud datasets in subpart a) and in subpart b)

4.4 Creation of a Digital Surface Model

A Digital Surface Model (DSM) is a digital representation of the topographic surface, which includes both natural terrain features and man-made structures located on the ground. It provides height information for all objects that are part of the ground surface and other objects that stand higher than their surroundings [20]. A DSM datasets are mostly generated from a variety of sources such as the airborne laser scanning, stereo processing of aerial or optical satellite imagery and radar interferometry techniques [21]. The traditional photogrammetric methods can extract elevation point values for each image pixel from the aerial photos acquired by an unmanned aerial system (UAS), which is generating a very high-resolution DSM dataset. The use of UAS-based photogrammetric terrain mapping has increased in the last decades [15].

Generally, point clouds and DSM datasets are valuable source for creating 3D representations of the Earth's surface, including trees, buildings and other objects [16]. Due to insufficient overlap of the original point clouds might not cover for some areas that makes appearance of gaps. One of the main advantages of the DSM data generation is filling gaps by interpolation of point cloud datasets [22]. Aerial images often have limitations in their spectral information that can make it challenging to distinguish between the ground surface and man-made structures [20]. The availability of high-quality DSM data plays an important role for building

extraction that provides detailed and accurate elevation information [23]. The main purpose of the DSM generation is creating a detailed 3D model of the visible ground surface with high accuracy by considering terrain discontinuities [24]. The LAS Dataset containing the entire point clouds were converted into a raster layer. This conversion is achieved through a fast-binning process, which involves of point dataset using the Inverse Distance Weighting (IDW) interpolation method for producing a DSM raster layer. The created DSM image is presented in Figure 5, which is strongly correlated with the orthomosaic map. The elevation values on the DSM data ranges from 1635 m to 1988 m above the mean sea level. A visual representation of small areas presented by two blue boxes in Figure 5, where the multi-store buildings in subpart (a) and the single-store buildings in subpart (b) are clearly represented. The separation of buildings from the ground surface and other natural objects are considered in the further image analysis processes.

4.5 Creation of a Digital Terrain Model

A Digital Terrain Model (DTM) is a digital representation of the ground surface, excluding any above-ground objects, while a DSM includes the elevation of all objects on the ground [21] and [25]. In open areas without objects higher than the ground, the elevation values of the DSM and the DTM will be very similar, representing the elevation of the natural terrain [15].

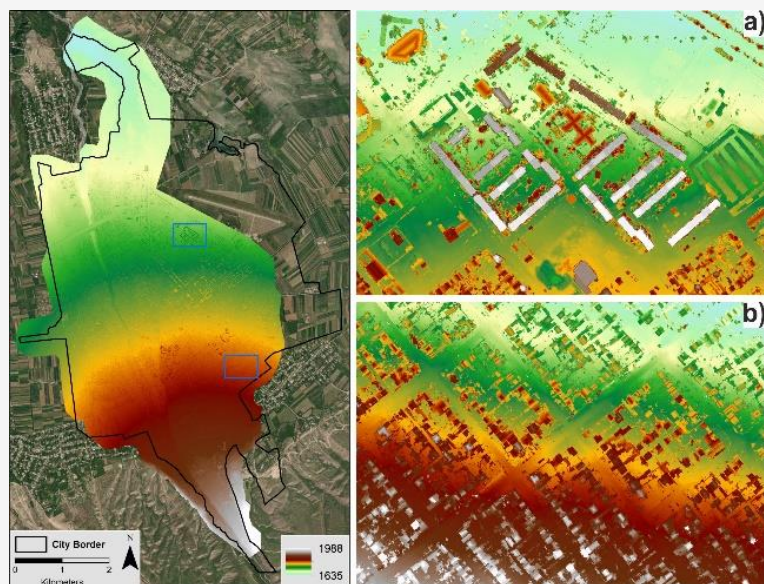


Figure 5: DSM image and the extended views in subpart a) and in subpart b)

A DTM provides highly accurate elevation information exclusively for the bare earth surface [25], which does not include the height of objects such as buildings and trees. Generally, the elevation data is mostly created through remote sensing, including LiDAR and UAV technologies, photogrammetry and GPS ground surveying methods [15]. These technologies capture elevation data points over the study area and are processed to generate a DTM dataset. In the past, several DSM filtering algorithms for DTM generation have been developed. Initially, these algorithms were applied for ground point filtering of the airborne laser scanner data, which is becoming a primary source of DSM data [21]. The algorithms were used to process DSM data and have played a crucial role in the extraction of ground information from the elevation datasets. The filtering algorithms have progressed over time and have been integrated into various Geographic Information System (GIS) software packages. This made it more accessible for GIS researchers to perform terrain analysis and generate DTMs from DSM data.

Morphological filtering methods, including the slope-based filtering that was developed by [26], which are widely used in the remote sensing analysis. It is based on the basic idea that the elevation difference of two adjacent ground points is unlikely to be affected by a steep sloped terrain. More likely, the higher elevation point is not ground point. Thus, for some elevation difference, the possibility that the higher elevation point value might be non-ground point increases when the distance between the two points decreases. Therefore, the slope filter determines a maximum elevation difference between

the two adjacent points as a function of the distance between the points. The set of points are classified as ground if there is no other point within a certain radius to which the elevation difference exceeds the allowed maximum elevation difference for the given distance between these two points [26].

DTM slope-based filter is implemented in SAGA software [27]. This algorithm was used to generate a DTM that represents the bare earth surface by removing non-ground objects from the DSM data. After several attempts, the optimal values for search radius of 40 and terrain slope of 5% were selected for the DSM filtering process. After applying the slope-based filter, the obtained results are two separated raster layers with detection of bare earth surface and the removed non-ground objects.

There is appearance of gaps in the created bare earth data. These gaps occur at the locations, where elevation data values are missing due to the removal of non-ground objects. The size of gaps increases as the search radius of the filter increases. To fill these gaps and create a continuous surface representing of the bare earth is necessary the use of interpolation method. By applying the tool Close Gaps with Spline within SAGA software, we can effectively to fill the gaps areas in the bare earth surface data and producing a continuous DTM surface for further terrain analysis applications. The effectiveness of the DTM slope-based filter may depend on the landscape characteristics of the ground surface and specific type of the elevation data. The created DTM image is presented in Figure 6, where the elevation values on the DTM data ranges from 1635 m to 1984 m above the mean sea level.

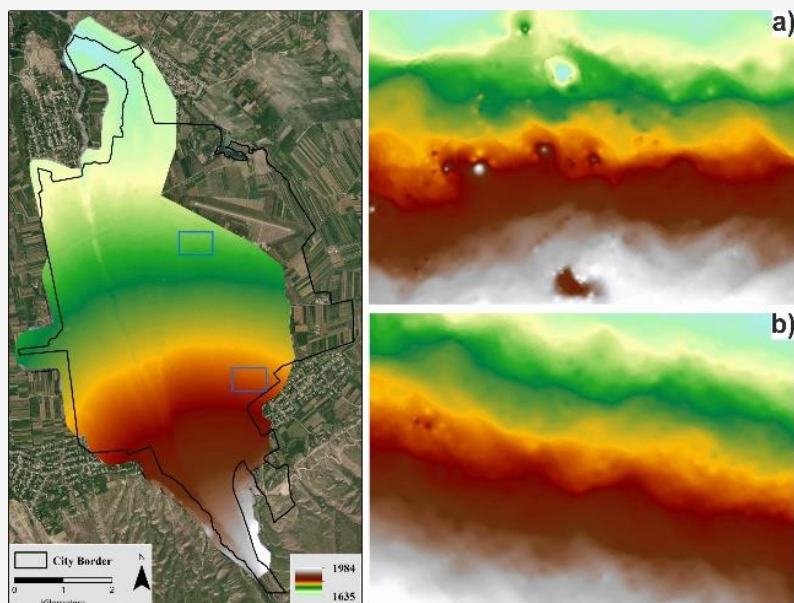


Figure 6: DTM image and the extended views in subpart a) and in subpart b)

4.6 Creation of a Normalized Digital Surface Model

When measuring building height, it usually refers to the vertical distance from the ground level to the highest point of the roof. The measurement of each building height in urban area with traditional field survey can provide detailed information but are very expensive and time consuming [28]. Therefore, a building height information can be obtained from UAS based data by generating a normalized Digital Surface Model (nDSM). The nDSM data is generated by subtracting the DTM data from the DSM data that derived as:

$$\text{nDSM} = \text{DSM} - \text{DTM} \quad \text{Equation 1}$$

This subtraction process effectively removes the elevation of the bare earth from the elevation of all the objects on the ground. The resulting nDSM data represents the relative heights of all objects above the ground surface and clearly illustrates the vertical characteristics of the landscape. The nDSM data provides information about the relative heights of man-made structures and natural features above the ground, including buildings and trees [29]. Generally, it is common to obtain negative height values in the nDSM due to accuracy issues in the DSM and the DTM datasets or interpolation processes [16]. Consequently, we investigated the places with negative height values comparing with the orthomosaic image and found that the most of them are appeared around the river side. These negative values in the nDSM data were replaced by 0

value that considered as zero-elevation ground surface. It was achieved by apply the CON function in the ArcGIS Raster Calculator as follows: $\text{Con}(\text{"nDSM"} < 0, \text{"nDSM"})$. The correction of the height values makes easier to identify ground and to measure building heights while reducing noise from the nDSM dataset. The final nDSM data covering the Karakol city is shown in the Figure 7. The height values of the nDSM ranges from 0 to 56m, where the highest elevation corresponds to the Tian-Shan Spruces. A visual demonstration of small areas presented by two blue boxes in Figure 7, where the multi-store buildings in subpart (a) and the single-store buildings in subpart (b) are clearly represented.

5. Object-Based Image Analysis

5.1 Object-Based Concepts

The advances in Earth observation sensors providing a high-resolution satellite and aerial images have smaller pixel sizes, where each pixel represents a small area on the Earth's surface [30]. When grouped together, these pixels form an image that visually represents real-world objects like buildings, trees, roads, etc. The objects in the remote sensing images are essentially perceived through the grouping of pixels with similar spectral values [31]. Therefore, the traditional pixel-based method analyzes images at the individual pixel level that has limitations, where the neighboring pixels may belong to the same spatial object. This problem is found as 'salt-and-pepper' effect that aimed of grouping pixels into image objects instead of individual pixels [32].

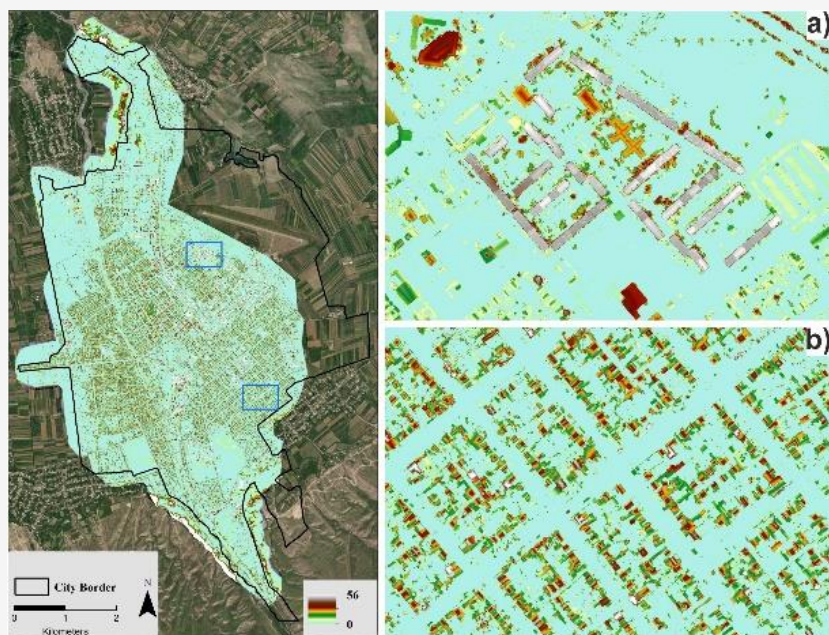


Figure 7: nDSM image and the extended views in subpart a) and in subpart b)

Consequently, it has led to focusing on the application of the object-based image analysis (OBIA) method that allows to use the spectral similarity, shapes, sizes, textures, neighborhood, contextual information and other spatial parameters of the image [30]. The paradigm shifts from analyzing individual pixels to meaningful objects allows for a more contextually and semantically meaningful interpretation of remote sensing imagery [31]. OBIA approach consists of two main steps: 1) image segmentation, 2) feature extraction and classification [33]. The main processing of OBIA are homogenous regions that called image objects [34] or segments as the basis entities for further classification [35]. As human vision generally tends to generalize images into homogenous areas first, that computer vision also processes of recognizing visual information through segmentation and meaningful object extraction [36], which should perfectly coincide with pattern of real-world objects [31]. Image segmentation is a form of partitioning an image into non-overlapping regions that each region is homogeneous [37], and the partitioned regions are representing the meaningful objects based on certain criteria in the image [31]. Segmentation have already been introduced in the mid-1970s [38] for identifying objects in an image processing through the available segmentation techniques [37]. Traditional image segmentation methods are categorized into four main approaches: (i) pixel-based, (ii) edge-based, (iii) region-based and (iv) hybrid methods [32] [33] and [39].

Image segmentation allows representation of image information at different scales are often referred to as multiscale segmentation method. It determines the type of a certain object at different scales, ranging from fine to coarse scales within an image [34]. Between these scales there are spatial and hierarchical dependency, as well as its connection to relationships within segments and classification procedures [40].

The feature extraction and classification in OBIA mostly depend on the quality of image segmentation [33]. The quality of segmentation determines how accurately and meaningfully objects are defined, contributing to the precision of further classification. Object-based classification can be rule-based or machine learning-based and a combination of both approaches. In rule-based classification, human expertise is used to define a set of rules [31] that determine how objects should be classified based on their features. In machine learning-based, algorithms are trained on labeled data to learn patterns and relationships between features and classes. Common machine learning approaches for OBIA include nearest-neighbor, fuzzy logic, and supervised classification methods [30].

eCognition, developed by Definiens AG, was one of the first commercial software packages designed for object-oriented image analysis (OBIA), which became available on the market on 2000 [34]. Alternative to commercial software, a free and open-source software (FOSS) in the field of Geographic Information System (GIS) that has very rapidly developed during the last decades.

The System for Automated Geoscientific Analyses (SAGA) is open-source GIS software for geoscientific data analysis and modeling [27]. The first public release SAGA 1.0 became available in 2004. SAGA utilizes a region growing algorithm for segmentation. The region growing algorithm starts from seed points and expands regions by merging neighboring pixels based on a similarity criterion, which creates homogeneous object primitives or segments [35].

5.2 Image Segmentation Strategies

The image segmentation and building footprint extraction processes were performed in SAGA software. The latest version of SAGA 9.1.0 used for our research work that available to download (<https://sourceforge.net/projects/saga-gis/>). This software offers various tools for remote sensing and geospatial data analysis. SAGA provides the Object Based Image Segmentation tool that uses a Seeded Region Growing (SRG) algorithm. SRG merges neighboring pixels to homogenous features or image objects. It starts from a limited number of single seed pixels, which are defined by its spectral and spatial features [35].

There are several parameters of the Object Based Image Segmentation tool in SAGA. The Band Width for Seed Point Generation is important parameter that quantifies the distance on the both spectral and spatial features in the neighborhood based on the homogeneity criterion. It maximizes the average homogeneity and minimize the heterogeneity of the image segments. According to the size of geographic objects on the image, where the given larger value for this parameter creates large image segment and the given smaller value creates smaller segments.

The neighborhood parameter gives an option to select between the Neumann (4-connected pixels) and the Moore (8-connected pixels) neighborhood types. The distance type allows selection for computation considering only in feature space or the both in feature space and position. It specifies the merging criteria of the adjacent clusters and the influence of spectral features and its spatial position.

Generalization parameter controls the intensity of the smoothing effect that applies a majority filter with defined search radius [35] and [41]. The multi-scale image segmentation is important for building footprint extraction such as single scale will lead to under- or over-segmentation [34]. The limitation of the SAGA that is not generating a hierarchical structure of image segmentation strategies such as the bottom-up and the top-down approaches. These approaches are possible using the multiresolution segmentation in eCognition software [42].

We are interested to apply the concept of the bottom-up approach in SAGA, where the initial segmentation generated small image segments at the fine scale Level-1. The small segments aggregated to larger group of spatial objects at the medium scale Level-2. Based on the multiscale segmentation process, small features are better represented by segments of a lower segmentation level, while some other features are better detected on a higher level. However, it is difficult to define the appropriate scale parameters for image segmentation. Human visual interpretation is the best way to evaluate the results of any image segmentation technique [37]. After apply several trial-and-error procedures in SAGA, the specific parameter values used for segmentation at two levels, where each level corresponds to a different size of resulting image objects from fine to medium detail that given in Table 2.

The orthomosaic image with Red (R), Green (G), Blue (B) bands and DSM data used as input features during the use of the Object Based Image Segmentation tool in SAGA. The images were resampled from the original spatial resolution 0.06 m to 0.3 m for increasing the image processing and saving the time purposes. The use of DSM data increases an accuracy of segmentation instead use only the RGB image. Because the segmentation causes some building roofs in separating from the roads, which have similar spectral properties. The result of segmentation process is a polygonal vector data. Generally, the mean spectral values of each input datasets calculated for all the image segments and stored on the attribute table of the vector data.

Table 2: Parameters of the Segmentation in SAGA

	Level-1	Level-2
Band Width for Seed Point Generation	1	10
Neighborhood	4 (Neuman)	4 (Neuman)
Distance	Feature space and position	Feature space and position
Variance in Feature Space	1	1
Variance in Position Space	1	1
Generalization	1	1

The creation of a hierarchical structure in image segmentation is necessary for representing images at multiple scales of detail. Because images often contain objects and features of varying sizes and complexities. The hierarchical structure allows for the identification of spatial relationships between image segments at different levels, providing variety of information such as spatial, size, spectral, textural and contextual [42]. At the bottom-up approach case, where the outer outlines of the image segments at higher levels are usually determined by those at the lower levels [43]. Unfortunately, the image segmentation procedure in SAGA software not preserving the outer borderlines when producing image segments at different hierarchical levels. Generating a hierarchical network of image segments characterized at both lower and higher levels in two separate steps. The multi-level segmentation strategy was applied, where the first lower Level-1 focuses on generating small segments with detailed spectral and shape information. The specific goal at this level is to separate ground and non-ground image objects based on height information. Therefore, the Grid Statistics for Polygons tool used to compute a Mean elevation value based on the nDSM layer, which represents the relative heights above the ground surface.

The statistical approach summarizes the height information values found within the image segments. All the polygonal image segments on the lower level were updated using elevation values. The image segments meeting the criteria based on elevation statistic values, where the mean $nDSM \geq 2.0m$ were

selected as the elevated objects. The selected segments, representing elevated objects such as buildings or a part of a buildings including trees were exported as a new vector layer. This new layer contains information about the spatial extent, the spectral and height characteristics of these elevated objects.

The second image segmentation process applied to generate medium sized image segments at the higher Level-2. Compare to lower Level-1, the increased number for the distance between seed points would result the larger segments and it led to a loss of finer details in the segmentation. In homogeneous areas, the segmentation groups a larger number of similar pixels into bigger segments. In areas with high heterogeneity creates smaller segments by preserving finer details and delineating the boundaries of different objects. The certain types of building roofs that may have similar spectral and shape characteristics to roads and being over-segmented. Therefore, the created new layer representing elevated objects is based on the lower Level-1 was overlaid with image segments from the higher Level-2 using the Union operator. It combines the geometries of the two layers, creating a new layer that represents the spatial union of features from both Level-1 and Level-2. Common border lines between larger and smaller segments are preserved in the resulting layer. Figure 8 shows the segmentation results at Level-1 and Level-2, where the elevated objects highlighted in red to make them visually distinguishable from other segments.

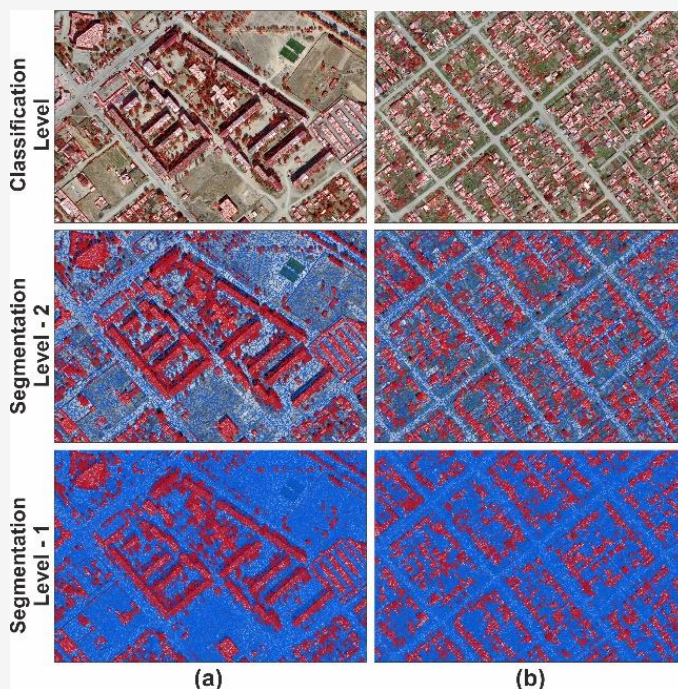


Figure 8: Image segmentation views in subpart (a) and in subpart (b)

Spectral and elevation values from both segmented layers are combined and stored in the attribute table of the newly created layer. Segments without elevation values that usually stored as No Data with the value of “-99999” in the attribute table are identified and classified as ground objects. Remaining segments with elevation values represent elevated objects, which are the main focus of the research. Both segmented layers' ID-fields are stored in the same attribute table to identify larger and smaller segments. Larger segments, which contain several smaller segments, need to be merged while preserving outer borders. The Dissolve operator is used to aggregate larger image segments based on specified ID-field attributes into a new single layer. The aggregated segments may include summaries such as the mean and the standard deviation of attribute values from the input layer. The resulting layer from the Dissolve operation is considered for the next classification process. The results obtained from multi-level image segmentation strategies on Level-1 and Level-2, and particularly focusing of the elevated image segments for classification are given at Classification Level in Figure 8.

5.3 Classification of Image Segments

Using the Grid Statistics for Polygons, the spectral values of the RGB image, DSM and nDSM datasets were computed for the elevated image segments using grid statistics. In addition, the polygonal shape indices for segments were calculated, which is providing information about the geometric properties of objects within an image. By extracting the shape indices such as the area, the perimeter, the ratio between perimeter and area (P/A ratio), we can obtain geometric characteristics of segments [5]. It provides information about the shape and size of the objects in an image.

The manually labeled image segments were used as training data for classifying the elevated segments into different categories. The training segments were selected randomly to classify buildings and not buildings such as trees and shadows across the entire study area. The class identifier codes were assigned to each class in the attribute table. This table contains information about the assigned class for labeled segments. Empty fields in the table indicate instances where training data were not specified. The attribute properties of the selected segments, especially pixel values can provide a detailed characteristic of the image datasets. By examining the distribution and range of pixel values for each class, we can identify key features that distinguish one class from another.

Image segments classification was implemented using the Supervised Classification (Shapes) tool.

This classification involves classifying image segments are based on the values of the mean and variance of input features. The computed features include Red, Green, Blue, nDSM and the P/A shape indices, which are describing the spectral, height and shape characteristics of each segment. The classification process involves training a model using class identifier field and selecting the Maximum Likelihood classifier to classify image segments based on the statistical distribution of 10 input features. The class identifier field contains labels for different classes that the classifier will learn to recognize classes. Maximum Likelihood classification is a statistical approach that assigns each segment to the class with the highest probability based on the statistical distribution of the input features. It assumes that the input features follow a certain statistical distribution for each class [15]. The classification result of image segments representing building footprints are given in the Figure 9, where two blue boxes are present. The first box in subpart a) appears to represent multi-store buildings, while the second box in subpart b) representing single-store building footprints.

5.4 Evaluation of Building Footprints

The process of object-based evaluation was performed by comparing the classification results of the building footprints to reference dataset. The evaluation provides a quantitative measure of how well the classified building footprint match the reference building. A number of research papers use the completeness and the correctness in object-based evaluation of classification results, which can be calculated as [44]:

$$\text{Completeness} = \frac{TP}{TP + FN} \quad \text{Equation 2}$$

$$\text{Correctness} = \frac{TP}{TP + FP} \quad \text{Equation 3}$$

In Equations (2) and (3), TP is representing the number of true positives, i.e., the number of image objects correctly classified as building footprints and also corresponds to buildings in the reference dataset. FN is representing the number of false negatives, i.e., the number of buildings in the reference data that were not correspond to classified building footprints. FP is denoting the number of false positives, i.e., the number of image objects that were classified as building footprints but do not correspond to buildings in the reference dataset [44].

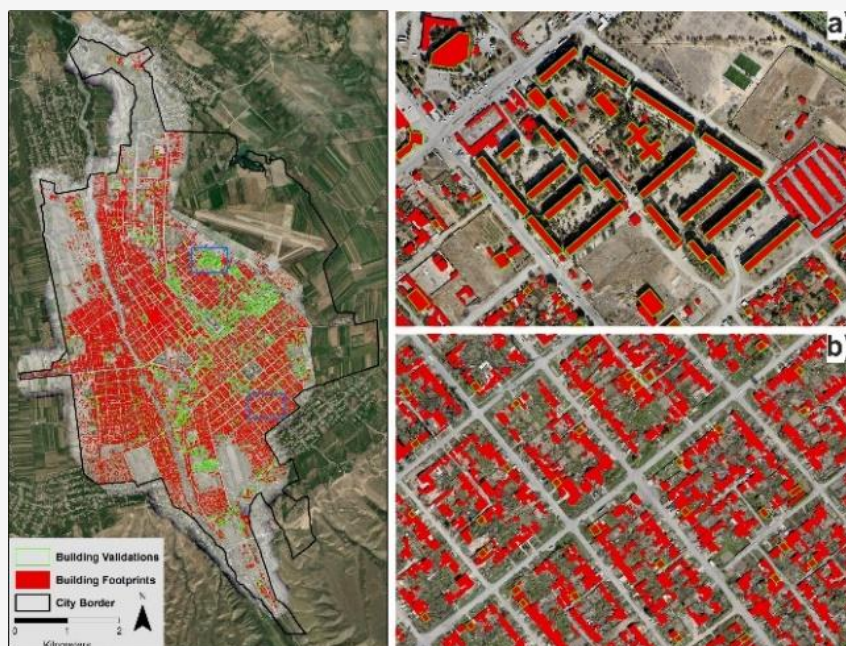


Figure 9: Building footprints and object-based validation views in subpart a) and in subpart b)

Completeness, often referred to as Producer's accuracy [45], which is related to the percentage of features in the reference data that were correctly detected on the classification map [46]. Correctness, often referred to as User's accuracy [45], that indicates how well the classified features match the reference datasets [46]. In object-based evaluation, the most criterion is based on substantial overlap between the classified building footprints and the reference buildings [45]. However, the topology of the reference data may not match with the classification results, particularly in densely built-up areas. Therefore, the topologies of the both datasets were spatially adjusted. Building footprints composed of multiple segments were merged to better represent individual buildings.

The reference dataset includes 1400 manually digitized buildings across the study area is given Figure 9. These buildings represent a variety of usage types, including single and multi-residential, commercial, administrative, etc. The building areas range from 50 to 4000 square meters and the mean area of the buildings in the dataset is 235m². The reference buildings were categorized into several classes based on their area parameters. The classification resulted building footprints were also categorized into the same classes as those used in the reference dataset.

Completeness is assessed separately for each category of the building in the reference dataset [45]. The categorization is based on the characteristics of the building area parameters. If the centroid of a

polygonal building in the reference data falls within the boundaries of the classified building footprint, the number of determined and not determined buildings are counted. The determined reference building is considered as a True Positive (TP) and other not determined building is considered as a False Negative (FN). Completeness of the dataset is then calculated based on the counts of TP and FN. Correctness is calculated for each category of the classified building footprints in the dataset [45]. Area of the building footprint is taken into consideration during the assessment. A classified building footprint whose centroid falls within the boundaries of the reference building is considered as a True Positive (TP). And other building footprint that is not defined in the reference dataset is considered as a False Positive (FP). Correctness of the dataset is then calculated based on the counts of TP and FP.

The object-based evaluation results for building footprints indicates with an overall completeness of 92.4% and correctness of 95.2%. Based on the results we can conclude that extracted building footprints have been generated accurately.

6. Conclusions

The use of Unmanned Aerial Systems (UAS) in remote sensing applications, specifically using the Trimble UX5 HP platform for aerial data collection over Karakol city in Kyrgyzstan. Aerial surveys resulted in the capture of 7180 images at an altitude of 200 meters with a very high ground sample distance (GSD) of 0.06 meters.

The acquired aerial images were adjusted precisely using the GNSS receiver data and ground control points. Photogrammetric technique used to identify and match common features in the overlapping aerial images to create an orthomosaic image and a point cloud. The point cloud datasets were interpolated to create a Digital Surface Model (DSM). A slope-based filtering algorithm was applied to the DSM data for generating a Digital Terrain Model (DTM). The normalized Digital Surface Model (nDSM) was derived from the DSM by subtracting the DTM, which represents the height information of above-ground features, such as buildings and vegetation.

Object-based image analysis (OBIA) approach applied to UAS data analysis that consisted of the segmentation and classification steps to correctly extract building footprints in an urban area. Combining orthomosaic image with DSM data in the segmentation process contributed to accurately delineating image objects based on spectral, spatial and height information characteristics. Image segmentation was conducted at two different levels, starting with finer details and progressing to larger scale features. The first level focused to create smaller segments for a detailed analysis. The image segments with a mean nDSM value greater than 2.0 meters were selected as the elevated image objects. The second level aimed to create larger segments, grouping together regions with similar characteristics.

Building footprint classification was performed using the supervised classification method. The spectral values of RGB image, relative height and geometric shape characteristics were computed for each larger segments. The manually labeled image segments were used as training data to train the classifier. The Maximum Likelihood classifier used to classify image segments based on statistical distribution of input features. The final classified map containing the accurate footprints of buildings within the urban area. The acquired UAS datasets and the corresponding building footprints are highly valuable for further vulnerability assessments of urban areas to natural disasters.

Acknowledgment

We are acknowledged for the Erasmus+ GeoTak Project (617695-EPP-1-2020-1-ES-EPPKA2-CBHE-JP) for the provided financial support for publication of these research outcomes.

References

- [1] Colomina, I. and Molina, P., (2014). Unmanned Aerial Systems for Photogrammetry and Remote Sensing: A Review. *ISPRS J Photogramm Remote Sens*, Vol. 92, 79–97. <https://doi.org/10.1016/j.isprsjprs.2014.02.013>.
- [2] Noor, N. M., Abdullah, A. and Hashim, M., (2018). Remote Sensing UAV/Drones and its Applications for Urban Areas: A Review. *IOP Conf. Ser.: Earth Environ. Sci.*, Vol. 169, 1-8. <https://doi.org/10.1088/1755-1315/169/1/012003>.
- [3] Gerke, M. and Przybilla, H. J., (2016). Accuracy Analysis of Photogrammetric UAV Image Blocks: Influence of Onboard RTK-GNSS and Cross Flight Patterns. *Photogrammetrie-Fernerkundung-Geoinformation*, 17–30. <https://doi.org/10.1127/pfg/2016/0284>
- [4] Mabdeh, A. N., Al-Fugara, A. and Jarah, M., (2018). Object-Based Classification of Urban Distinct Sub-Elements Using High Spatial Resolution Orthoimages and DSM Layers. *Journal of Geographic Information System*, Vol. 10, 323–343. <https://doi.org/10.4236/jgis.2018.104017>.
- [5] Mattivi, P., Pappalardo S. E., Nikolić, N., Mandolesi, L., Persichetti, A., De Marchi, M. and Masin, R., (2021). Can Commercial Low-Cost Drones and Open-Source GIS Technologies be Suitable for Semi-Automatic Weed Mapping for Smart Farming? A Case Study in NE Italy. *Remote Sens.*, Vol. 13, <https://doi.org/10.3390/rs13101869>.
- [6] Djenaliev, A., Kada, M. and Chymyrov, A., (2016). Building Inventory Data Development for Pre-Earthquake Evaluation. *International Journal of Geoinformatics*, Vol. 12(4), 41–47. <https://journals.sfu.ca/ijg/index.php/journal/article/view/990>.
- [7] Angela, B. V., Norbert, H. and Jochen, S., (2013). Building Extraction from Remote Sensing Data for Parameterising a Building Typology: A Contribution to Flood Vulnerability Assessment. *Joint Urban Remote Sensing Event, Brazil*, 147–150. <https://doi.org/10.1109/JURSE.2013.6550687>.
- [8] Cosyn, P. and Miller, R., (2013). *White Paper. Trimble UX5 Aerial Imaging Solution*. A New Standard in Accuracy, Robustness and Performance. Trimble Survey, Westminster, CO, USA. 1–7.

- [9] Trimble, (2016). *User Guide*. Trimble UX5 HP Aerial Imaging Solution. 1–141.
- [10] Pauly, K., (2016). *White Paper*. Trimble UX5 HP – Increasing Your Productivity, 1–11.
- [11] Abdykalykov, A., Baidjumanov, D., Osmonaliev, A., Tulegabylov, N. and Kim, A., (2010). *National Statistical Committee of the Kyrgyz Republic. Population and Housing Fund*, Bishkek, Kyrgyzstan.
- [12] Trimble, (2022). Trimble Business Center. Processing Aerial Survey Data with a Trajectory, 1-28.
- [13] Wierzbicki, D., Kedzierski, M. and Fryskowska, A.. (2015). Assessment of the Influence of UAV Image Quality on the Orthophoto Production. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XL-1/W4, 1–8. <http://doi.org/10.5194/isprs-archives-XL-1-W4-1-2015>.
- [14] Westoby, M. J., Brasington, J., Glasser, N. F., Hambrey, M. J. and Reynolds, J. M. (2012). Structure-from-Motion Photogrammetry: A Low-Cost, Effective Tool for Geoscience Applications. *Geomorphology*, Vol. 179, 300–314. <http://doi.org/10.1016/j.geomorph.2012.08.021>.
- [15] Green, K., Congalton, R. G. and Tukman, M., (2017). *Imagery and GIS: Best Practices for Extracting Information from Imagery*, ESRI Press, California, 1-418.
- [16] Dong, P. and Chen, Q., (2018). *LiDAR Remote Sensing and Applications (1st ed.)*, CRC Press, 1-220. <http://doi.org/10.4324/9781351233354>.
- [17] Hirschmüller, H., (2005). Accurate and Efficient Stereo Processing by Semiglobal Matching and Mutual Information. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), San Diego, USA*, Vol. 2, 807-814. <http://doi.org/10.1109/CVPR.2005.56>.
- [18] Konecny, G., (2014). *Geoinformation Remote Sensing, Photogrammetry, and Geographic Information Systems*, Second Edition, CRC Press, 1-472. <http://doi.org/10.1201/b15765>.
- [19] Kulhavy, D., Hung, I. K., Unger, D. R., Viegut, R. and Zhang, Y., (2021). Measuring Building Height Using Point Cloud Data Derived from Unmanned Aerial System Imagery in an Undergraduate Geospatial Science Course. *Higher Education Studies*, Vol.11(1). <http://doi.org/10.5539/hes.v11n1p105>.
- [20] Haala, N. and Brenner, C., (1999). Extraction of Buildings and Trees in Urban Environments. *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 54(2-3), 130–137. [http://doi.org/10.1016/S0924-2716\(99\)00010-6](http://doi.org/10.1016/S0924-2716(99)00010-6).
- [21] Krauß, T., Arefi, H. and Reinartz, P., (2011). Evaluation of Selected Methods for Extracting Digital Terrain Models from Satellite Born Digital Surface Models in Urban Areas. *SMPR 2011*, 1-7, Tehran, Iran. <https://core.ac.uk/download/pdf/11148575.pdf>.
- [22] Asal, F., (2019). Comparative Analysis of the Digital Terrain Models Extracted from Airborne LiDAR Point Clouds Using Different Filtering Approaches in Residential Landscapes. *Advances in Remote Sensing*, Vol. 8, 51–75. <http://doi.org/10.4236/ars.2019.82004>.
- [23] Weidner, U., (1997). *Digital Surface Models for Building Extraction*. In: Gruen, A., Baltsavias, E.P., Henricsson, O. (eds) Automatic Extraction of Man-Made Objects from Aerial and Space Images (II). Birkhäuser Basel, 193–202. http://doi.org/10.1007/978-3-0348-8906-3_19.
- [24] Baltsavias, E., Mason, S., Stallmann, D. (1994). Use of DTMs/DSMs and Orthoimages to Support Building Extraction. In: Gruen, A., Baltsavias, E.P., Henricsson, O. (eds) Automatic Extraction of Man-Made Objects from Aerial and Space Images. Birkhäuser Basel, 199-210. http://doi.org/10.1007/978-3-0348-9242-1_19.
- [25] Chen, Z., Gao, B. and Devereux, B., (2017). *State-of-the-Art: DTM Generation Using Airborne LIDAR Data*. *Sensors*, Vol 17, <http://doi.org/10.3390/s17010150>.
- [26] Vosselman, G., (2000). Slope Based Filtering of Laser Altimetry Data. *IAPRS*, XXXIII, 935-942.
- [27] Conrad, O., Bechtel, B., Bock, M., Dietrich, H., Fischer, E., Gerlitz, L., Wehberg, J., Wichmann, V. and Böhner, J., (2015). System for Automated Geoscientific Analyses (SAGA) v. 2.1.4. *Geosci. Model Dev.*, Vol. 8, 1991–2007. <http://doi.org/10.5194/gmd-8-1991-2015>
- [28] Zhou, J., Liu, Y., Nie, G., Cheng, H., Yang, X., Chen, X. and Gross, L., (2022). Building Extraction and Floor Area Estimation at the Village Level in Rural China Via a Comprehensive Method Integrating UAV Photogrammetry and the Novel EDSANet. *Remote Sens.* Vol. 14. <http://doi.org/10.3390/rs14205175>.

- [29] Yu, B., Liu, H., Wu, J., Hu, Y. and Zhang, L., (2010). Automated Derivation of Urban Building Density Information Using Airborne LiDAR Data and Object-Based Method. *Landscape and Urban Planning*, Vol. 98(3), 210–219. <http://doi.org/10.1016/j.landurbplan.2010.08.004>.
- [30] Campbell, J. B. and Wynne, R. H., (2011). *Introduction to Remote Sensing*. 5th Edition, The Guilford Press, 1-667.
- [31] Blaschke, T., Hay, G. J., Kelly, M., Lang, S., Hofmann, P., Addink, E., Feitosa R. Q., Meer, F., Werff, H., Coillie, F. and Tiede, D., (2014). Geographic Object-Based Image Analysis - Towards a New Paradigm. *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 87, 180–191. <http://doi.org/10.1016/j.isprsjprs.2013.09.014>.
- [32] Blaschke, T., (2010). Object Based Image Analysis for Remote Sensing. *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 65(1), 2–16. <http://doi.org/10.1016/j.isprsjprs.2009.06.004>.
- [33] Hossain, M. D. and Chen, D., (2019). Segmentation for Object-Based Image Analysis (OBIA): A Review of Algorithms and Challenges from Remote Sensing Perspective. *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 150, 115–134. <http://doi.org/10.1016/j.isprsjprs.2019.02.009>.
- [34] Benz, U. C., Hofmann, P., Willhauck, G., Lingenfelder, I. and Heynen, M., (2004). Multi-Resolution, Object-Oriented Fuzzy Analysis of Remote Sensing Data for GIS-ready Information. *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 58, 239–258. <http://doi.org/10.1016/j.isprsjprs.2003.10.002>.
- [35] Böhner, J., Selige, T. and Ringeler, A., (2006). *Image Segmentation Using Representativeness Analysis and Region Growing*. In: Böhner, J., McCloy, K., Strobl, J. (Eds.), *SAGA - Analyses and Modelling Applications*. Göttinger Geographische Abhandlungen, Vol. 115, 29-38.
- [36] Blaschke, T. and Strobl, J., (2001). What's Wrong with Pixels? Some Recent Developments Interfacing Remote Sensing and GIS. *GIS Zeitschrift für Geoinformationssysteme*, Vol. 14 (6), 12-17.
- [37] Pal, N. R. and Pal, S. K., (1993.) A Review on Image Segmentation Techniques. *Pattern Recognition*, Vol. 26(9), 1277–1294. [http://doi.org/10.1016/0031-3203\(93\)90135-J](http://doi.org/10.1016/0031-3203(93)90135-J)
- [38] Haralick, R. M., Shanmugam, K. and Dinstein, I., (1973). Textural Features for Image Classification. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-3, Vol. 6, 610–621. <http://doi.org/10.1109/TSMC.1973.4309314>.
- [39] Adams, R. and Bischof, L., (1994). Seeded Region Growing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 16(6), 641-647. <http://doi.org/10.1109/34.295913>.
- [40] Lang, S., (2008). *Object-Based Image Analysis for Remote Sensing Applications: Modeling Reality – Dealing with Complexity*. Object-Based Image Analysis, Berlin, 3–27. https://doi.org/10.1007/978-3-540-77058-9_1.
- [41] Bechtel, B., Ringeler, A. and Böhner, J., (2008). Segmentation for Object Extraction of Trees using MATLAB and SAGA. *SAGA – Seconds Out, Hamburger Beiträge Zur Physischen Geographie Und Landschaftsökologie*, 1–12.
- [42] Baatz, M. and Schape, A., (2000). Multiresolution Segmentation: An Optimized Approach for High Quality Multi-Scale Image Segmentation. In: Strobl, J., Blaschke, T. and Griesbner, G., Eds., *Angewandte Geographische Informations-Verarbeitung*, XII, 12-23.
- [43] Hofmann, P., (2001). Detecting Urban Features from IKONOS Data using an Object-Oriented Approach. *RSPS2001 Proceedings*, 79–91. <https://pennur.upenn.edu/uploads/media/Hoffman.pdf>.
- [44] Rottensteiner, F., Trinder, J., Clode, S. and Kubik, K., (2005). Using the Dempster–Shafer Method for the Fusion of LIDAR Data and Multi-Spectral Images for Building Detection. *Information Fusion*, Vol. 6, 283–300. <http://doi.org/10.1016/j.inffus.2004.06.004>.
- [45] Rutzinger, M., Rottensteiner, F. and Pfeifer, N., (2009). A Comparison of Evaluation Techniques for Building Extraction from Airborne Laser Scanning. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, Vol. 2, 11–20. <http://doi.org/10.1109/jstars.2009.2012488>.
- [46] Story, M. and Congalton, R. G., (1986). Accuracy Assessment: A User's Perspective. *Photogrammetric Engineering and Remote Sensing*, Vol. 52, 397–399. https://www.asprs.org/wp-content/uploads/pers/1986journal/mar/1986_mar_397-399.pdf.