Master Thesis

submitted within the UNIGIS MSc programme
Interfaculty Department of Geoinformatics - Z_GIS
University of Salzburg

Evaluation of the 'Regularize Building Footprint' function for the generalization of refugee dwellings based on very high-resolution satellite imagery

by

Franziska Halbritter

106926

A thesis submitted in partial fulfilment of the requirements of the degree of

Master of Science – MSc

Advisor:

Assoz. Prof. Dr. Stefan Lang

Ingolstadt, 25 September 2023

Science Pledge

I hereby declare that this thesis is entirely the result of my own work. I have cited all sources I have used in this thesis; I have always indicated their origin. This thesis was not previously presented to another examination board and has not been published.

Ingolstadt , den 25. Sep. 2023 Franziska Halbritter

Acknowledgements

First, I'd like to say thank you to the entire UNIGIS Team for all the support and positive interaction during my studies. Especially to Prof. Dr. Lang, who always found time in his busy schedule for discussions and made this work possible in the first place.

A big thank you goes to my family. They have always supported me and believed in me. Finally, a special thank you to my boyfriend. He always cheered me up with loving words. Thank you for your patience, for proofreading the thesis and for your suggestions throughout the entire process.

Abstract

Because of natural disasters, wars and violence, more and more people are forced to leave their home. These refugees or internal displaced persons usually find protection in refugee camps. The access to these camps is often difficult for aid organizations because of violent conflicts or difficult local conditions. It is therefore complicated for them to estimate how many people are living there to be able to provide targeted humanitarian aid.

Earth observation and image analysis systems have rapidly changed in the last few years. These improvements offer new opportunities to identify information about the infrastructure of the camp or the number of living people in the camps. New approaches can be used to estimate the population from a building outline from a high-resolution satellite image. However, these new technology improvements result in further challenges. For example, in the case of extracted building outlines, the lines contain many redundant vertices and artefacts because of the high resolution. These influence the shape and size of the objects and represent not the real building outlines. However, to provide the organizations reliable information about the camp, the improvement of the cartographic representation of the outlines has to be explored.

This work is therefore about the generalization of extracted outlines of refugee camp shelters from very high-resolution satellite imagery. The aim is the improvement of the outlines for the representation in cartographic maps. The function 'Regularize Building Footprint' from ESRI provides a way to generalize buildings in urban areas. Whether the function also offers an alternative in the context of refugee accommodation is investigated in this study.

The results show a good way to represent large and rectangular accommodations applying the generalization algorithm. For more complex structures, the method is less suitable. It also shows that the parameters used for generalization depend on the size and shape of the extracted outlines.

Abbreviations

AA .... *ANY_ANGLE*

C .... *CIRCLE*

CDL .... *Christian Doppler Laboratory*

CNN... *Convolutional Neural Networks*

D .... *densification*

DL .... *Deep Learning*

DP .... *diagonal penalty*

EO .... *Earth Observation*

GEOHUM... *Geo- and EO-based Humanitarian Technologies*

IDPs .... *internal displaced persons*

MFS .... *Médecins Sans Frontières*

OBIA .... *Object-Based Image Analyzes*

PEAK.. *polynomial approximation with exponential kernel*

RA.. *RIGHT_ANGLE*

RAaD .... *RIGHT_ANGLE_AND_DIAGONALS*

RBF.. *Regularize Building Footprint*

T .... *tolerance*

UNHCR .... *United Nations High Commissioner for Refugees*

VHSR .... *Very High Spatial Resolution*

List of figures

List of tables and charts

Table of Contents

1. Introduction

1.1. Problem statement and research question

During the extraction of building outlines from satellite imagery several factors can influence the accurate identification of the objects. One of the main problems is that the reconstructed lines are generated on the grid of the images. The consequence is, that the new polylines have many unnecessary vertices, and the outlines are inaccurate. If the earth observation system produces very high-resolution images, buildings can be described more precisely. This leads on the other side into a growing number of redundant points and "ragged" polylines (cf. Figure 1). This aspect can show wrong representations of the dwellings.

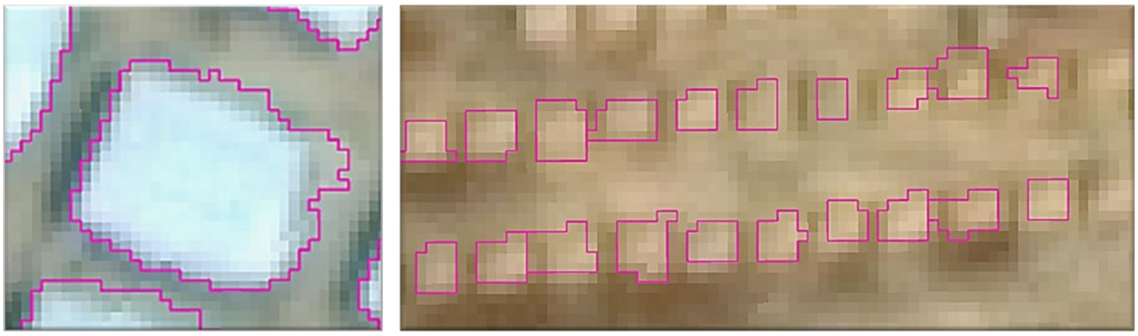


Figure 1: “Ragged” polygons based on raster images.

Another challenge during the extraction is the similar spectral reflectance of pixels from nearby objects. For example, the reflection properties of an object can be affected by solar radiation or shading. Through these pixel values the shadow of a dwelling can appear like the building and is difficult to separate. Another effect appears because of the dry climate in the region of some refugee camps. Dust can collect on the roof of the buildings, or people cover them with blankets and straw to protect themselves from the heat. These factors can also lead to a similar reflection of the roof and the ground and therefore to a wrong classification of the pixels (cf. Figure 1 (right images)). Thus, vegetation can also lead to wrong outlines or artefacts. In this case, the area of a tent can be covered by the crown of a tree. As a result, this part of the building cannot be completely identified on the image (Gao et al., 2022, Tiede et al., 2017).

A further problem, especially for refugee dwellings, is the size of the buildings. In comparison to buildings in urban areas some of them are very small and have only an area of 1 m². For those tiny buildings it is difficult to determine the accurate shape in combination with misleading spectral properties and a not optimal resolution of the image. Other buildings like the Facility Buildings can be identified easily. They are big enough (over 90 m²) and have good spectral reflection to separate them.

All these factors can result that the buildings are not clearly identified from their surroundings (Füreder et al., 2014). Therefore, a method must be found to generalize these outlines to represent them more accurately in maps.

There are a variety of methods to improve the extraction of information from an Earth observation image. Meanwhile also a wide range of methods for solving generalization problems is available. However, every algorithm follows different goals and therefore uses other mathematical functions. For example, Douglas and Peucker (1973) describe in their studies, how they remove unnecessary vertices of a line. At the end they create a new polyline that has significant fewer vertices but keep the shape of the line. This method is often used to generalize objects for different map scales and for data compression. Other methods have the special aim to simplify urban buildings. For this purpose, Sester investigates in her studies the least square adjustment theory (Sester, 2001). Alexander Gribov has developed a complex algorithm which is looking for a line that has the smallest possible sum of vertices, but also represents the original shape. The function contains additionally geometrical conditions like right angles and parallel segments. However, the new line has to be within a certain tolerance of the original line (Gribov, 2018).

So far, these methods have only been applied to urban buildings. It should therefore be investigated whether these methods can also be applied for generalization of refugee accommodation with their specific characteristics.

The method from Gribov's offers a good way to normalize building footprints from aerial or drone images. The algorithm is integrated as an extension tool in ArcGIS, which is called 'Regularize Building Footprint' (RBF). His specific aims for the target polygons are also relevant for shelters in refugee camp.

This master thesis investigates the use of the tool 'Regularize Building Footprint' for the generalization of refugee dwelling outlines. To evaluate the results manual recorded reference objects from the satellite image are determined. The output results are than assessed for applicability by checking whether all objects are edited and whether they correspond to the references in shape, area, perimeter and orientation. For this purpose, various parameters are tested on different types of dwellings. It will also be checked if recommendations of the parameters can be determined for using the tool in other refugee camps.

With this intention, the following research questions will be investigated.

- What benefits does the 'Regularize Building Footprint' function have for the generalization of refugee dwelling outlines based on very high-resolution satellite imagery?
- What relationship can be found between the best fitting parameters of the 'Regularize Building Footprint' function and the area of the extracted outlines?

To answer these research questions, the extracted building outlines of the Minawao camp are verified. The output of the method is evaluated quantitatively and qualitatively with the help of a scoring model.

1.2. Background

1.2.1. Refugee Movement

Until the end of 2022, 108.4 million people worldwide had to leave their homes because of wars, human rights violations, famine, persecution and climate disasters (UNHCR, 2023). These forcibly displaced people are called internal displaced persons (IDPs) or refugees. The IDPs leave their home region mostly because of natural disasters. However, they remain in their home country with the aim of returning as soon as possible. In contrast to internal displaced persons, refugees leave the borders of their home country. In most cases this is because of violent conflicts. During their journey, people are registered in the new country and are therefore legally accepted. Both groups find protection in refugee camps for a shorter or longer period. They often only have a few valuables of their own and therefore depending on shelter and food as well as additional support from aid organizations during this time (Kemper and Heinzel, 2014).

Especially during large or dynamic refugee movements, the number of refugees in camps can fluctuate rapidly. Because of violent conflicts, difficult access to effected areas, and a deficit of human resources, organizations and authorities are often unable to provide fast and specific support. This can lead to shortages of food and water supply, sanitation as well as shelter (Benz et al., 2019). Even when providing emergency aid for several years, international organizations such as UN Refugee Agency (UNHCR) and Médecins Sans Frontières (MSF) quickly reach their limits because they have no information about the camps. Reliable reports on the situation, number of people or size of refugee or IDP camps are therefore essential for these organizations. These data are crucial to be able to provide medical care, infrastructure or vaccination programs. For this reason, aid organizations depend on quick and current news about the situation in the camps (Tiede et al., 2017, Lang et al., 2020).

In recent years Earth observation (EO) data like "Very High Spatial Resolution Satellite Imagery" (VHSR) has become a very useful tool. They offer reliable and present satellite images of the situation in refugee camps. Currently a big range of optical sensors are available for this purpose. They provide information with high spatial, temporal and spectral resolution and offer a huge potential for observing the Earth. However, the rapid development and constant improvement of Earth observation systems require equally rapid and targeted processing of the data and lead to new challenges for research. Furthermore, the information from these Earth observation data must be fast available for aid institutions as a basis for decision-making. Mostly they are available in form of geodata products such as maps or 3-D models (Lang et al., 2015).

1.2.2. Christian Doppler Laboratory

The Christian Doppler Laboratory (CDL) for Geo- and EO-based Humanitarian Technologies (GEOHUM) at the Department of Geoinformatics - Z_GIS at Paris Lodron University Salzburg is taking up this challenge. In cooperation with the MSF the team is investigating the extraction of refugee accommodation from satellite images to estimate the camp population based on building outlines (GEOHUM, 2023a).

The CDL focusses on three research areas with “Img2Info” on the one hand and “ConSense” and “Info2Comm” on the other hand. These three overarching topics are shown in Figure 2 with further sub-areas.

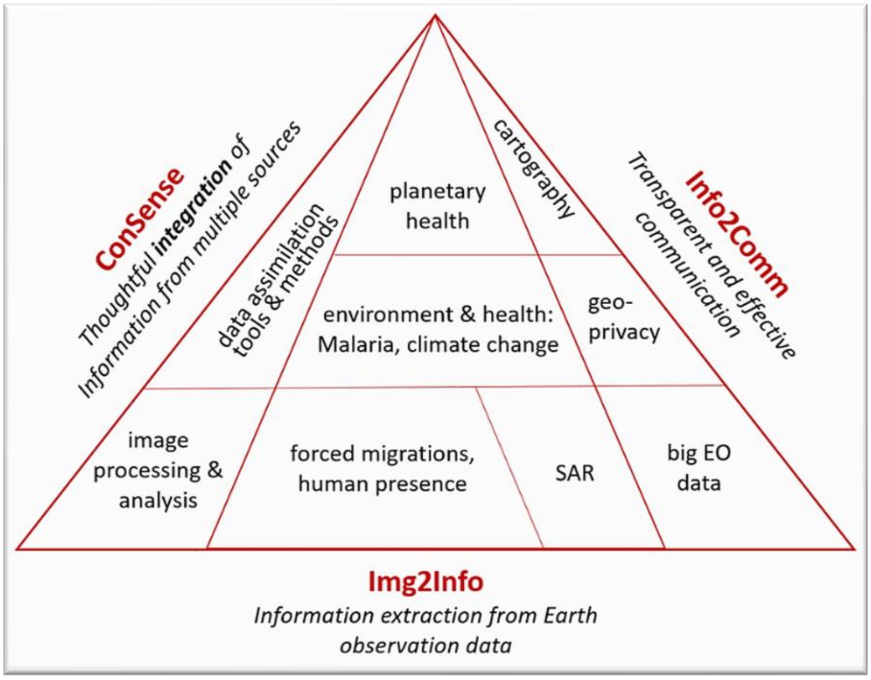


Figure 2: Research areas of the CDL (GEOHUM, 2023b)

The first research area "Img2Info" deals with the processing of Earth observation data. This is important for obtaining information from images or lidar sensors. The satellite images are analyzed, for example, by an Object-based image analysis (OBIA) or by new approaches such as Deep Learning (DL) in Convolutional Neural Networks (CNN). This makes it possible, to extract the building outlines in a camp from satellite images with very high spatial resolution and to identify and classify the type of dwelling on the basis of certain features in a quick step (GEOHUM, 2023b).

The second research area "ConSense" deals with the further processing and linking of the results with additional data and information. For example, the results of the primary Earth observation data are linked with statistical values or with Open Street Map data (GEOHUM, 2023b). In the situation of dwellings, the building outlines are compared and linked with a catalogue from UNCHR, which indicates the maximum number of people for each tent. Based on this information an estimate of the entire camp can be made (cf. Figure 3).

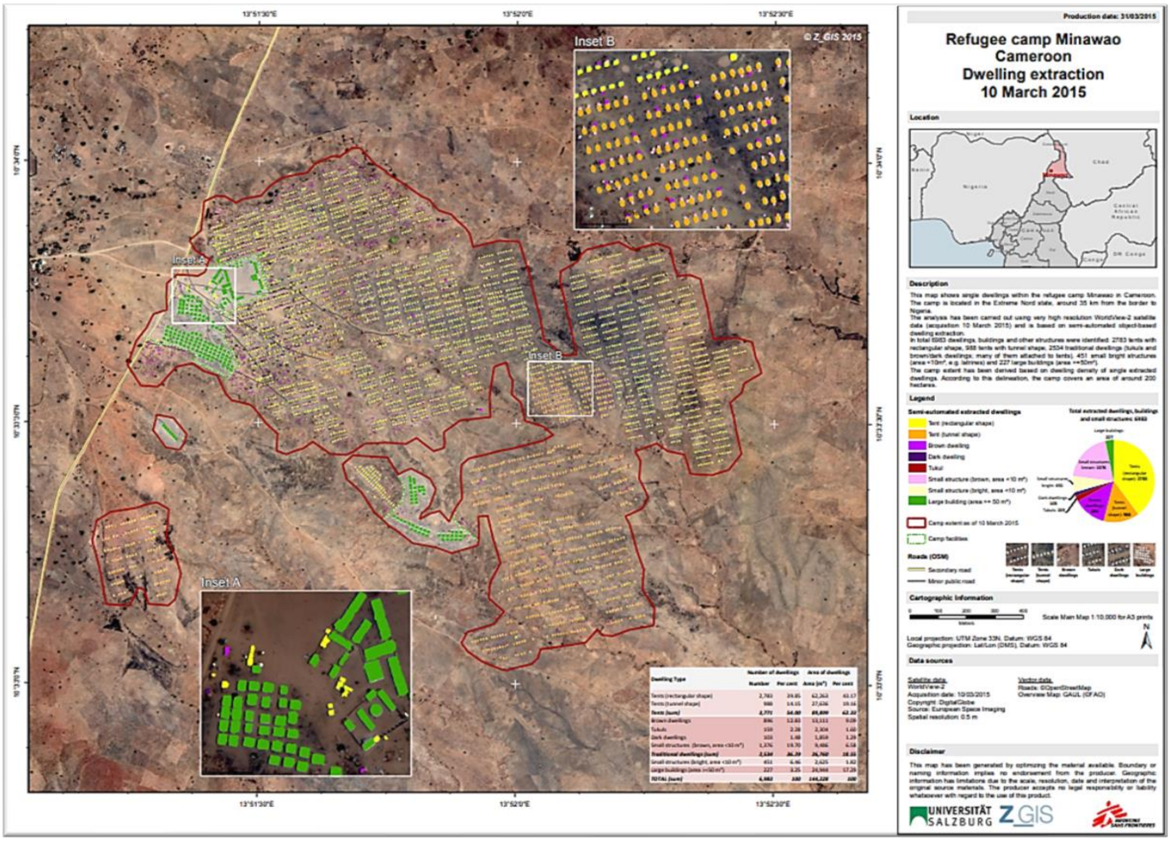


Figure 3: Estimation of the camp population (Z_Gis and MSF, 2015)

The third research area is the "Info2Comm". Because of a huge amount of data from the first two pillars, a valuable way must be found to show this information clearly for organizations. This part deals with the presentation and supply of data for aid organizations. It is therefore necessary to prepare these results in such a way, that they can be read and understood easily and without any background knowledge. Afterwards the products can offer an assistance for aid organizations and should therefore not lead to confusion among decision-makers because of an inaccurate presentation (GEOHUM, 2023b).

One specific research field of the GEOHUM and MFS is the Minawao camp. It is in the extreme north of Cameroon and was established in 2013 by the UNHCR. At that time many Nigerians were fleeing from the violent Islamist group Boko Haram. The terror continues to this day and the flow of refugees does not want to end. Therefore, people are always looking for shelter and support. As a result, the number of refugees in Minawao continues to rise. Targeted support from aid organizations is therefore essential for the operation of the camp (UNHCR, 2022). Several studies have already been conducted at the CDL to gather information about this refugee camp using satellite imagery. The outlines of buildings from Minawao camp were extracted from high-resolution satellite imagery in a semi-automatic procedure for population estimation (Lang et al., 2020). For several reasons, these outlines contain errors that make it impossible to present the dwellings accurately in maps.

This work builds up on this previous research of the CDL and is part of the third pillar of the research project. This paper deals with the generalization of the extracted building footprints for accurate representation of tents and tukuls in maps and other geoinformation products. For this purpose, the ‘Regularize Building Footprint’ function is applied as a generalization method to the extracted building outlines.

1.3. Study area and data

1.3.1. Study area

The study area of this research is the Minawao camp in Afrika. It is located in the Extreme North of Cameroon, about 70 kilometers from the border to Nigeria (cf. Figure 4). Since the violent right-wing extremist group Boko Haram starts to terrorize, many Nigerians were fleeing to neighboring Cameroon. In 2013 the UNHCR set up the camp to offer protection to the refugees. Until today in 2023, the terror of Boko Haram's continues, and many people still arrive in Minawao. Since the opening the number of refugees has doubled to almost 75,000 people, however the camp is not designed for this number of refugees. Organizations such as the MFS and the UNHCR support and protect the refugees during their stay (UNHCR, 2022).

This study uses data from 2015. At this time the camp accommodated about 40,000 people. The camp enclosed an area of about 200 ha with a total of 6983 emergency shelters (Z_Gis and MSF, 2015).

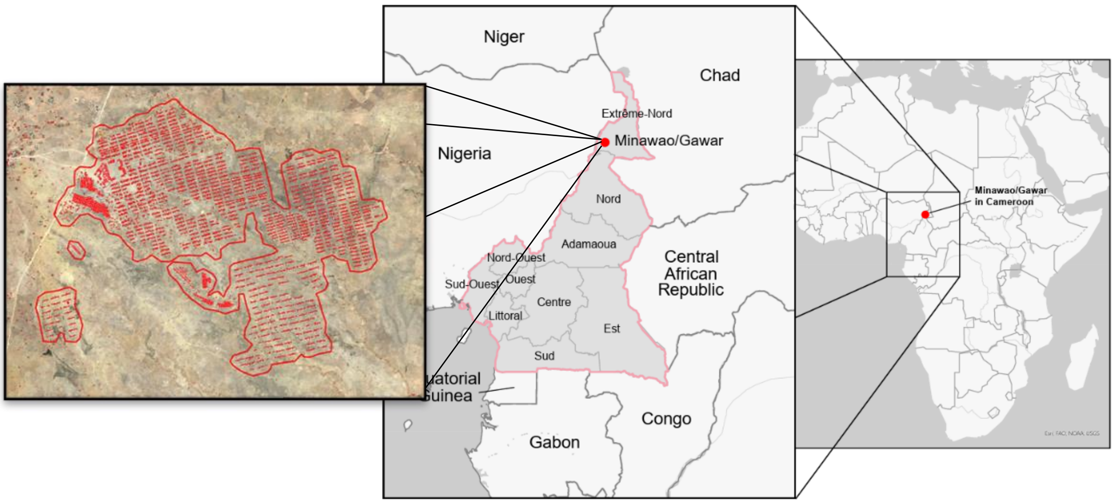


Figure 4: Study area – Minawao camp in Cameroon

As a result of the overcrowded camps, sanitation is often inadequate and there are outbreaks of disease. For example, a cholera epidemic broke out in 2022. This was caused by contaminated food and water (Kindzeka, 2022). The climate of the region plays in this context an important aspect. The rainy season in the region lasts from April to October. With a maximum precipitation of 228 mm, there is barely rainfall. The rest of the year, from November to March, temperatures in northern Cameroon are between 25° and 30 °C (Bank World, 2023). This semiarid climate causes terrible water quality, which leads to a lower level of hygiene and outbreaks of diseases.

This tough condition complicates the support. Many aid organizations often reach the limits of their capacity. They are not able to offer adequate accommodation in a fast way. Also, to improve the hygiene standards and to avoid preparing food with contaminated water is difficult.

For example, the Land Life project is concerned with the improvement of the camp life. The aim is to create better conditions for the people by reforesting the region in the far north of Cameroon. The UNHCR and Land Life have been involved in this project since 2017. They are supported by the Dutch National Postcode Lottery. Together they want to create opportunities to cook in a more environmentally way by using sustainable fuel from trees. Another aim is to replace old plastic tents with sustainable shelters. An important point is to educate the refugees in the tree nursery. By planting trees and plants refugees can gain valuable experience and knowledge (Land Life, 2022).

Therefore, it is important that aid organizations receive sufficient data about the refugee camp. This must be in a clearly way to improve the living conditions.

1.3.2. Data

To identify the refugee shelters in Minawao a WorldView-2 satellite image and an Object-based image analysis (OBIA) were used. The WorldView-2 satellite mission was launched 8th October 2009. Although the rapid development of Earth observation and geo-information systems has enabled already higher spatial and temporal resolutions, these images are still part of the highest resolution satellite images which are freely available. Technical specifications for WorldView-2 (ESA, 2023):

- Orbit Altitude: 770 km
- Orbit Type: Solar synchronous
- Orbit Period: 100 minutes
- Revisit Time: up to 1.1 days
- Equator Crossing Time: 10:30 descending node
- Spectral Resolution: 1.84 m
- Panchromatic Resolution: 0.46 m

The satellite image in this study was taken 10th March 2015. It is available as an RGB Geo-TIFF file with a pixel size of 0.46 m, 16-bit RGB and 4 channels. For the coordinate system the EPSG:32633 - WGS 84 / UTM Zone 33N was used. In preparation for the analysis, the image was pan-sharpened (see Figure 5). This allows a detailed analysis during further assessments. In this procedure the higher resolution of the panchromatic band (left image) is adapted to the image with the lower resolution, in this case the color image (middle image). At the end the spectral image has the same resolution as the panchromatic image (see right image)(Garzelli et al., 2004).

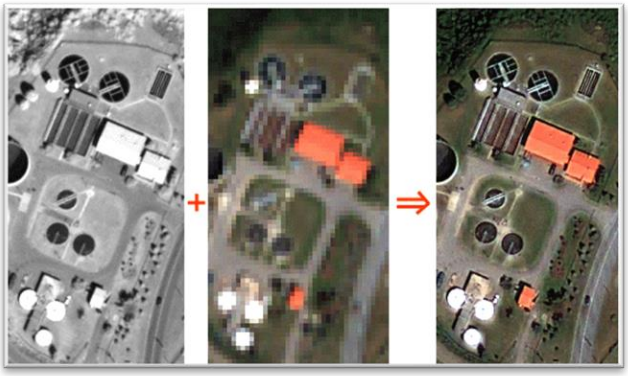


Figure 5: Adaption of the resolution (geosage, 2023)

Extraction methods

As described above, there are various ways to extract information from satellite images.

Current approaches use Deep Learning and Convolutional Neural Networks to support the automatic extraction of building footprints. These new systems enable an easier way to understand the earth's surfaces in a quick way. On the other hand, these developments lead to a large amount of data and therefore to new challenges for data processing (Zhou et al., 2018, de Lima and Marfurt, 2020, Gella et al., 2023). Traditional analysis, such as visual image analysis, are time-consuming. Therefore, their use in the humanitarian field, which needs fast information, is not possible (Lang et al., 2020). Other methods like semi-automatic analysis are significant faster. For example, OBIA or the pixel-based image analysis. They work with the spectral properties of the image. The pixel-based image analysis uses the information of each individual pixel. In contrast, OBIA uses information from a group of similar pixels (see Figure 6). Based on similar values of spectral properties and their surrounding context, a group of pixels is defined as an object (Blaschke, 2010). As described in the introduction, there are a lot of effects that lead to a less accurate result than with a visual interpretation (Gao et al., 2022).

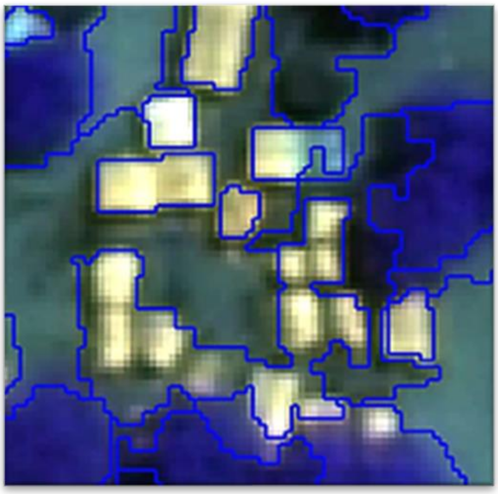


Figure 6: OBIA segmentation of similar pixels

1.3.3. Dwelling extraction as an input for this study

In this study, the results from a building extraction algorithm developed by Tiede and Lang was used. It is based on a method by Lang that had been further developed over several years (Lang, 2008, Tiede et al., 2010, Tiede et al., 2013). At first the important areas based on the satellite image were determined. Then the objects for classification, in this case dwellings, were identified. For this step the brightness or the relative contrast of the pixels compared to the surrounding was used. If similar pixels were nearby, they were grouped into objects or groups. Based on the individual determined objects, a supervised classification was used with a Support Vector Machine. For this purpose, 10 examples of each object group were selected to be the reference for the wanted classification. For this step statistical values like the mean of the spectral bands as well as the shape and size of the objects were considered. The final step of the classification was to divide the generated dwelling types into different shelter classes based on a population density (Tiede et al., 2017).

With this information a total of 6983 shelters, which can be divided into seven different shelter types, were identified (cf. Figure 7).

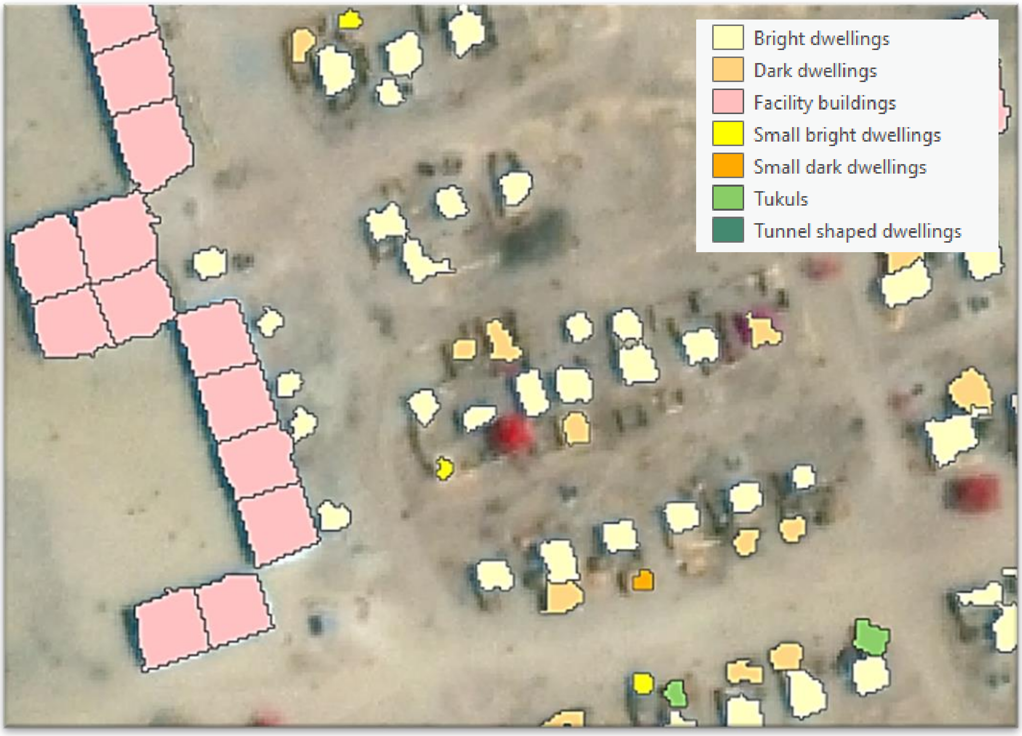


Figure 7: Section of the extracted dwellings

The bright rectangular tents are the largest class with 2783 buildings. Furthermore, 988 tunnel tents and 2534 traditional dwellings were counted. These are tukuls and brown/dark buildings. They are often attached to rectangular tents. In addition, 451 buildings were calculated with a size less than 10 m². Small buildings as well as latrines are included in this category. 227 larger buildings were counted with an area of 50 m² or more. These were mainly used by aid organizations as supply rooms (Z_Gis and MSF, 2015).

The UNHCR has published a catalogue of standard tents for refugee shelters. There is a wide variety of shelters. According to need the size and function of a tent can differ. For example, there are octagonal family tents with dimensions of 4 x 6.6 m, 4.15 x 4 m, or 4.3 x 4.3 m. The size of the tukuls, which are mostly round shelters, can also vary enormously (cf. Figure 8). This catalogue describes also how many people maximum can stay in each tent (UNHCR, 2016).

Figure 8: Different types of dwellings (left = tukul, right = family tent) (UNHCR, 2016)

2. Generalization methods

In recent years, not only earth observation sensors and image analyses techniques have been further developed, but also functions for generalization of polylines have been improved. They are used for raster vector conversion, map generalization or the reduction of data volumes. For this application simplification or smoothing methods are used.

Several algorithm build on elimination of specific features such as curves (for example arms of rivers), others on removing unnecessary vertices or approximation of lines by mathematical functions (Bodansky et al., 2002). There are different approaches with various algorithms. Some functions are especially useful for generalization of building footprints.

When simplifying building outlines, it can be disadvantageous if the new lines lie on the vertices of the original. This should be avoided especially if the building outlines contain many noises. Furthermore, specified criteria and geometric options that are important for the generalization of building outlines must be considered. These could be straight or parallel lines as well as right or other angles between two sides.

The following chapter describes common methods such as the "Minimum Bounding Geometry", the "Douglas-Peucker" algorithm and the "PEAK" method. They are mainly developed for data compression and map generalization, but also can be used for normalizing building outlines. Furthermore, the method RBF is shown, which was created specifically for building generalization and contains explicit conditions to normalize building footprints. Each of the methods is available as a tool in ArcGIS. The specific advantages and disadvantages of the technique in terms of simplification and approximation outlines are discussed in the following chapter.

2.1. Minimum Bounding Geometry (MBG)

One method to generalize a building footprint is the "Minimum Bounding Geometry". The aim is to get the smallest possible wrapping around the original polygon (cf. Figure 9).

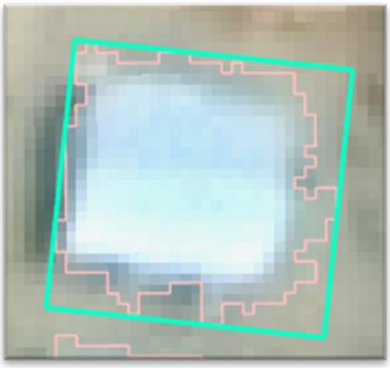


Figure 9: Regularization of a Facility Building with the MBG (cyan line)

The algorithm is included in the standard installation of ArcGIS and therefore freely available. It is also existing as a tool in the open-source software QGIS. Throughout the use different types of shapes can be defined for the resulting polygon. Depending on the requirement, for example, the smallest possible circle or the smallest possible rectangle around all vertices of a polygon can be identified. During this process the given line is not modified.

The algorithm is based on investigations by Shamos. He was searching for a convex hull around a group of data points (Shamos, 1978). In further researches, the principle of the rotating caliper was used for the calculation of the smallest surrounding rectangle (Toussaint, 1983). Through an iterative procedure, the possible polygons are formed around the initial points. The polygon with the best result is then returned.

The method below uses the principle of the rotating caliper for the calculation of a minimum oriented bounding rectangle (cf. Figure 10):

1. Calculation of the convex hull (black line) of a polygon (green line).
2. Searching for the extreme points of the convex hull of the polygon.

$$P1, P2, P3, P4$$

3. Creation of two vertical support lines at $P1 - P2\ and\ P3 - P4$ and two horizontal lines at $P2 - P3\ and\ P4 - P4$ (red lines).
4. Then the supported lines are rotated until one line coincides with an edge line of the convex hull. This step is repeated until all edges of the convex hull have matched with

one of the support lines (blue rectangles). During this process, the area of each rectangle is adapted and calculated and at the end the smallest rectangle is determined.

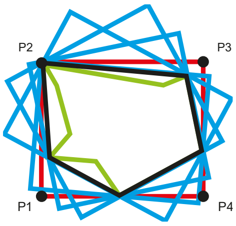


Figure 10: Searching for the minimum oriented bounding rectangle.

This ArcGIS tool can create different shapes depending on the situation. These are (cf. Figure 11) (ESRI 2023):

- CONVEX HULL - The convex hull is the smallest possible hull of a polygon that contains all its vertices.
- RECTANGLE BY AREA - This function calculates the rectangle with the smallest area.
- RECTANGLE BY WIDTH - This function calculates the rectangle with the smallest possible width.
- CIRCLE - This function searches for the smallest circle that fits around the hull of an input feature.
- ENVELOPE - The target is the smallest undirected rectangle around a group of data points (Step 1-3 of the described method above).

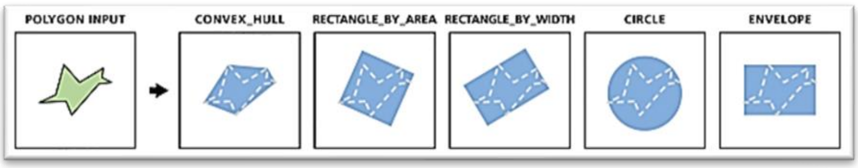


Figure 11: Target shapes of the bounding geometry (ESRI, 2023c)

This method is not really a compression tool, but an easy technique to generalize the shape in a quick way. The advantage of this method is that the determined shapes of the new polygons can be orthogonal or have right angles. The disadvantage of this technique is that the area is always larger because all the vertices have to be inside of the box. Furthermore, only simple forms are permitted. The method works well when the original objects are rectangular or square and contain few minor "artifacts". However, if the building structure is more complex, this method is not suitable. For example, specific angles cannot be modified or detailed features are also discarded. In this case an approximation of the original object is therefore very unlikely.

2.2. Douglas-Peucker-Algorithm

In 1973, David Douglas and Thomas Peucker developed a method that simplifies a line while trying to keep the size and shape of the original object (cf. Figure 12). Important data points that characterize the shape are retained. Redundant points are removed (Douglas and Peucker, 1973).

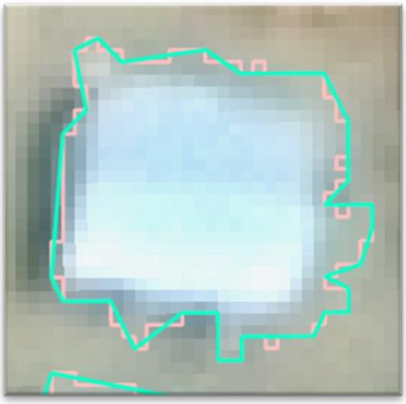


Figure 12: Regularization of a Facility Building with the Douglas-and-Peuker-Algorithm (cyan line)

The algorithm is included in the standard installation of ArcGIS and is freely available. It is also part of the open-source software QGIS. Figure 13 shows the iterative process for removing vertices according to a specified tolerance.

a. In a first step the end points of a line (0) are connected with a straight line (red line 1). The following part is to determine the distance from each vertex to the new line. Then, the point with the highest distance from the new line is determined (blue point 1). At the end the points which are closer to the trend line than the tolerance is eliminated (yellow).
b. During the next step, the trend line is divided at the point which is farthest away from the original line. Then the vertex with the highest distance to the new trend line is searched for again. At the end the points which lie within the tolerance (2) are eliminated again.
c. This process is repeated until the goal of eliminating points out of the tolerance is achieved (3) and the optimal polyline is determined (4).

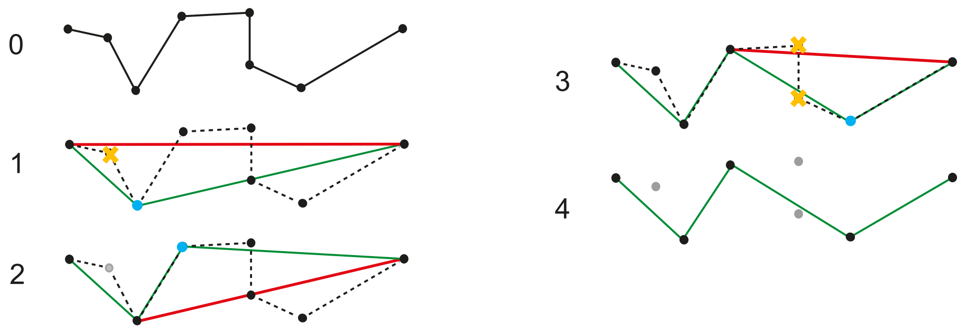


Figure 13: Steps of the algorithm

Before using the algorithm, the tolerance can be specified. It determines the maximum allowed orthogonal distance between the support points and the new line. This influences the intensity of the simplification. A large value is required for a rough result, a small value for an accurate effect (ESRI, 2023c).

This algorithm a good way to reduce the data of a line or polygon. With this method the shape and thus the area and perimeter of the original object can be preserved. The orthogonality and other geometrical conditions that are desired for building generalization are difficult to retain. Another disadvantage is that the resulting lines may contain unsightly corners.

2.3. Polynomial approximation with exponential kernel (PEAK)

This method is available as a free tool in ArcGIS as well as in the open-source software QGIS and uses the "continuous local weighted averaging" algorithm defined by Bodansky et al. (2002). In this procedure a smoothed line is find by calculating the parametric continuous averaging (cf. Figure 14). Thus, the new line does not have to go through the original points. With a tolerance the intensity of the smoothing can be defined.

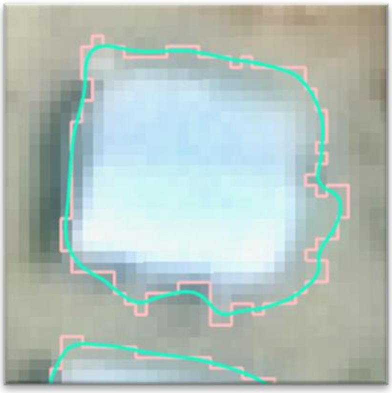


Figure 14: Regularization of a Facility Building with the PEAK algorithm (cyan line)

During this process for all vertices of the initial line the average of the coordinates is calculated. Throughout this process the distance from a point is considered and weighted (ESRI, 2023a). That means, that a point which is far from the other points is less weighted (red points). While vertices which are nearby (green points) are more weighted (cf. Figure 15). With the tolerance the maximum length (pink line) of the line of which the points are assessed can be chosen. Therefore, with a higher tolerance the line (dark green) is smoother than with a lower value (bright green). This function is also using a second-degree polynomial approximation method to create the new points. The result is a line which points are not necessarily part of the original line (Bodansky et al., 2002).

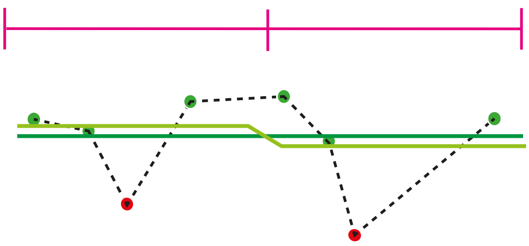


Figure 15: Schematic illustration of the PEAK

2.3. Polynomial approximation with exponential kernel (PEAK)

In this case a generalized line which considers the form while reducing noises and specific characteristics is calculated. The disadvantage of this technique is that the determined appearances of the new polygons cannot be orthogonal or have right angles. In this case specific angles are not possible and necessary features are also discarded.

2.4. Regularize Building Footprint

The RBF was developed by Gribov et al. in 2019. The specific aim was to create a technique to normalize footprints of building outlines from Earth Observation imagery. Several properties (such as orthogonality) can be considered with this function (cf. Figure 16). The tool is available as an extension (3D Analyst) in ArcGIS and is therefore not available freely.

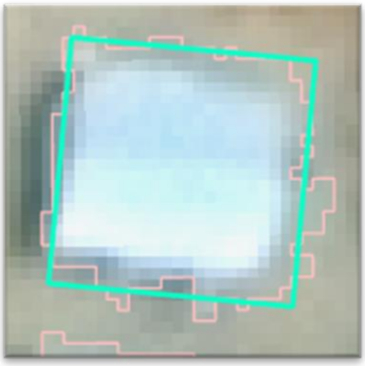


Figure 16: Regularization of a Facility Building with the RBF (cyan line)

Alexander Gribov has been working for several years on the simplification and smoothing of lines and polygons after the raster vector conversion and building extraction. First studies were carried out in 2002 in cooperation with Bodanksy. They developed the "continuous local weighted averaging" algorithm, which was used for the PEAK method described above. This simplification method is a good way to reduce the size of the data before using a compression algorithm to improve the accuracy (Bodansky et al., 2002). Further investigations deal with searching for the best circle, which means the best approximation of “round” outlines (Bodansky and Gribov, 2004). In 2004 they present a new technique for generalization. They solve the problem of some compression methods, like that from the algorithm of Douglas and Peucker, where the target line contains the noises from the original outlines. That implies, that if the polygon has outliers, they are also transferred to the new line. In "A new Method of polyline approximation" Gribov and Bodanksy describe how they minimize the errors with a piecewise linear approximation. By dividing the line into clusters with similar direction in distance of critical points, they replace it by a straight line (Gribov and Bodansky, 2004). To find these critical points within a line a threshold is used. In other studies, they worked on a process to generate specific geometrical conditions for the target buildings. Considering the fact, that buildings have right angles and straight sides this leads to the point, where all lines of a polygon have only two possible directions. Gribov called this the ‘cardinal directions’. This means that any two lines are either parallel or perpendicular to each other (Gribov and Bodansky, 2006). The algorithm builds on a study of 2004 and enable to create buildings with orthogonal sides. To calculate a rectangle the process starts from the beginning point of a polygon. In a clockwise way the lines are fragmented in horizontal or vertical lines. During this the new line is built in combination with the least square adjustment theory (cf. Figure 17). Whether the lines are perpendicular or straight is indicated by a threshold.

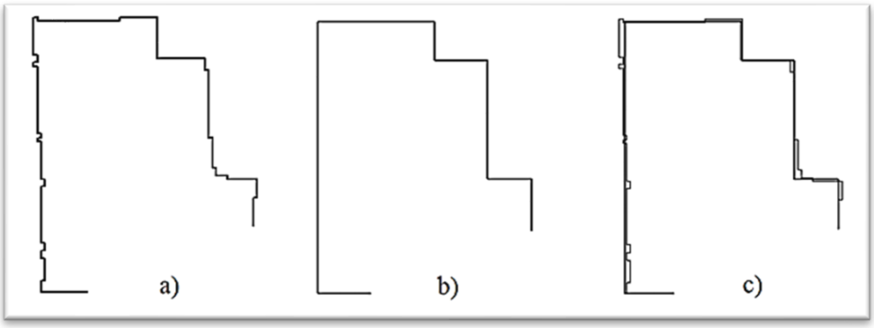


Figure 17: a) original polygon, b) new polygon, c) combination of both (Gribov and Bodansky, 2006)

For the RBF function Gribov (2018) creates another new approach in combination with previous methods. One aim of the function is to find an ‘optimal polyline’ (Gribov, 2018). This ‘optimal polyline’ must lie within a specified tolerance to the original polyline and at the same time, the result has to contain the lowest number of vertices. In this case, only a subgroup of the points of the original polyline is applied to create a new line. If the points lie within the tolerance (red) of the original polygon, they are used to build the newline. If they lie outside the tolerance, they are discarded (blue point) (cf. Figure 18).

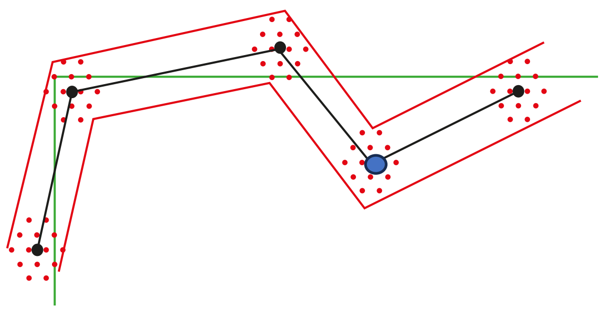


Figure 18: Optimal line within a tolerance

On the other hand, he describes the opportunity to get not only orthogonal segments but also polygons with different angles. This is important because the geometric properties between the sides of the buildings are also crucial. In a further approach Gribov described the improvement of the last algorithm to create buildings not only with right angles but also in combination with other angles (cf. Figure 19)(Gribov, 2019). For example, tangential changes between straight and curved sides of a building. Another point which is mentioned in the study is the alignment of the orientation of the buildings to lines like streets oder other references.

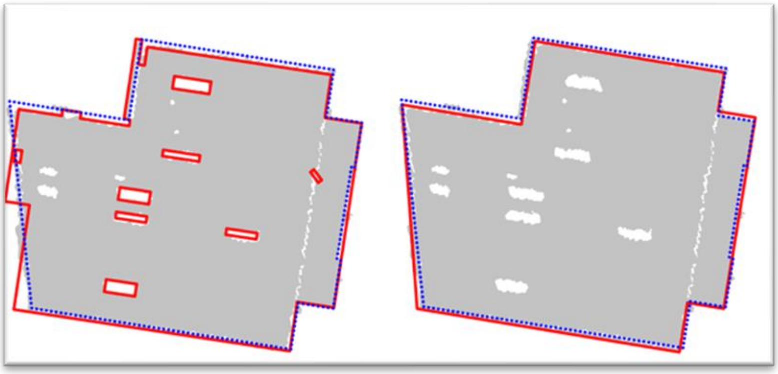


Figure 19: A building with right angles in combination with any angles (Gribov, 2019)

The RBF algorithm describes a lot of conditions that are also necessary for dwelling normalization. The following part gives an overview of the various parameters of the ArcGIS tool for processing the input data and defining the conditions for the target object (ESRI, 2023b).

Target shape methods:

- RIGHT_ANGLES - Angles of 90° are formed between the segments. This means that the segments are perpendicular to each other.
- RIGHT_ANGLES_AND_DIAGONALS - This setting creates shapes consisting of angles of 45 and 90 degrees between adjacent edges.
- ANY_ANGLES - This function is used to create shapes with any angle between adjacent edges.
- CIRCLE - Constructs the circle that fits best to the entered objects. The ratio of the entered features is calculated here. A circle has a ratio of 1, the closer to 1, the more likely the circle is.

Tolerance

The distance between the new and the original line is important for the application (cf. Figure 18). This is determined by the tolerance. It is specified with the units of the coordinate system of the input feature. When the CIRCLE method is used, the tolerance is also interpreted as the difference between the original feature and the result line.

Densification

This parameter decides whether the smoothed feature is straight or curved. It can be seen as the testing interval. The value must be less or equal to the tolerance value. This parameter cannot be used for circles and methods where the result is not a rectangle. With this setting the details of the object can be defined. If the densification is smaller the chance is higher that the line is bent.

Precision

This allows to correct the accuracy of the building footprints. The values can be between 0.05 and 0.25. The precision defines the spatial grid which is used for regulation. This plays a more noticeable role at higher resolution levels, like in the case of residential buildings and cadasters.

Diagonal penalty

This value indicates the likelihood of constructing right angles or diagonals between two adjacent segments. These values should be small. The higher the value, the more likely it is that the defined angles like 45° or 90° will be used. It is used in combination with any angles as well as right angles and diagonals.

The ‘Regularize Building Footprint’ (RBF) can consider specific properties that a normalized building outline must fulfill. The strength of the function lies in the generalization of building outlines made from aerial or drone images. Because of that the tool considers a lot of necessary properties for the generalization of dwelling outlines. The different settings allow a variety of applications, but also require a deeper learning of the tool. Therefore, in this work, the function with its specific properties is analyzed in more detail. The function to align buildings to roads or other features is at the moment not available.

3. Methods

3.1. General workflow

The goal of this research is to evaluate the ‘Regularize Building Footprint’ (RBF) function for generalization of building footprints in the context of refugee accommodation. The tool was selected based on the conditions described above.

An evaluation of the new polylines would be easy if in situ data of the real buildings were available. With this data the differences to the normalized polygons could be easily determined. However, in this case official information are not accessible. Comparing the results with the extracted contours is also not useful. Since many errors are already contained in the polygons, no accurate assessment of the precision of the new curves can be made. The goal is to adapt the forms to the real shapes and not to the outlines, which have errors and are in most cases too big (cf. Figure 20). So, in this study another way must be found to evaluate the accuracy of the function.

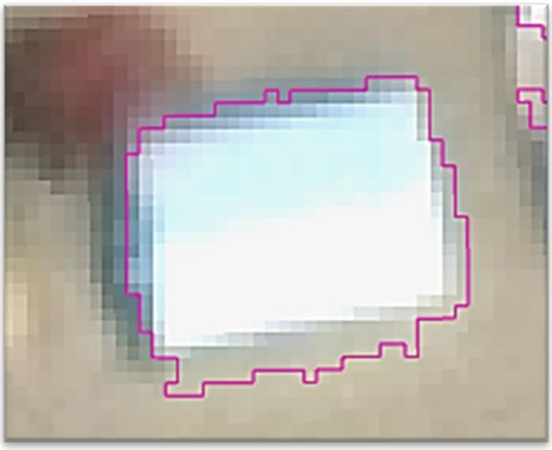


Figure 20: The extracted area is too big (pink line)

To answer the research question, an approach is developed that focuses only on a specific selection of objects in the camp. With this subdivision different parameter of the algorithm were tested and evaluated. The idea is that the best parameters can also be transferred to dwellings with similar sizes and shapes. The collection contains objects for each class and based on the visual homogeneity of the buildings. Objects with equal orientation, size and shape are clustered in a first step. In a second step, twenty objects for each group are created from this sample class manually. It should be note, however, that the visual interpretation cannot provide accurate building outlines either. The main problem in obtaining reference data is related to the same challenges as during the Object-based image analysis (OBIA). The outlines may not be clearly visible due to spectral characteristics. In addition, some of the buildings are very small and can be difficult to capture in the image analysis. Again, this leads to wrong shapes which are too big, too small or the arrangement is wrong. Overall, in this case the manual handling is more accurate for the reference data than the automatic analysis. The goal of the next part is to find the "standard" accommodations of the selected areas. After the buildings have been identified manually, the reference objects are created based on average values of the group. These references are used for the verification.

In order to be able to apply the optimal target shape method for each type of dwelling, further pre-selections and categorization of objects must be made in a first step. The algorithm can then be applied with suitable parameters on these areas of the Minawao camp. The evaluation and comparison of the results with the references is done with a scoring model. In the final step, the best settings with the highest number of points are determined for each class and are applied to the entire camp. The transfer of the parameters is verified by a visual interpretation of the results. Figure 21 shows the described workflow.

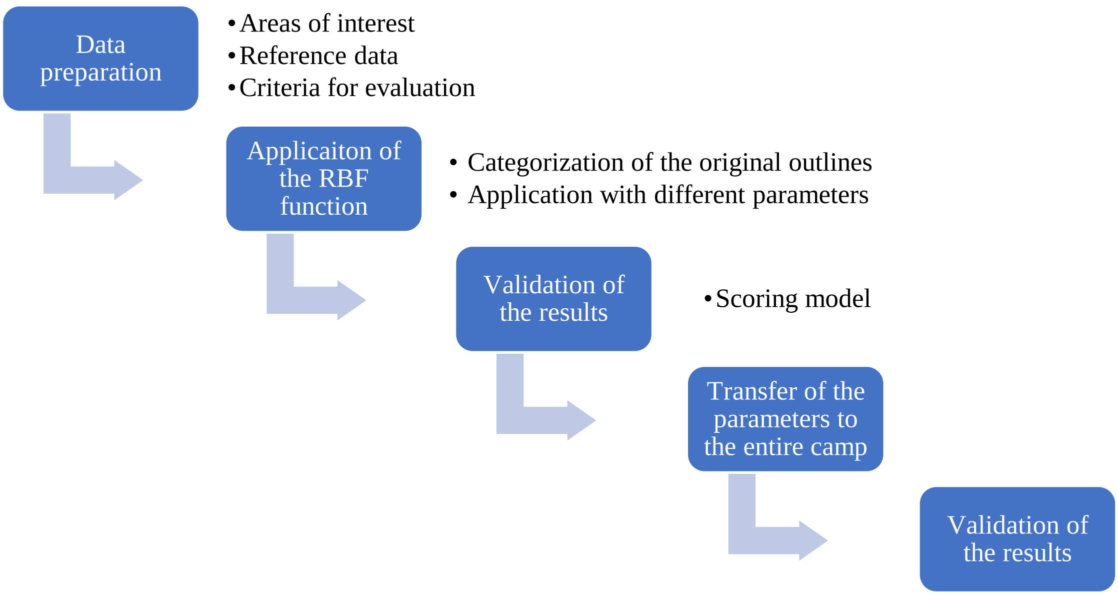


Figure 21: General workflow of the evaluation

3.2. Data preparation

3.2.1. Areas of interest

One of the thesis challenges is to validate the resulting function output with a suitable source. Additionally, a wide variety of different size, orientation, and shape of the dwellings leads to increased complexity. With the available resources the approximately 7000 dwellings of Minawao camp are too much to compare in a single way with manual comparison. As a result, a systematic assessment of the entire camp is difficult. Therefore, only one sample group per dwelling class is evaluated in order to enable the quality check of the resulting lines. One representative group shall consist of twenty objects. It is assumed that the best parameters can be transferred to similar objects with minimal differences in size and form. As conclusion one reference object for each group shall be defined. This is likely by the same properties of dwelling size and orientation (cf. Figure 22). A precise inspection of the region is necessary to use the particular arrangement of the camp to find these unique buildings.

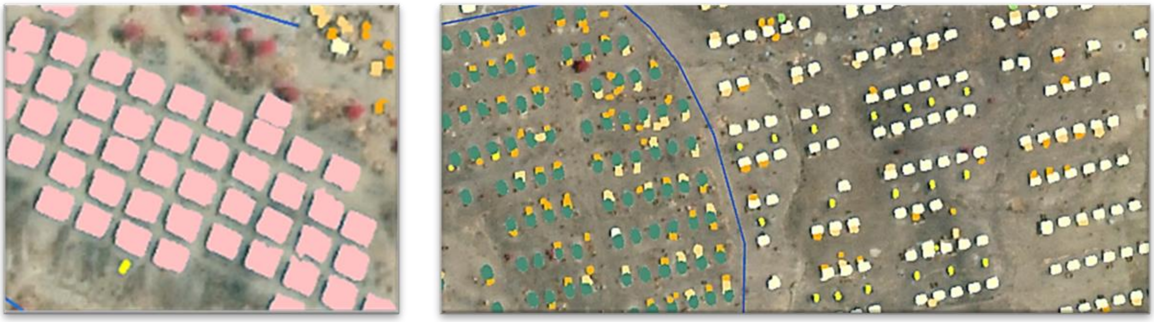


Figure 22: Particular arrangement of the camp

The initial data was already divided into 7 classes by the semi-automatic image analysis . Bright Dwellings, Facility Buildings, Dark Dwellings, Small Bright Dwellings, Small Dark Dwellings, Tukuls and Tunnel Shaped Dwellings are distinguished. As mentioned, there are also differences in size, shape, and orientation of the buildings within each class. Anyway, for some classes the satellite image and the extracted outlines show a uniform shape for many objects. For example, Facility Buildings, Bright Dwellings and Small Bright Dwellings are classes that have a huge number of dwellings with a uniform shape. Therefore, a group of these dwellings can be chosen for the reference samples. Otherwise, classes such as Dark Dwellings do not have any regularity appearance. Therefore, different sizes of objects are chosen in this case.

Another important point to find suitable reference data is the orientation of the buildings. Therefore, objects with the same size only in combination with the same orientation are selected to enable an evaluation of the direction. For this it is important to know that the buildings are arranged in certain patterns. There are areas for Facility Buildings, which are mainly used by institutions as supply or collection rooms. There are special areas for family accommodation and areas for tunnel tents. These were built up to house the people who live in the camp. Dark tents are often placed between Bright Dwellings. Without any special arrangement, Small Dark and Bright Dwellings and Tukuls are scattered throughout the area. The orientation of the buildings depends in most cases on the zone or type of building (UNHCR, 2016).

For the selection of the samples, objects that have the same shape, arrangement, and orientation are determined. The blue lines are the different zones of the accommodation classes. From the areas whose object classes have virtually the same orientation and arrangement sub-areas are visually chosen. The red areas are the zones in which sample parts are selected (cf. Figure 23).

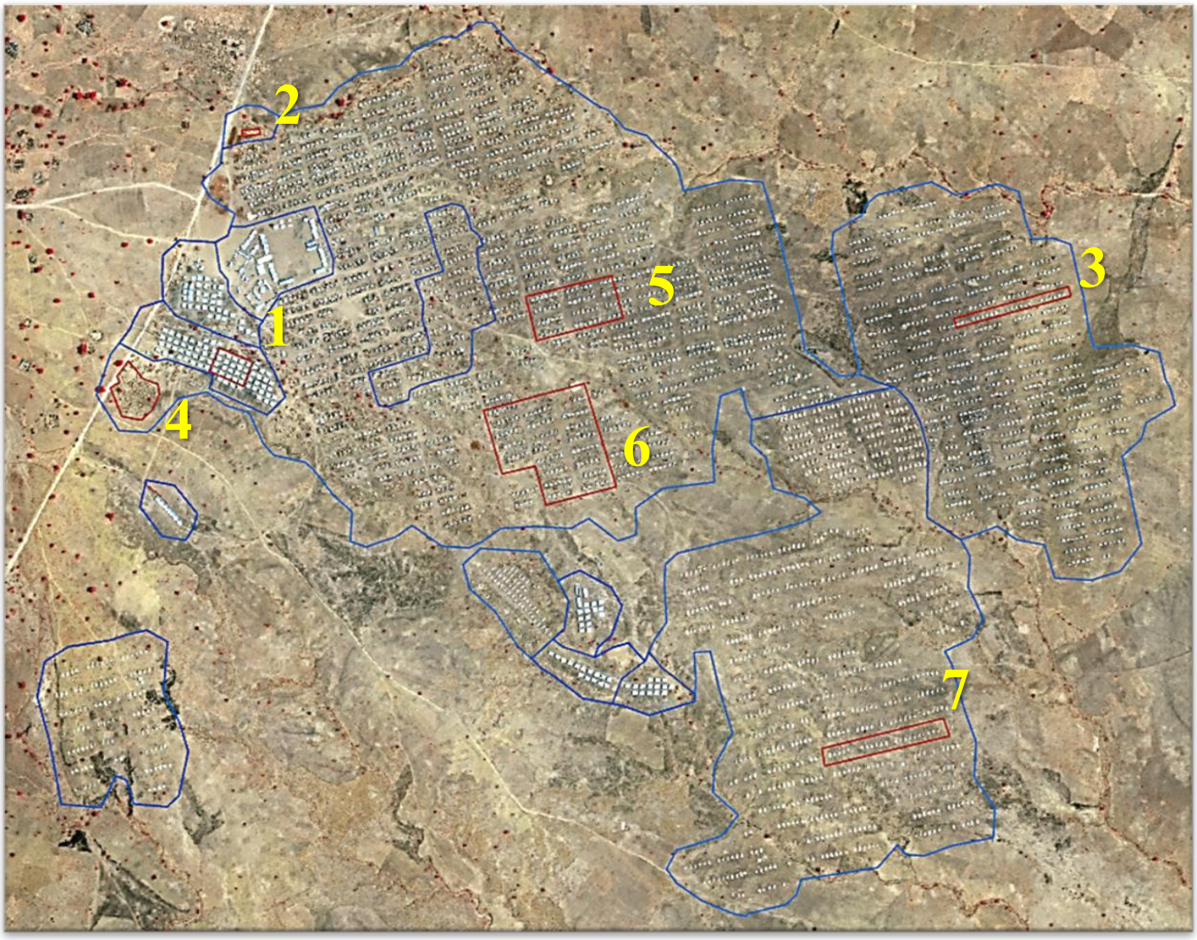


Figure 23: Splitting of the area (blue polygons); Sample data (red polygons)

Figure 23 shows the individual areas for each accommodation type. In the following Figure 24 the twenty objects of each class are chosen as a basis for the further steps.

Figure 24: Areas of interest (1: Facility Buildings, 2: Small Dark Dwellings, 3: Bright Dwellings, 4: Dark Dwellings, 5: Small Bright Dwellings, 6: Tukuls, 7: Tunnel Shaped Dwellings)

3.2.2. Reference data

For the validation of the function output reference objects for each sample class shall be used. The reference objects are described by shape, dimensions, circumference, area, and orientation. The shapes are determined by a visual interpretation of the satellite image and catalogue reference, whereas the size, orientation, circumference, and area of the reference objects are calculated based on mean values from the sample data.

In addition, the twenty objects of each class were recorded manually. For the large, bright objects such as Facility Buildings and Bright Dwellings there are recorded similar shapes. They stand out clearly from their surroundings and have a rectangular shape. Small, dark objects, on the other hand, are more difficult to recognize. They are determined to have a square or rectangular shape based on visual interpretation (cf. Figure 25). According to the UNHCR catalogue the Tukuls are assumed to be circular but vary in size. The Tunnel Shaped Dwellings are also uniformly octagonal with different sizes. This type is also not clearly determined from the image and original data but is expected to match with the catalogue.

Figure 25: Examples of manual analyzed references.

In the assumption, that the dwellings in the sample group are “similar”, the average is calculated. Subsequently, the mean values for length and width as well as circumference and area are calculated for each object type. The orientation is determined by the Minimum Bounding Box (MBB). The orientation of the longest imaginary line connecting the vertices of the antipodes (red points) of the rectangle is determined and this angle is calculated according to the azimuth (cf. Figure 26).

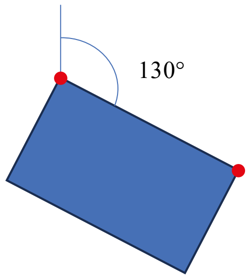


Figure 26: Calculation of the orientation with the MBB

In the following Table 1, the shape, the dimensions, the orientation, the average perimeter and area of the references, and the average perimeter and area of the original objects are presented. The table shows that the area from the most buildings is larger than their reference values. Comparing these values, this is an important point which has to be noticed for the evaluation. Should the resulting polygon be larger, the approximation would be less optimal. This is an indication that the resulting polygon should be smaller. Dark Dwellings have no similar shape, but according to the assessment, the differences of the entire samples are evaluated to see how much they alternate.

Table 1: Parameters of the reference data and the original data

Dwellings	Shape	Dimensions reference in m	Orientation reference in degree	Perimeter reference in m	Area ref. in m²	Perimeter original in m	Area original in m²
Bright dwellings	rectangular	3.71 x 5.25	76.68	17.90	19.28	22.10	23.48
Dark Dwellings	rectangularly	3.13 x 5.29	101.13	16.81	16.58	19.40	15.01
Facility Buildings	rectangular	8.47 x 10.58	114.77	38.09	89.60	50.86	102.90
Small Bright Dwellings	rectangular	1.89 x 3.94	162.12	11.65	7.47	12.55	6.35
Small dark dwellings	square or rectangular	2.11 x 2.30	142.11	8.82	4.86	9.75	5.09
Tukuls	circular	Ø 4.03	-	12.72	13.00	17.35	13.26
Tunnel Shaped Dwellings	uniformly octagonal	4.54 x 6.84	173.45	19.35	26.46	25.46	29.31

The Bright Dwellings, Facility Buildings and Small Bright/Dark Dwellings are determined as rectangular objects with angles of 90°. Dark Dwellings have also a rectangle shape but with more than four angles. For tunnel-shaped flats an octagonal elongated reference object is chosen, therefore angles over 90° are applied. According to UNHCR the Tukuls are supposed to be round. Figure 27 presents the shapes of different types.

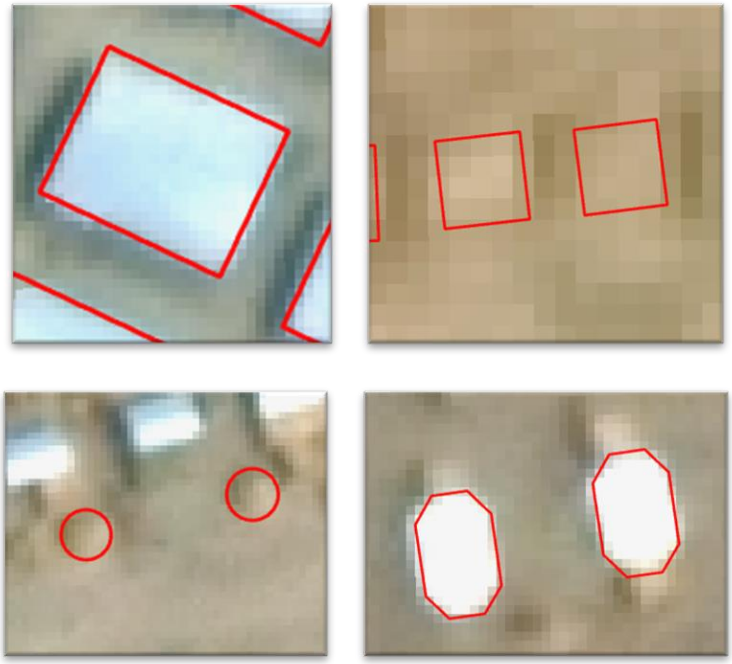


Figure 27: Shapes of the reference dwellings (rectangle, square but also rectangularly, circle and octagon reference data)

3.2.3. Criteria for evaluation

In order to investigate and evaluate the quality of the generalization of building outlines, the objectives and characteristics for the evaluation must be determined.

The investigation and evaluation objective are to answer the research questions as defined in chapter 1.

- Which benefits has the ‘Regularize Building Footprint’ function for the generalization of building outlines of refugee shelters which were extracted from very high-resolution satellite imagery?
- Can a relationship be found between the best fitting parameters of the ‘Regularize Building Footprint’ function and the area of the extracted outlines?

The research questions build the basis for the definition of the comparison and evaluation characteristics. The generalization results of the sample data will be systematically compared to the reference object per class and evaluated afterwards. The comparison will result in quantified values as well as non-measurable only visible judgable values.

Derived from the research questions the following sub questions are in focus to be answered and the assessment based on their required conditions to match with the reference objects per class. These terms are:

- The resulted polygon corresponds to the desired reference shape.
- The orientation matches to the reference shape.
- The area and perimeter differ only minimal from the reference group.

To be able to assess the RBF function results a scoring model is used. This is a proven method to evaluate all relevant aspects and rank the results by object or alternatives. Each object is evaluated against defined criteria and compared in detail. With a scoring model non-measurable values can be included in the assessment (Kühnapfel, 2021). Especially the shape is an important criterion but is difficult to quantify.

The scoring model is based on the following procedure:

1. Define objects and alternatives with different parameters.
2. Define criteria for evaluation.
3. Weight the criteria.
4. Evaluate the criteria and score.
5. Multiply and sum up the scores.

In this case the results of the different parameters are defined as alternatives (see Table 1). The evaluation criteria in direct link to the before discussed sub questions are area, perimeter, shape, and orientation. The points are awarded according to the desired weighting. The most important criterion is the shape. If it does not match the reference, the parameter is not optimal. For this reason, the weighting of the shape is the highest at 50%. If 75,1-100% of the objects have the desired shape, the method gets 5 points. If 50,1-75% have the wanted shape, it gains 4 points. With 25,1-50% difference the function gets just 3 points and with 0-25,1% only 2 points. Should none of the objects have been edited it gets only 1 point. The second important criterion is the orientation. The higher difference to the references, the more the quality of the method decreases. The alignment is measured according to the deviating number of degrees. If less than +-5° it awarded 5 points. For every further +-5°, one point is deducted. With more than +-20° misalignment only one point is scored. The area and perimeter play a subordinate role. They are heavily based on the original data and can therefore deviate considerably from the references. The parameter gets 5 points if the percentage disagreement is less than +-5%. With every more +-5% misalignment, 1 point is subtracted. From +-20% onwards, no more points are given. Area and circumference thus play a minor role and are not decisive factors when comparing the shapes.

Finally, the importance of the properties must be defined. For this study it is assumed, as mentioned above, that the shape is the most significant part and has a weighting of 50%. The second is the orientation with 30%, and the area and circumference with 10% for each value. This results in the following evaluation matrix presented in Table 2.

Table 2: Matrix for evaluation

	Weighting	1 point	2 points	3 points	4 points	5 points
Shape	50 %	Not Edit	<25,1%	25,1-50 %	50,1-75%	75,1-100%
Orientation	30 %	>20°	+-15,01 - +-20°	+-10,01°- +- 15°	+-5,01°- +- 10°	<+- 5°
Area	10 %	>20%	+-15,01 - +-20%	+-10,01 - +-15%	+-5,01 - +-10 %	<+-5 %
Perimeter	10 %	>20%	+-15,01 - +-20%	+-10,01 - +-15%	+-5,01 - 10 %	<+-5 %

At the end, the maximum score of a parameter combination is 5, except for the tukuls (3.5) where the orientation is not evaluated. With regard to the first research question, the higher the score of the parameter combination is, the better is the quality of the normalized outlines. The relationship between the best settings and the extracted dwellings should be providing information about the parameters to use them in other refugee camps.

3.4. Application of the ‘Regularize Building Footprint’ function

3.4.1. Workflow of the application

The parameters of the ‘Regularize Building Footprint’ (RBF) function allow many variations of target polygons and must therefore be chosen individually depending on the dwelling class. Due to different dwelling sizes and shapes it is not possible to choose one value per parameter for all dwellings. Next to the target shape, also tolerance, densification and diagonal penalty have to be set. Therefore, several decisions must be made in advance in order to select the correct parameters and methods within the tool.

Figure 28 shows the workflow including application of the fitting parameter values for each sample class.

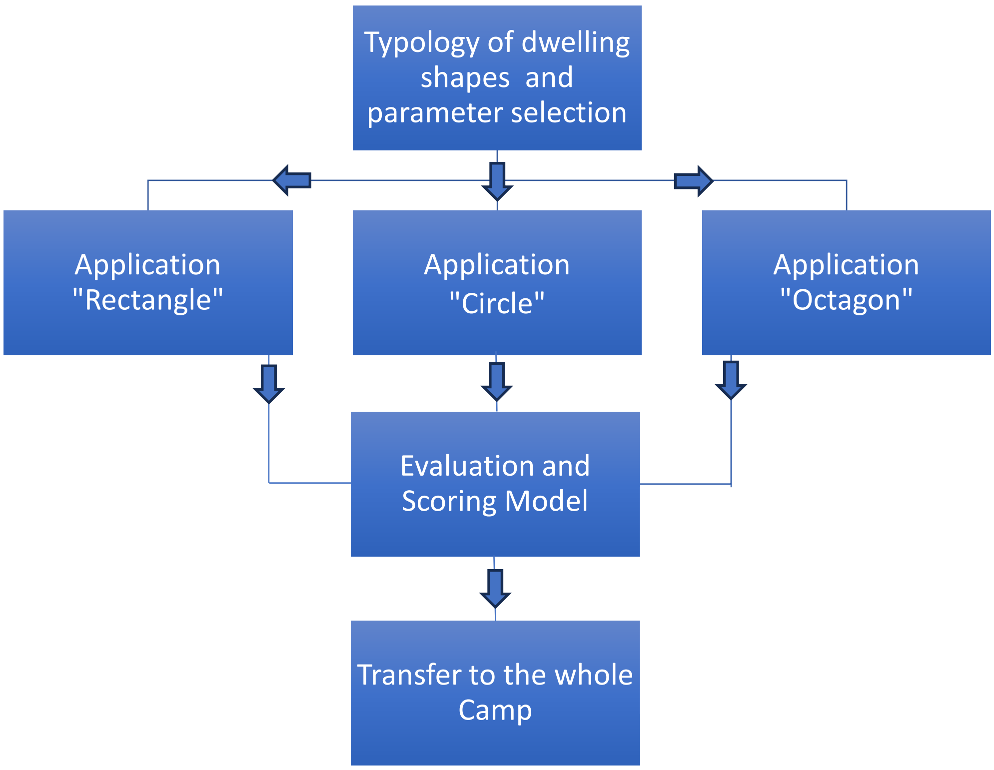


Figure 28: Workflow of the application

3.4.2. Typology of dwelling shapes

As a preliminary to the ‘Regularize Building Footprint’ (RBF) function it is important to determine which angles are needed for the target outlines. This information will help to choose the right method and parameters during the further studies. The algorithm is specifically designed for explicit angles like 90° to lead to orthogonal results. But also, other angles and circles can be generated with this tool.

The first step is to group the dwelling classes according to their shape and size. After this the function target shape method (RIGHT_ANGLE (RA), CIRCLE (C), ANY_ANGLE (AA) or RIGHT_ANGLE_AND_DIAGONALS (RAaD)) can be applied for each dwelling class accordingly. With the information from chapter 3.2.2. the objects are grouped by their special angles. In addition, the rectangles are separated into large (type 1) and small dwellings (type 2). All rectangular forms with more than 4 edges are included in type 3. Type 4 are the Tukuls with the target shape method CIRCLE and type 5 are the Tunnel Shaped Dwellings with the target shape method ANY_ANGLES. Because of the visual interpretation, this step is carried out quickly. The types are shown in Table 3.

Table 3: Typology of the dwelling classes

Type 1 **RIGHT ANGLES**	Type 2 **RIGHT ANGLES**	Type 3 **RIGHT ANGLES**	Type 4 **CIRCLE**	Type 5 **ANY ANGLES**
Bright Dwellings	Small Bright Dwellings	Dark Dwellings	Tukuls	Tunnel Shaped Dwelling
Facility Buildings	Small Dark Dwellings			

3.4.3. Application

Types 1 + 2 + 3

During the pre-categorization of the rectangle dwellings, it is obvious that this groups are orthogonal. In combination with the target shape method RIGHT_ANGLE, a tolerance of 0.5, 1 and 2 with the same densification for each is set for types 1 and 2. The tolerance is determined from the differences of the reference data to the extracted outline. In this case the minimum deviation is 0.5 m. On the other hand, the maximum disagreement (blue arrow) between the references and the original outlines is ca. 2 m (cf. Figure 29). To be able to evaluate an intermediate value tolerance 1 is also chosen.

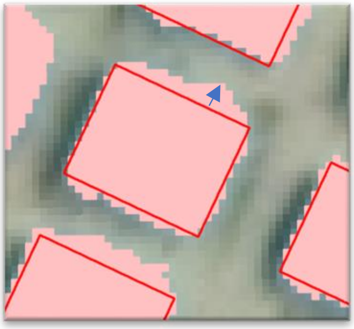


Figure 29: Disagreement of the reference (red line) and extracted data (ca. 2 m)

In contrast to types 1 and 2 more than four angles are wanted for type 3. In this case advanced considerations must be made. In order to get more edges and details, it has been additionally used a densification of 0.5 for the tolerance 1 and 2. The shorter lengths of the adjusted lines allow more corners and edges.

Types 4

For type 4 the target shape method CIRCLE is used. It is important to know that the resulting area adapts to the area of the extracted polygons. Therefore, it is crucial to enable a circle with the same area through the tolerance. In worst case, the tolerance is too small, and the circumference of the circle cannot adjust to the original area. This will be tested with a tolerance of 0.5, 1 and 2.

Furthermore, it can be chosen that only certain areas within a range are going to be edited. However, this is not desired because all objects have to be processed. In this case the standard parameter value 0.1 to 1000000 is used.

It should be noted, that in terms of the entire camp, it is important to find out if there are any other round objects than tukuls. In the approach to the algorithm, it is possible to discover mathematically whether the extracted objects are round or not. Therefore, the compactness ratio can be calculated. It is computed as following:

$$(4 * \pi * area) \,/\, length^2$$

Usually, round buildings are represented by a value of 1. Nevertheless, it can be assumed that the closer to 1, the more likely it is to be a round object. With this information, the compactness ratio is tested on the entire camp to determine if further buildings in other classes than tukuls are circles or not. In case of the extracted building outlines none of the object has a compactness ratio equal to 1. With the marked object in Figure 30, there is a building which is clearly a circle. This object has a compactness ratio of 0.61. But even clear rectangles in the entire camp have a value like this. As well the Tukuls which are known that they are round have a smaller or an equal ratio. However, for this thesis, the calculation of the compactness ratio does not give a clear answer. Therefore, it is not possible to determine other round objects mathematically in order to adapt the function in a better way.

Figure 30: Clear circle with a compactness ratio of 0.61.

Types 5

With the assumption that the objects of type 5 have an octagonal form the target shape method RIGHT_ANGLES cannot work. RIGHT_ANGLES_AND_DIAGONALS does also not offer the necessary preconditions because the objects have angles which are bigger than 90° (cf. Figure 31). Therefore, the method ANY_ANGLES is selected. In this case, the diagonal penalty is the important parameter and must be determined in a first step. With this setting it can be regulated, which angles are used. With a smaller value the likelihood to generate any angles is higher. On the other side a higher value is going to prefer angles as 45° or 90°. Therefore, the values of 0, 0.1 and 0.5 are used. In addition, the tolerance must also be adjusted if the angles are not as desired. In this case the tolerance is modified to 0.25, 0.5 and 1. Since the distance to the reference data is also reduced, the tolerance is kept smaller. Thus, different parameters are tested to determine the chances for an octagonal form.

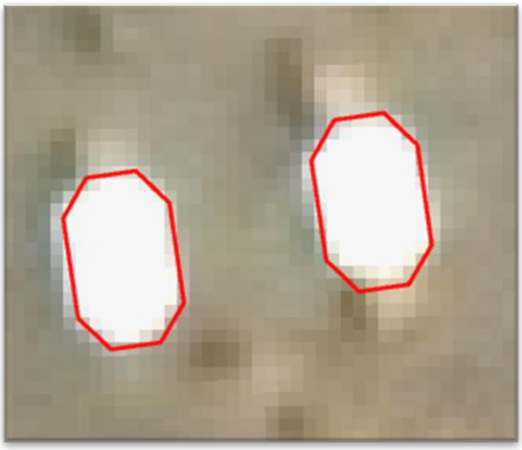


Figure 31: The octagonal shape with angles > 90°

3.4.4. Overview of the parameters

In summary, a tolerance of 0.5 m, 1 m and 2 m is chosen for larger dwellings and 0.25 m, 0.5 m and 1 m for smaller dwellings. These were determined based on the smallest and greatest distance from the original lines to the reference data. The densification is the same as the tolerance. For dark objects, also a smaller densification is additionally chosen because they have more than four edges. For Tukuls a tolerance of 0.5, 1 and 2 is used. In the last types, a low diagonal penalty in combination with a densification of 0,25 and tolerances like 0.25, 0.5 and 1 are selected for the tunnel class to check if the shape fit to the reference. The following Figure 32 gives an overview of the settings:

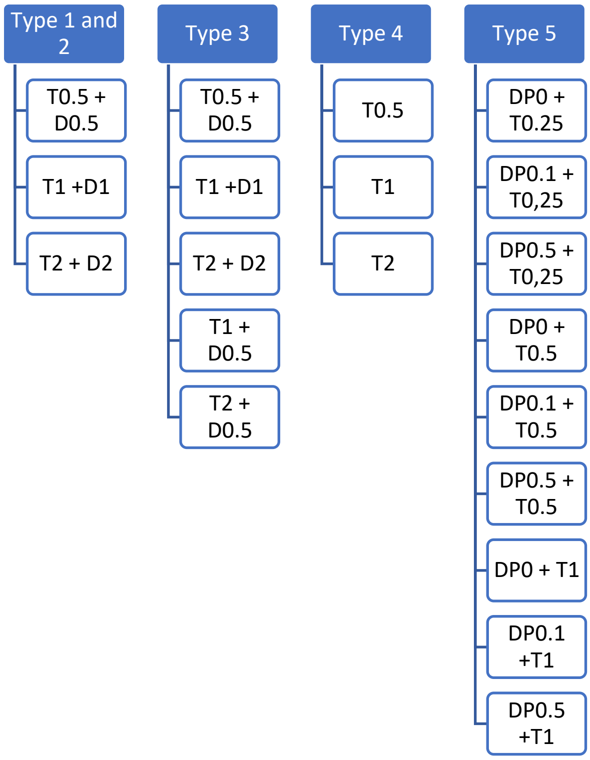


Figure 32: Overview of the tolerance, densification, and diagonal penalty values for each type (T = Tolerance, D = Densification, DP = Diagonal Penalty)

The further setting “precision” plays a minor role in this thesis and is checked in a last step after finding the best settings for the parameters. As the objects are very small and the accuracy of the image is too low, it is not necessary to align the results to cadastral points. Therefore, at the end of the scoring model it is used for completeness to test which effect this parameter has on the results.

4. Results

4.1. Results of the ‘Regularize Building Footprint’ function

In the following section, a comparison of the results of the RBF and the references is made. For all objects the differences between the shape, orientation, area, as well as perimeter and the respective tolerances (T), densification (D) and diagonal penalty (DP) are assessed (cf. Table 4).

Table 4: Results of the regularized objects and the disagreement to the reference data

	Method	Orientation RBF in °	Perimeter RBF in m	Area RBF in m²	Dif. Orientation in °	Dif. Perimeter in %	Dif. Area in %
Bright Dwellings (Without the errors)	T0.5+D0.5	75.93	19.02	21.84	-0.94	7.50	16.22
	T1+D1	76.27	19.33	22.98	-0.60	9.23	22.27
	T2+D2	77.11	19.61	23.57	0.24	10.83	25.42
Bright Dwellings	T0.5+D0.5	85.32	19.76	23.26	8.64	10.42	20.69
	T1+D1	86.23	19.66	23.89	9.55	9.82	23.96
	T2+D2	77.25	20.15	25.08	0.56	12.59	30.09
Facility Buildings	T0.5+D0.5	111.46	42.18	102.78	-3.31	10.72	14.72
	T1+D1	112.76	40.76	102.86	-2.00	7.01	14.80
	T2+D2	108.36	41.44	106.50	-6.41	8.77	18.87
Small Bright Dwellings	T0.5+D0.5	166.90	10.96	6.45	4.78	-5.93	-13.74
	T1+D1	168.52	10.89	6.23	6.40	-6.54	-16.71
	T2+D2	126.22	11.09	5.88	-35.90	-4.79	-21.40
Small Dark Dwellings	T0.5+D0.5	65.54	8.93	4.99	-76.57	1.19	2.54
	T1+D1	60.84	8.94	5.01	-81.27	1.39	3.06
	T2+D2	36.27	9.05	4.65	-105.84	2.55	-4.40
Dark Dwellings	T0.5+D0.5	112.17	17.39	14.96	12.53	7.53	-3.45
	T1+D1	113.00	16.26	15.20	13.35	0.54	-1.90
	T2+D2	110.17	16.05	15.00	10.52	-0.77	-3.19
	T1+D0.5	108.98	16.26	15.23	9.33	0.54	-1.70
	T2+D0.5	108.43	16.26	15.30	8.78	0.54	-1.25
Tukuls	T0.5	0.00	16.71	13.26	0.00	31.40	2.04
	T1	0.00	13.69	13.26	0.00	7.59	2.04
	T2	0.00	12.86	13.26	0.00	1.14	2.04
Tunnel Shaped Dwellings	T0.25+DP0	164.79	21.19	29.32	-8.66	9.54	10.78
	T0.5+DP0	151.16	21.79	29.50	-22.29	12.66	11.48
	T1+DP0	107.52	23.49	29.39	-65.93	21.42	11.05
	T0.25+DP0.1	165.38	21.41	29.35	-8.07	10.68	10.92
	T0.5+DP0.1	150.48	21.77	29.44	-22.97	12.52	11.25
	T1+DP0.1	107.52	23.49	29.39	-65.93	21.42	11.05
	T0.25+DP0.5	158.29	21.50	29.37	-15.16	11.13	10.98
	T0.5+DP0.5	150.48	21.75	29.40	-22.97	12.43	11.09
	T1+DP0.5	107.52	23.49	29.39	-65.93	21.42	11.05

First of all, the results are interpreted by a visual analysis. In this context, the quality of the shape, but also the quantity of the edited objects is evaluated. In the next step, the orientation, the area, and the perimeter are compared with the values from the reference data (cf. Table 4). The comparisons are based on the mean property values of the 20 sample dwellings per dwelling class described in chapter 3.2.2. Afterwards, the best results are determined by using the scoring model. Finally, the best settings are applied to the entire area and checked by visual interpretation and by calculating the number of unprocessed features.

Type 1 - Bright Dwellings

Along with the Facility Buildings the Bright Dwellings are one of the largest dwellings in this camp. The reference object of the class has an area of 19.28 m², a perimeter of 17.90 m and an orientation of 76.68°. The resulting objects differ visually only slightly from the references. In the cases with tolerances of 1 and 2, all objects have the desired shape. With a look at the results more closely, it can be seen that with T0.5 a total of 13 buildings are rectangles. Seven show unwanted details, so only 65% of results have the correct form (cf. Figure 33 – left image).

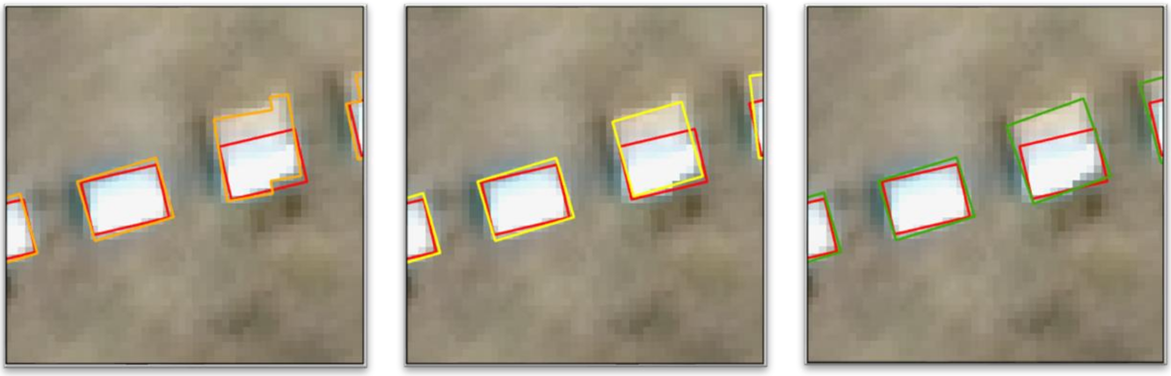


Figure 33: Results of the Bright Dwellings
(red = reference, orange = T0.5; yellow = T1; green = T 2)

It is noticeable, that some objects within each parameter are significantly larger than the reference (cf. Figure 33). This is due to the fact that through the semi-automatic image analysis outbuildings or areas next to the tent are classified as a dwelling and therefore are incorrectly outlined. In total five buildings are identified, which are affected of this problem (cf. Figure 34). In this case the resulting polygon is bigger.

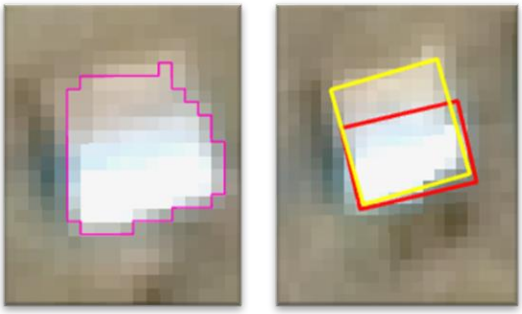


Figure 34: Incorrect extraction during OBIA (left image),
reference = red and resulting outline = yellow polygon (right image)

An example for good approximation can be seen in case of the wrong extraction due trees. One building is partially covered by a tree, which results in a small notch in the extraction (cf. Figure 35 left image). With all tolerances this is compensated positively (cf. Figure 35 right image).

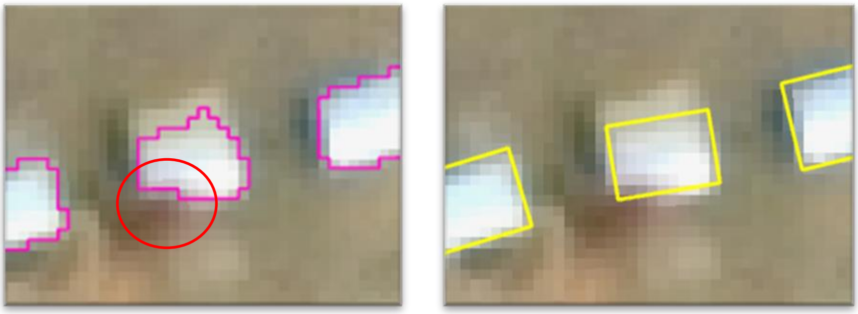
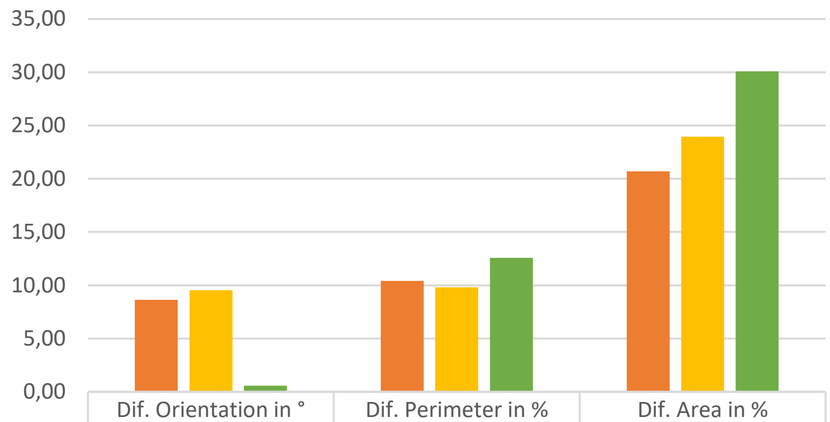


Figure 35: Correction of artefacts

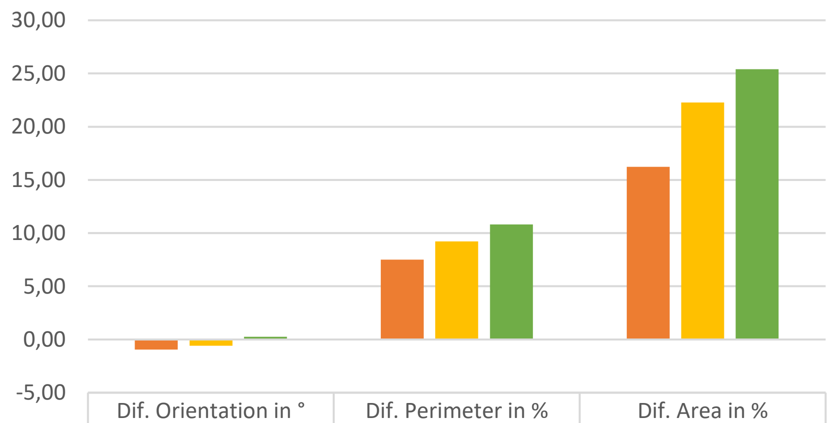
The statistical results **Fehler! Verweisquelle konnte nicht gefunden werden.**show that the d eviation of the orientation from the reference data is marginal for all three parameter settings (cf. Chart 1). T2 has the smallest disagreement with a difference of 0.56°. The other two have a slightly bigger difference in the orientation. Because of the five buildings mention above, in sum all tolerances have a taller shape. However, this confirms the visual assessment described above. With T2 the objects have with 25.08 m² the biggest area. This is in average 30.09% larger than the reference data with 19.28 m². In contrast, the difference decreases slightly at T1 with +23.96% and with +20.69% by using T0.5. In connection with the perimeter, a different picture appears. The average deviation is 1,76 m at T1, which corresponds to an increase of +9.82%. In addition, the disagreement at T0.5 is somewhat higher with 1,86 m (= +10.42%) and at T2 with 2,25 m (= +12.59%).



	Dif. Orientation in °	Dif. Perimeter in %	Dif. Area in %
T0,5+D0,5	8,64	10,42	20,69
T1+D1	9,55	9,82	23,96
T2+D2	0,56	12,59	30,09

Chart 1: Disagreement of the Bright Dwellings

If the five buildings that are mistakenly extracted are excluded, the result shows a smaller deviation of the area and the perimeter. This can be seen in Chart 2**Fehler! Verweisquelle konnte nicht gefunden werden.**. However, the perimeter differs between +7.5% and +10.83% and the area between +16.22% and +25.42% to the reference data. In sum T2 shows the largest disagreements and T0.5 the smallest. Thus, the orientation would be +-1° for all three objects. Since the same initial data were used, this is less relevant for the comparison of the parameter settings.



	Dif. Orientation in °	Dif. Perimeter in %	Dif. Area in %
T0,5+D0,5	-0,94	7,50	16,22
T1+D1	-0,60	9,23	22,27
T2+D2	0,24	10,83	25,42

Chart 2: Disagreement of the Bright Dwellings without the wrong extracted outlines

Type 1 - Facility Buildings

These buildings are the largest ones in the camp with a reference object orientation of 114.77°, a perimeter of 38.09 m and an area of 89.60 m². In this case, all results show objects which are larger than the reference object. Looking at the three parameters, tolerance 2 comes closest to the reference object in terms of shape. At the end 100% of the resulting objects have the desired shape. At T1only two buildings have additional small notches. With T0.5 every building has some unwanted noise (cf. Figure 36).

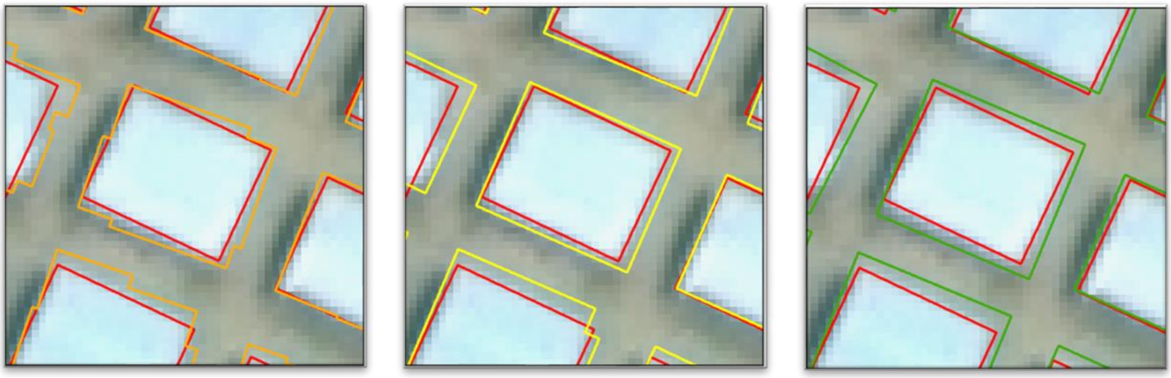
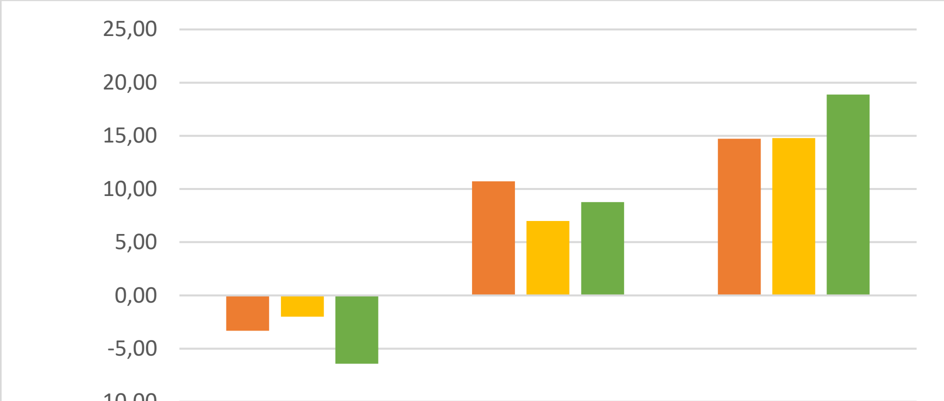


Figure 36: Results of the Facility Buildings
(red = reference, orange = T0.5; yellow = T1; green = T 2)

The statistics show similar behavior for the largest buildings (cf. Chart 3). The orientation indicates only a minimal deviation from the reference data. For tolerance 2 it is -6.41°. For tolerance 1 the disagreement is even smaller (= -2°) and for tolerance 0.5 it is -3.31°. All in all, they are close to the reference data. Between the tolerances there are no significant differences in the orientation. In the case of the area, they show more bigger disagreements. T2 has the largest difference. In sum the results have an average area of 106.50 m² (= +18.87%) and are slightly bigger than the other two. In terms of the perimeter, they are about the same. Tolerance 1 has the largest circumference with 40.76 m (= +7.01%), followed by T2 and T0.5 with a slightly smaller difference of +8.7% and +7.01%.



	Dif. Orientation in °	Dif. Perimeter in %	Dif. Area in %
T0,5+D0,5	-3,31	10,72	14,72
T1+D1	-2,00	7,01	14,80
T2+D2	-6,41	8,77	18,87

Chart 3: Disagreement of the Facility Buildings

Type 2 - Small Bright Dwellings

In this next section, the smaller buildings are inspected. The corresponding reference object has an area of 7.47 m² and a perimeter of 11.65 m. The orientation is 162.12°. When the results are compared per visual interpretation, it becomes clear that with T2 25% of the bright building outlines are not altered. They have the same shape as before. All other objects show the desired rectangular shape. On the other hand, with the tolerances of 0.5 and 1, 100% of the resulting objects show rectangular shapes without errors (cf. Figure 37).

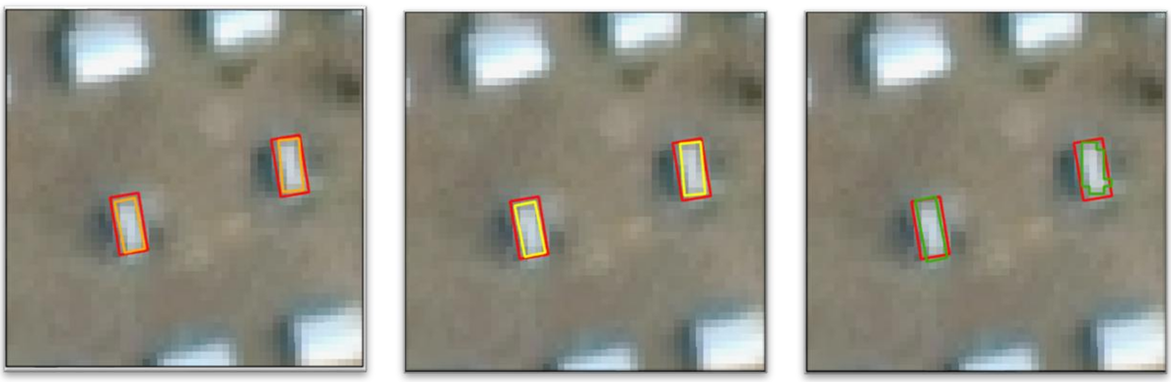


Figure 37: Results of the Small Bright Dwellings

(red = reference, orange = T0.5; yellow = T1; green = T 2)

During the inspection of the statistic values, it is immediately clear that the alignment of the Small Bright Dwellings with tolerance 2 deviates significantly from the reference properties. They have in sum a deviation of -35.90° which corresponds to an average orientation of 126.22°. This is due to the five objects which were not generalized. Because the orientation for the resulting outlines is calculated with the minimum bounding box, this results in an orientation of 0° (cf. Figure 38). The reference object has an orientation of 162.12°, which leads to a significant distortion of the results. Considering the orientation of the other two tolerances they have a minimum disagreement. Tolerance 1 shows a slightly larger deviation of +6.40° and about 168.52° in comparison with T0.5 with a plus of 4.78° and 166.90°.

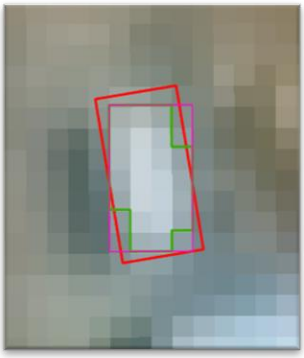
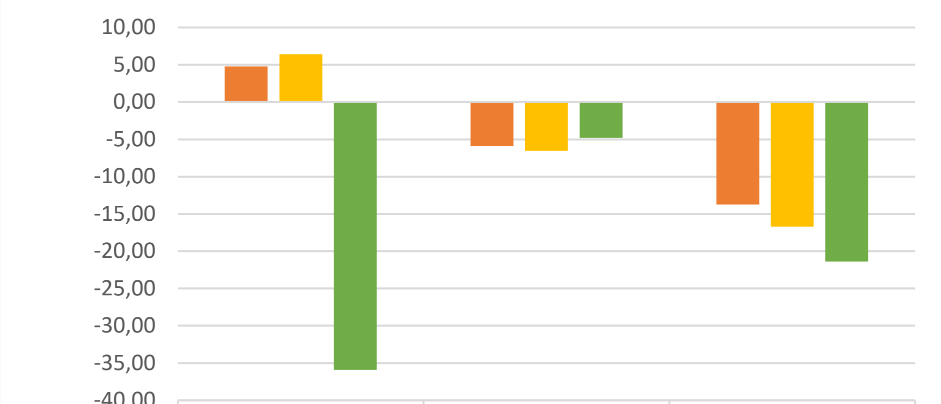


Figure 38: Orientation of the MBB (purple box); red = reference object, green = original outline

In contrast to the large buildings, the statical values show for the small one’s different structures (cf. Chart 4**Fehler! Verweisquelle konnte nicht gefunden werden.**). All three results have a smaller area than the reference. The larger the tolerance, the smaller the area and perimeter. Tolerance 0.5 has the smallest area with 6.45 m² (= -13.74%), followed by T1 with -16.71%, which corresponds to an area of 6.23 m² and T2 with 5.88 m² (= -21.40%). The disagreements from the areas relate similar to the perimeters. They are also smaller than the reference. There are only minimal significant differences in the perimeter with the different tolerances (T2 = -4.79% and 11.09 m perimeter; T1 = -6.54% and 10.89 m perimeter; T0.5 = -5.93% and 10.96 m perimeter).



	Dif. Orientation in °	Dif. Perimeter in %	Dif. Area in %
T0,5+D0,5	4,78	-5,93	-13,74
T1+D1	6,40	-6,54	-16,71
T2+D2	-35,90	-4,79	-21,40

Chart 4: Disagreement of the Small Bright Dwellings

Type 2 - Small Dark Dwellings

This is the smallest class with a reference object area of 4.86 m², a circumference of 8.82 m and an orientation of 142.11°. The visual interpretation shows that with a tolerance of 0.5 and 1 all objects have the determined rectangle or almost a square shape. 30% do not have a rectangular shape and contain clear errors with tolerance 2 (cf. Figure 39).

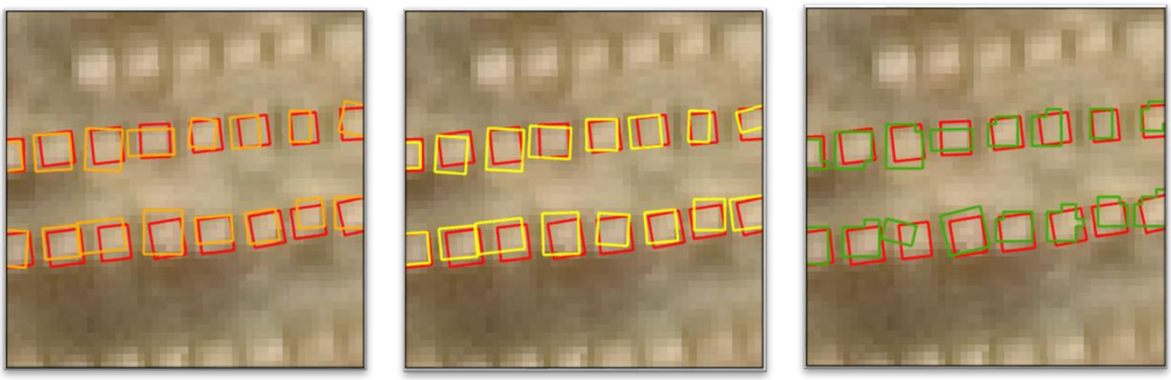
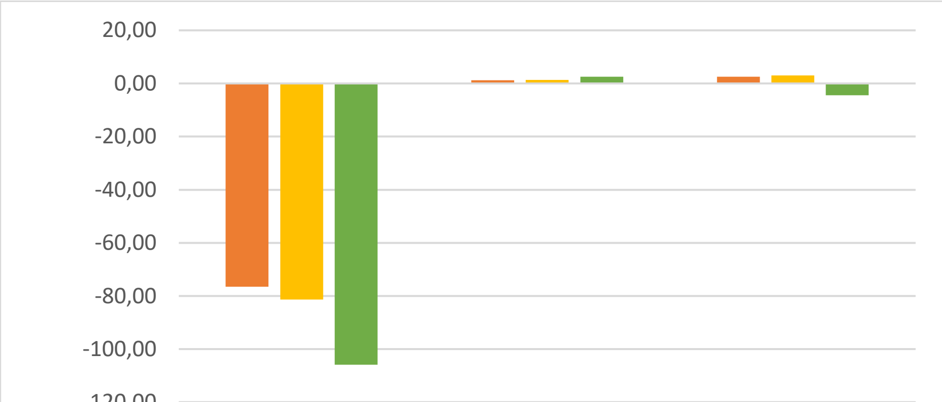


Figure 39: Results of the Small Dark Dwellings
(red = reference, orange = T0.5; yellow = T1; green = T 2)

In terms of orientation, all tolerances show significant deviations. Tolerance 2 appears with the largest disagreement of -105.84°, followed by tolerance 1 with -81.27° and tolerance 0.5 with -76.57°. Again, the large deviations are the result of the MBB. Due to the almost square shapes, it can happen that the sides of the object differ minimally. This means that the measurement of the longer side may be oriented to the wrong side. This can lead to disagreements with 90°. The further assessment of the small dark areas shows an average deviation of the area and circumference in minimum range with less than +-5 % (cf. Chart 5).



	Dif. Orientation in °	Dif. Perimeter in %	Dif. Area in %
T0,5+D0,5	-76,57	1,19	2,54
T1+D1	-81,27	1,39	3,06
T2+D2	-105,84	2,55	-4,40

Chart 5: Disagreement of the Small Dark Dwellings

Type 3 - Dark Dwelling

These buildings are the only ones which have not only a rectangle shape. Some of them have further edges and differ in size and orientation. Once again, it should be noted that the visual interpretation can only be an approximation and an exact statement about the shape is not possible. In addition, they have an average orientation of 101.13°, a perimeter of 16.81 m and an area of 16.58 m². With T0.5+D0.5, only 3 objects have an appropriate shape. Visually considered, with tolerance 1 and 2 and the same densification 17 dwellings have the same shapes as the references but differ sometimes in the orientation. This is also the case with T1 and T2 in combination with a densification of 0.5. In comparison, with T1 and T2 all objects are edited and show better results than with T0,5. At the end they still do not match with the reference in some terms of the additional edges (cf. Figure 40).

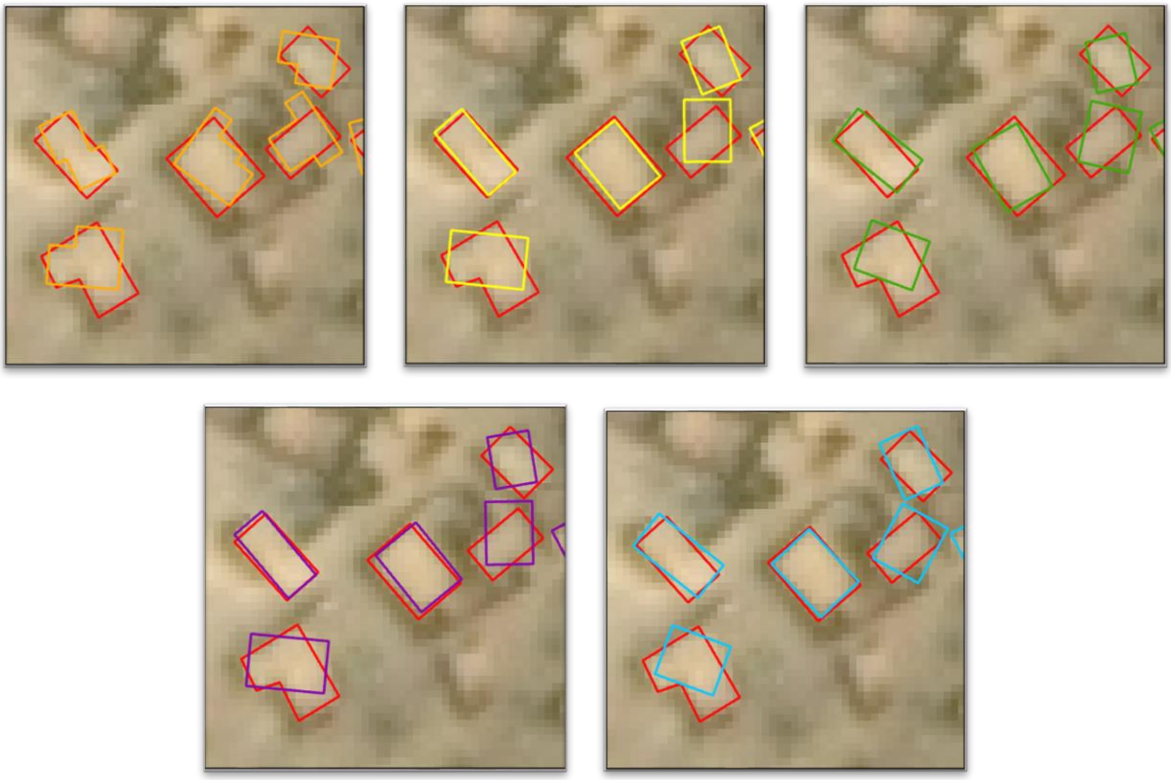
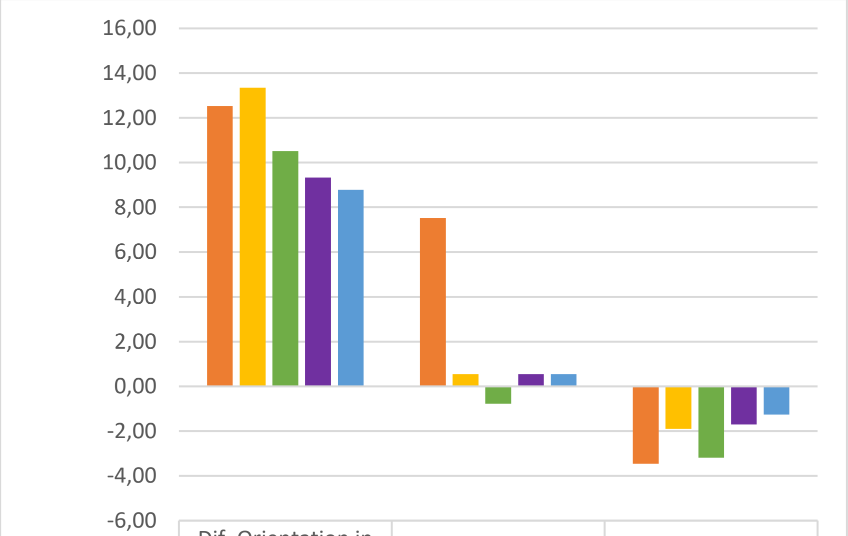


Figure 40: Results of the Dark Dwellings (orange = T0.5+D0.5; yellow = T1+D1; green = T2+D2; purple = T1+D0.5; blue = T2+D0.5; red = reference)

In sum there are found significant disagreements from the reference values in the orientation. T2+D0.5 has the lowest and T1+D1 the biggest deviation. In this case there are a number of objects that deviate strongly from the reference and contribute to this result. The slightly larger dark buildings behave similarly to the Small Bright Dwellings in context to the area and perimeter. Furthermore, all objects have a very small difference with less than -5% for the area. Also, the perimeter differs only slightly, with just a minimal disagreement of +-5% for all parameters. Only the deviation for T0.5 with D0.5 is significant larger (cf. Chart 6).



	Dif. Orientation in °	Dif. Perimeter in %	Dif. Area in %
T0,5+D0,5	12,53	7,53	-3,45
T1+D1	13,35	0,54	-1,90
T2+D2	10,52	-0,77	-3,19
T1+D0,5	9,33	0,54	-1,70
T2+D0,5	8,78	0,54	-1,25

Chart 6: Disagreement of the Dark Dwellings

Type 4 - Tukuls

The reference circle has a diameter of 4.03 m, a circumference of 12.72 m and an area of 13.00 m². The visual inspection shows that with T0.5 only 15% of the objects were edited, in contrast to the tolerance 1 with in sum 85% and T2 with 100% modified objects (cf. Figure 41).

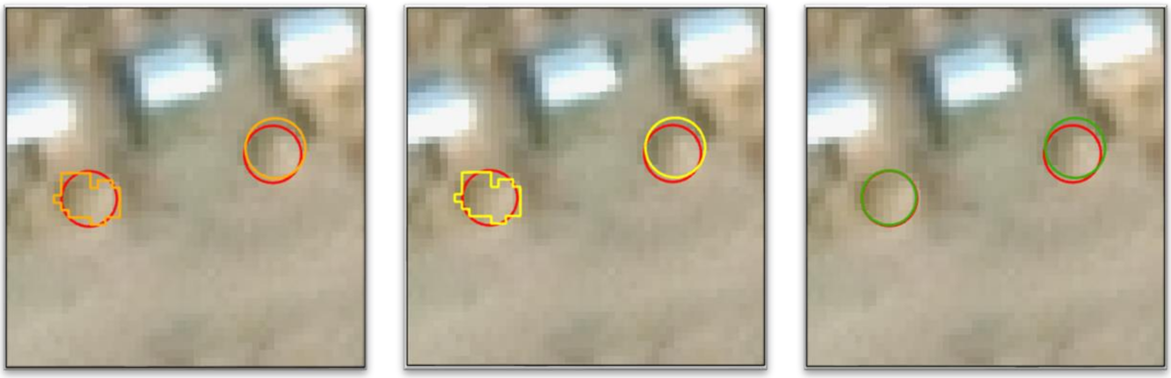


Figure 41: Results of the Tukuls (red = reference, orange = T0.5; yellow = T1; green = T 2)

For the areas of the Tukuls the statistical results show that they deviate only slightly from the reference data with all tolerances (cf. Chart 7). In this case the reference differs only minimal from the extracted outline which have an area of 13.26 m². Therefore, the disagreement is only +2.04% from the reference. The outlines with 16.71 m, 13.69 m and 12.86 m for T0.5, T1 and T2 (T0.5 = +31.4%; T1 = +7.59%; T2 = +1.14%) show also disagreements with all three tolerances. In this case the T0.5 presents a larger difference than the other two. The reason for this is that not all objects with these tolerances are processed.

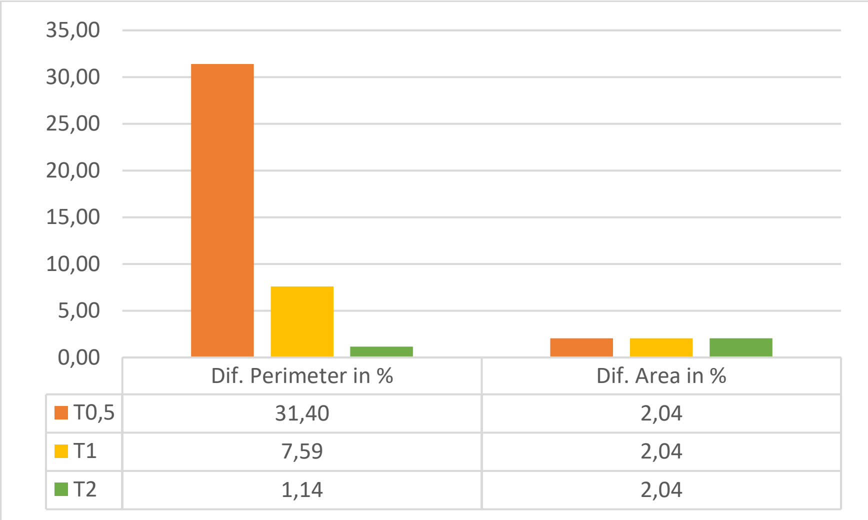


Chart 7: Disagreement of the Tukuls

Type 5 - Octagon

During the inspection it is noticeable that all objects still do not have an octagonal form. The closest to the reference object is the function with a diagonal penalty of 0 independent of the tolerance. With a diagonal penalty of 0.1, there are clearly more objects that have some right angle. At a diagonal penalty of 0.5, the disagreement increases significantly. These forms have sharp angles, which are also undesirable in this case. A high deviation is identifiable with all tested values for each parameter (cf. Figure 42). The tolerance does have a minor impact to the shape. More important is the diagonal penalty which influences the shape significant.

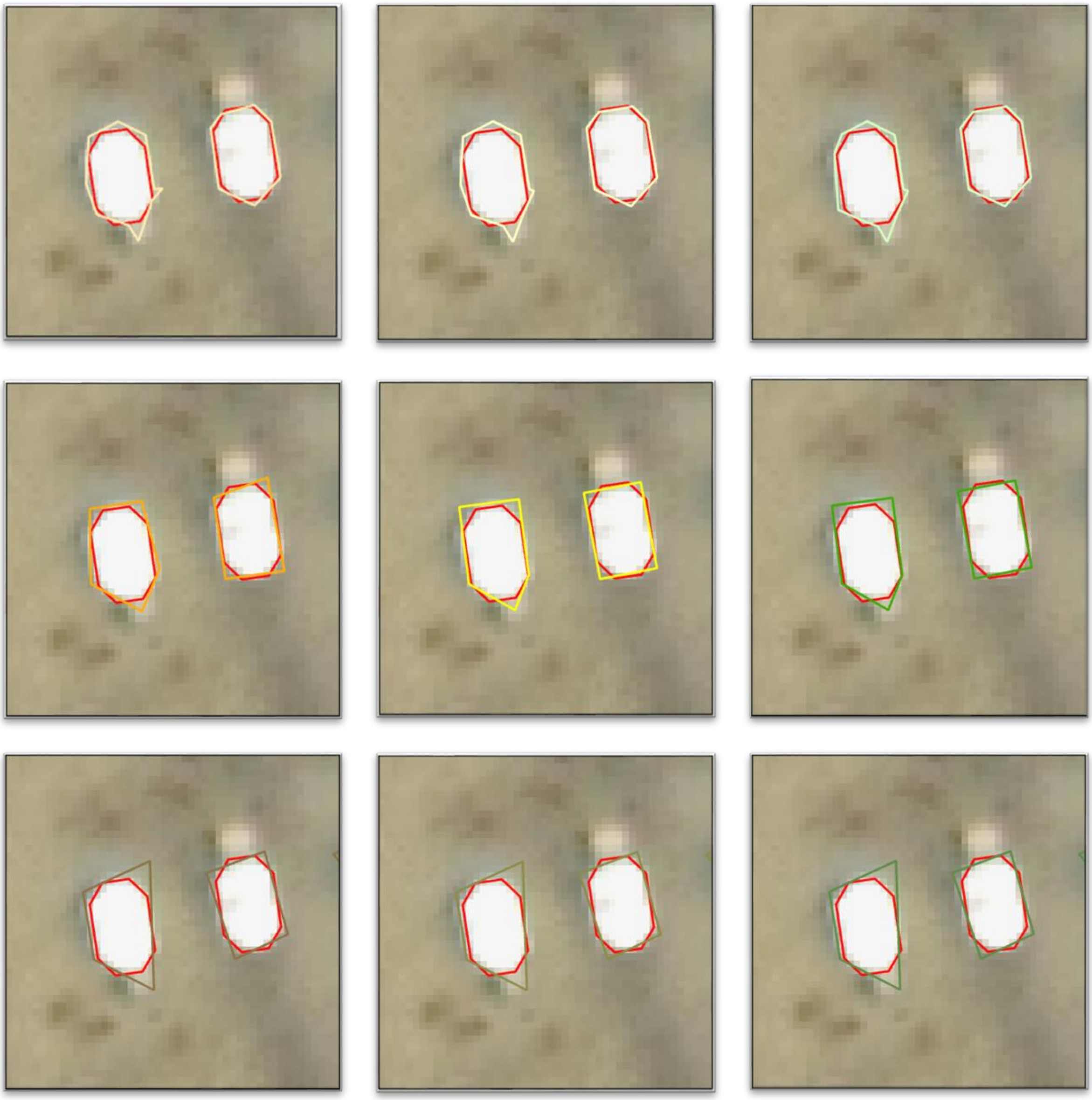
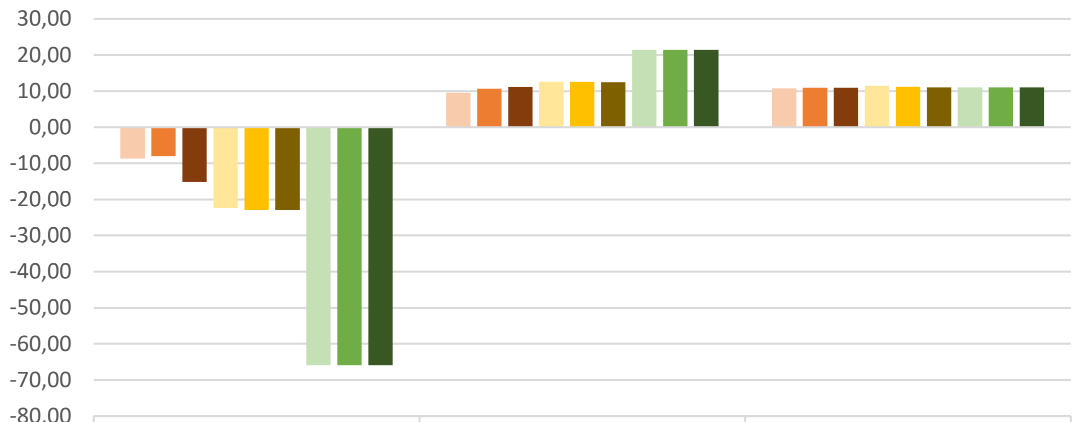


Figure 42: Results of the Tukuls (upper row left = DP0+T0.25, middle = DP0+T0.5, right = DP0+T1;middle row left = DP0.1+T0.25, middle = DP0.1+T0.5, right= DP0.1+T1; upper row left = DP0.5+T0.25, middle = DP0.5+T0.5, right = DP0.5+T1; red = reference)

For the last type, the orientation of the minimum bounding box of the reference is 173.45°. The perimeter is 19.35 m, and the area is 26.46 m². The statistical values show strong disagreements in the orientation as well as in surface and circumference (cf. Chart 8). The higher the tolerance the higher is the deviation in this case. In sum with a diagonal penalty of 0 and a low tolerance the smallest disagreement of the result is identified. This is verified again in connection with the clearly unwanted shapes.



	Dif. Orientation	Dif. Perimeter	Dif. Area
T0,25+DP0	-8,66	9,54	10,78
T0,25+DP0,1	-8,07	10,68	10,92
T0,25+DP0,5	-15,16	11,13	10,98
T0,5+DP0	-22,29	12,66	11,48
T0,5+DP0,1	-22,97	12,52	11,25
T0,5+DP0,5	-22,97	12,43	11,09
T1+DP0	-65,93	21,42	11,05
T1+DP0,1	-65,93	21,42	11,05
T1+DP0,5	-65,93	21,42	11,05

Chart 8: Disagreement of the Tunnel Shaped Dwellings

4.2. Scoring Model

After the individual inspection the best value for each parameter per dwelling class is determined. In the scoring model each result is assessed by its properties, afterwards the judged values are weighted and summed. The respective settings are evaluated according to the scoring matrix. In Table 5 the points are presented, and the green values determine the parameter combination with the best score.

Table 5: Results of the scoring model

	Tolerance/ DP	Area	Perimeter	Orientation	Shape	Σ
		10%	10%	30%	50%	
Bright Dwellings	RA T0.5 D0.5	0.1	0.3	1.2	2	3.6
	RA T1 D1	0.1	0.4	1.2	2.5	4.2
	RA T2 D2	0.1	0.3	1.5	2.5	4.4
Facility Buildings	RA T0.5 D0.5	0.3	0.3	1.5	1	3.1
	RA T1 D1	0.3	0.4	1.5	2.5	4.7
	RA T2 D2	0.2	0.4	1.2	2.5	4.3
Small Bright Dwellings	RA T0.5 D0.5	0.3	0.4	1.5	2.5	4.7
	RA T1 D1	0.2	0.4	1.2	2.5	4.3
	RA T2 D2	0.1	0.5	0.3	2	2.9
Tukuls	RA T0.5	0.5	0.1	-	1	1.6
	RA T1	0.5	0.4	-	2.5	3.4
	RA T2	0.5	0.5	-	2.5	3.5
Dark Dwellings	RA T0.5 D0.5	0.5	0.4	0.9	1	2.8
	RA T1 D1	0.5	0.5	0.9	2.5	4.4
	RA T2 D2	0.5	0.5	0.9	2.5	4.4
	RA T1 D0.5	0.5	0.5	1.2	2.5	4.7
	RA T2 D0.5	0.5	0.5	1.2	2.5	4.7
Small Dark Dwellings	RA T0.5 D0.5	0.5	0.5	0.3	2.5	3.8
	RA T1 D1	0.5	0.5	0.3	2.5	3.8
	RA T2 D2	0.5	0.5	0.3	2	3.3
Tunnel Shaped Dwellings	AA T0.25 DP0	0.3	0.4	1.2	1	2,9
	AA T0.25 DP0.1	0.3	0.3	1.2	1	2.8
	AA T0.25 DP0.5	0.3	0.3	0.6	1	2.2
	AA T0.5 DP0	0.3	0.3	0.3	1	1.9
	AA T0.5 DP0.1	0.3	0.3	0.3	1	1.9
	AA T0.5 DP0.5	0.3	0.3	0.3	1	1.9
	AA T1 DP0	0.3	0.1	0.3	1	1.7
	AA T1 DP0.1	0.3	0.1	0.3	1	1.7
	AA T1 DP0.5	0.3	0.1	0.3	1	1.7

For the Bright Dwellings, T2 shows the best result with 4.4 points. The orientation of T2 is minimally better than T1, which has with 4.2 points also a good result. For the largest dwellings, the Facility Buildings, T1 has the best result with 4.7 points. It has an orientation that is minimal better than T2. However, it should be considered that with T2 all objects were edited in contrast to T1. Nevertheless, both have with 4.7 and 4.3 points fine results. For the slightly Smaller Bright Dwellings, the method with T0.5 gives the best results. With a score of 4.7, it is better than T1, which achieves with 4.3 points also a good result. According to the scoring model the Dark Dwellings have an average score of 4.7 with T1 and T2. However, T2 shows a minimal better shape, although not all objects have the right shape. For the Small Dark Dwellings, T0.5 and T1 have the best fit and a score with 3.8 points. At this point the high disagreement of the orientation due to use of MBB must be noted. For the circles, T2 has the best results with 3.5 points, since all objects are processed. T1 has also a score of 3.4, but it should be noted, that not all objects are processed. In the case of the Tunnel Shaped Dwellings all results do not correspond to the shape and therefore have more edges and less points overall. The best is still the method with a diagonal penalty of 0. At the end it has only 2.9 points with a tolerance of 0.25. All others show 90° angles or triangles and therefore differ greatly in orientation. They have an average score between 1.7 and 2.8.

In addition, the best results are achieved with the highest number of points per class. Comparing these results with the original areas, the following Chart 9 appears.

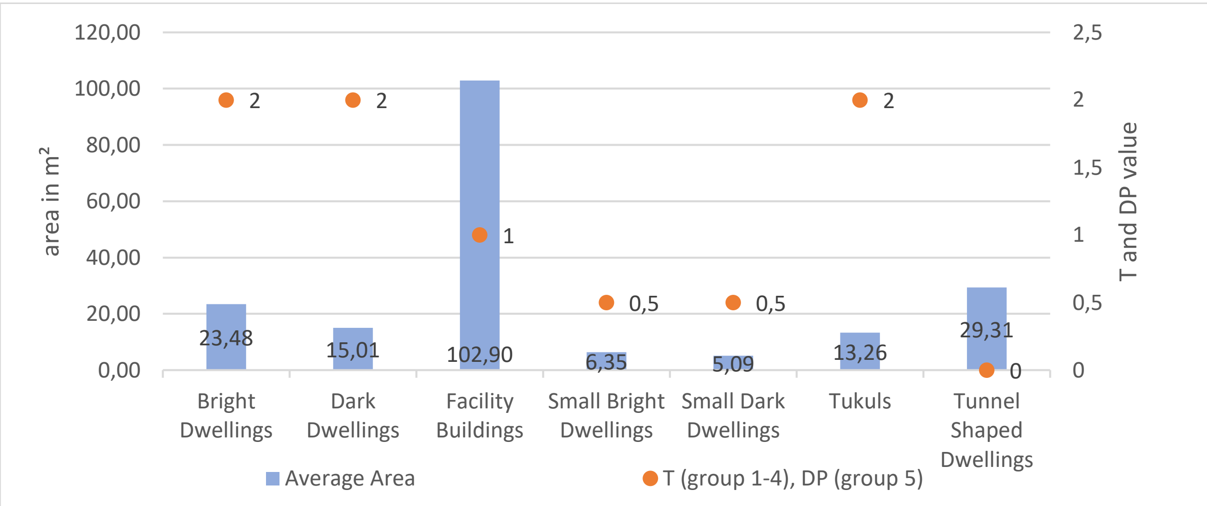


Chart 9: Area in comparison to the best tolerance and diagonal penalty

In summary, for the big Bright Dwellings, Facility Buildings and also for the Dark Dwellings and Tukuls, which have an area around 15 m² and more, the tolerances of 1 and 2 show the best results. Considering the fact, that for the Facility Buildings all objects with T2 have the desired shape and with T1 only 90%, it could be possible, that for the other buildings in the entire camp T2 would lead to a better result. This is because they have a bigger area in the average. This will be reviewed in the next step when the parameters were applied to the entire camp. For Small Bright Dwellings the tolerances 0.5 and for Small Dark Dwellings tolerances with 0.5 and 1 are the best parameter settings. They have both an area with less than 6,5 m² in the extracted data. The best result for Tunnel Shaped Dwellings shows the function with a diagonal penalty of 0. While comparing these results with the original data, it could be found out that the tolerance does only influence the shape in this case minimally.

Improving of the precision

As mentioned above, the function offers another parameter to improve the accuracy. During the last procedure a precision with 0.25 was checked. In this context, the precision for the best results is tested at 0.15 and 0.05 to see what happens with the accuracy. In the results it can be seen that there are only minimal disagreements in the orientation, area and perimeter. An example for Bright Dwellings can be seen in Figure 43.

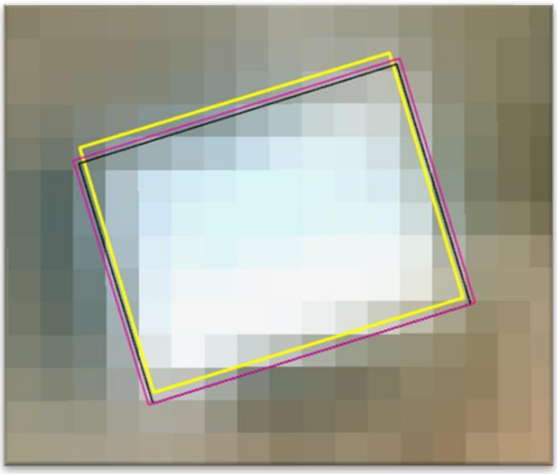


Figure 43: Marginal differences with precision of 0.25 (yellow), 0.15 (red) , 0.05 (blue)

This evaluation leads to the conclusion, that for large objects, which means sizes up to 15 m^2 and more, a large tolerance greater than 1 is necessary to regularize the outlines in a good way. For smaller and more complex ones, a lower tolerance like 0.5 or a diagonal penalty of 0 must be chosen if angles over 90° are wanted. The precision plays in this case a minor role. At the end the parameter for generalization depends on the size of the objects. Therefore, in summary, the smaller the area, the smaller the tolerance should be chosen to achieve good result. Conversely, the parameter to be defined should be also larger for bigger dwellings.

4.3. Transfer to the entire camp area

At the end the best results for each parameter per class are applied to the entire camp area to check whether the new outlines can also be suitable for objects that do not correspond to the reference objects. During the visual inspection, several points are noticed.

Because many objects are not edited in the previous tests, it is useful to check if all outlines are modified. The function creates a new attribute field during the regulating process that identifies objects which are not reduced. In this case in total 5 outlines are not edited. These objects are mainly found in the class of the Small Dark Dwellings. They have a very expansive outliers and are thin at the same time. This has led to the fact, that the objects were not modified (cf. Figure 44).

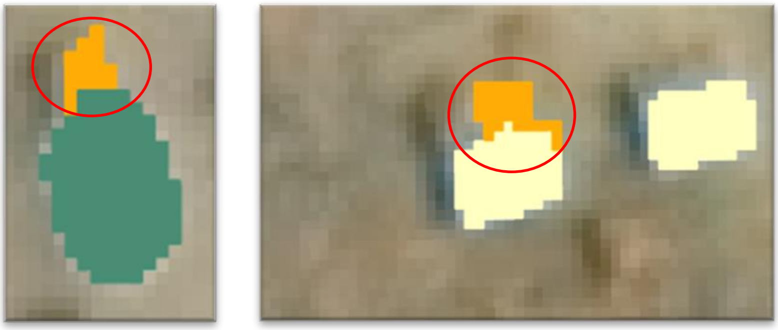


Figure 44: Objects which are not edited (orange areas).

Furthermore, one outline of the Tukuls, was not adapted. This one has a very elongated shape. As a result, the new area deviates significantly from the original shape if the tolerance is too small (cf. Figure 45).

Figure 45: One Tukul which was not edited (green object).

A further important point is mentioned above. In the case of Facility Buildings, some results have notches and unnecessary edges. The original extents of this buildings are much larger than the area of the sample buildings. So, in this case and with the information from the scoring model the higher tolerance 2 presents better results. At the end they show a rectangle shape without unwanted edges (cf. Figure 46).

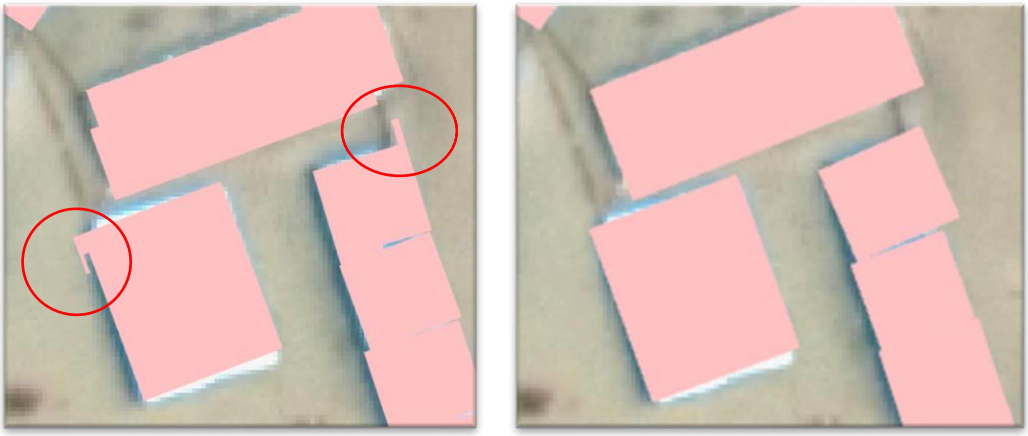


Figure 46: Results with T1 (left) and with T2 (right)

One disadvantage which appears in the entire camp is, that the resulting data contains overlapping objects. The objects which are close to each other's and do not have the same orientation leads into overlays (cf. Figure 47). It is assumed that the buildings connect in real but do not intersect. In the visual inspection, many small dwellings cover larger tents. For example, Small Dark and Dark Dwellings often intersect and correlate with Bright or Tunnel Shaped Dwellings. Also, the Facility Buildings overlap in some cases.

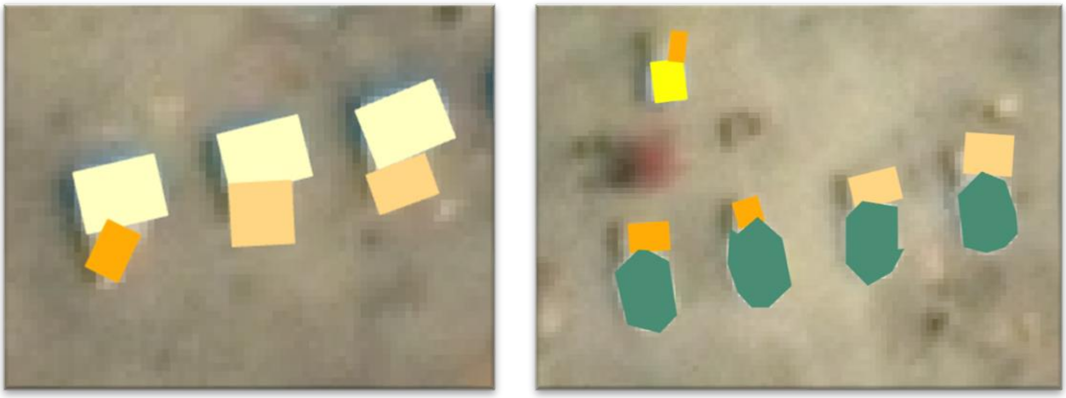


Figure 47: Overlapping buildings

Another point is mentioned above in connection with the compactness ratio. During the workflow, clear round objects in type 1 + 2 + 3 do not have a circular form in the resulting data (cf. Figure 48). In this case the calculation of the compactness ratio to determine round objects in the pre categorization does not work. This leads to the fact, that only a visual interpretation is useful to select the clear circles and to choose the right method for these buildings.

Figure 48: Wrong method for round Facility Buildings

Nevertheless, the transfer of the parameters shows throughout the visual inspection good results. At the end some objects are not edited and the parameter for the Facility Buildings was not optimal. Thus, there are some dwellings which need a deeper interpretation and therefore an adaption of the parameter. But with the information of the scoring model the tolerance can be adapted with a higher or lower tolerance, so that the outlines also show the real shapes.

The resulted dwellings shapes, i.e., the best results for each class, were presented in a map and can be seen in the Appendix 1.

5. Discussion

Overall, this study is challenged by its complex character. While processing the data various classes of different objects and parameters, which need independent handling and judging, are handled with multiple values and results. At the very beginning the input data of the camp imagery and the OBIA dwelling outlines is screened. Because of the diversity of dwellings with different forms and sizes, they require individual supervision for processing. Furthermore, the RBF function with its multiple parameters and corresponding values for setting is researched. The lack of in situ data for comparison of the function results adds to the list of challenges. For this purpose, reference objects are identified as the basis for the validation. Additional limitation is caused by low image resolution and the spectral properties of the objects. Thus, it is partly difficult to visually identify the right size and shape of the reference objects. Moreover, the evaluation of the RBF function results in comparison with the reference objects is challenged by the quantification of the shapes. In conclusion a scoring model is used to assess the disagreement of shape, orientation, area, and perimeter. The scoring model also helps with the consolidation of the evaluation results for each property into one overall score to find the best parameter value combination per dwelling class. However, the scoring model has its limitations and cannot compensate for the lack of a validated comparison base and the difficulties with the measurement of shapes. The outcome of the scoring model highly depends on the set premises and quantified scoring. Therefore, the personal perception and rating of the property weights and judged shapes highly impact the overall score. At the end it was able to transfer the results to the entire camp, but it was also difficult to check the quality.

In conclusion, the 'Regularize Building Footprint' function works was suitable for rectangular and round buildings with different sizes. For this, different tolerances and densification parameter settings are necessary and depend on the size of the dwelling. On the other side for more complex dwellings like octagons, the RBF function is less suitable.

The data analysis and the scoring model in this study show that for rectangular objects the target shape method of RIGHT_ANGLE in combination with a tolerance of 1 and 2 is a good way to generalize the outlines. For Bright Buildings, Facility Buildings and Dark Buildings, the tolerance 2 provides the wanted results. In all cases the shapes and orientations are approximated to the references (cf. Figure 49).

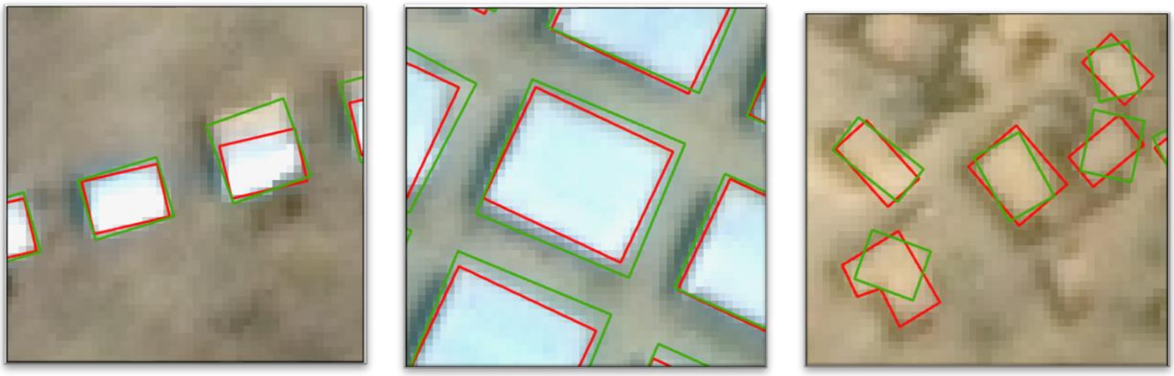


Figure 49: Best parameters for big buildings (red = reference, green = result of the RBF)

For the smaller dwellings a different situation is occurred. The Small Dark and Small Bright Dwellings show better results with a little tolerance of 0.5 (cf. Figure 50). However, the RBF function results depend strongly on the quality of the OBIA extraction results as input. For the Small Dwellings the resolution of the image is too low and in combination with the difficulties from the spectral properties it is not easy to identify the exact size and form of the buildings.

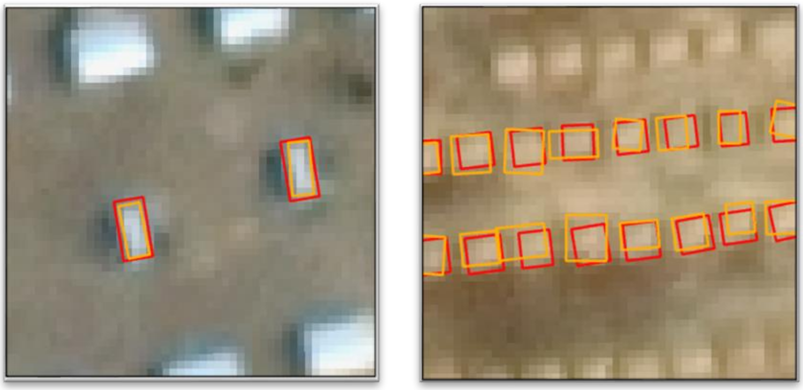


Figure 50: Best parameters for small dwellings (red = reference, orange = result of the RBF)

For round Tukuls, the target shape method CIRCLE in combination with a tolerance of 2 creates also good results. All objects, except one, are edited in the entire camp. For more complex structures, such as octagons, no optimal parameter values for the RBF function could be found. In this case the angles above 90° are not exactly modified through the algorithm. However, it must be noted that due to the resolution and difficult spectral properties, the extracted outlines of the OBIA as input are suboptimal. With the target shape method ANY_ANGLES, a diagonal penalty of 0 and a tolerance of 0.25 the results show the best approximated form (cf. Figure 51).

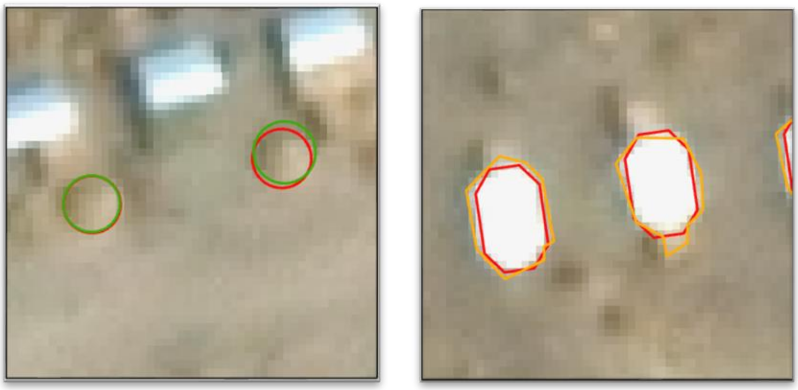


Figure 51: Best parameters for Tukuls and Tunnel Shaped Dwellings (red = reference, green and orange = result of the RBF)

Further graduation of the tolerances between 0.25 and 2 or over 2 were not tested. This study focused on a parameter indication based on the available imagery of Minawao camp. However, with different tents and different image resolution the results can vary. In the case of refugee camps the setting of precision was not necessary, because the camps do not correspond to boundary or infrastructure like streets.

Generally, this work provides new insights into the relationship between the size of the extracted dwellings and the parameter values to be applied while using the 'Regularize Building Footprint' function in the context of refugee camp analysis. The most important result of the study is that the RBF function is a good generalization method for rectangular shapes and circles under the permissions which mentioned at the beginning of the chapter. Considering the fact that the RBF function offers a huge potential, the resolution, the difficult spectral properties and small objects limit the function results, especially for octagonal shapes. In addition, the result depends strongly on the quality of the extracted results. If this is accurate, the method can also be better applied.

6. Conclusion and outlook

In the master thesis “Evaluation of the ‘Regularize Building Footprint’ function for the generalization of refugee dwellings based on very high-resolution satellite imagery”, the potential of the function was investigated. In this context it was searched for the advantages of the ‘Regularize Building Footprint’ function for the generalization of refugee dwelling outlines based on very high-resolution satellite imagery. Furthermore, it was explored what correlation between the best fitting parameters of the ‘Regularize Building Footprint’ function and the area of the extracted outlines occurs.

To answer the research question, it was evaluated if the ‘Regularize Building Footprint’ function is suitable to generalize the extracted outlines of refugee shelters. The study area was the refugee camp in Minawao. The buildings were extracted from a very high-resolution satellite image using a semi-automatic image analysis. The aim was to generalize the outlines of dwellings from the Minawao camp so that they are similar to the real dwellings and can be used as a basis for maps and 3D objects.

The challenge of this work was to find a way to validate the results of the different parameters. Because of a lack of in situ data the results could not be compared with the real data. In the same way the results cannot be systematically compared with the original extracted data in terms of area and perimeter. This is because of the non-accurate extraction. The outlines have a lot of artefacts and errors which results in larger or smaller object which would not lead to an authentic conclusion. Since there are limits in the resolution and the extraction analysis, a method must be found to check the new outlines. For this reason, manually composed reference data were collected as support. For larger objects, which were easy to distinguish on satellite images due to their good appearance, the data obtained for this purpose provides a good reference. On the other hand, for smaller objects, the original shape and orientation were vague. At the end standard objects are defined which were based on their mean values for area, perimeter, shape, and orientation from about 20 sample buildings of each dwelling class. This enables a systematic comparison of the function results with the reference objects. During the process of the research, several assumptions were made to use the method optimal.

After the pre-considerations it was explored whether different parameter settings in terms of shape, orientation, area, and perimeter provide good results compared to the reference data.

This was done through a quantitative and a qualitative evaluation. For this purpose, a scoring model analysis was carried out. Criteria for verification are determined by the perimeter, shape, orientation, and area.

In summary, the study indicates that the RBF function is a good method for generalizing the contours of dwellings. The disagreements from the reference data in terms of shape and orientation are minimal for the large rectangularly objects. For the rectangular buildings over 15 m² a tolerance of 2 and for smaller buildings with a tolerance of 0.5 show good result. For smaller objects it is difficult to make a concrete statement, because the original size of a form is difficult to determine in the image. But even so the new data provides good results in comparison to the manually collected references. For more complex shapes such as the octagons, however, this method is only limited suitable. It can be said that the method can deliver good results for the tunnel shaped dwellings with a diagonal penalty of 0 and a tolerance of 0.25, but the desired shape is only rudimentarily achieved.

With regard to the research questions, the function shows a huge benefit for the generalization of rectangular buildings. However, it should be noted, that the evaluation method was subject by many challenges, such as the lack of reference data, a variety of different input data or the subjective weighting, which can lead with other settings in different results. For more complex dwellings with different angles this function lead in not optimal fittings. On the other hand, it can be found a relationship between the best fitting parameters and the area of the extracted lines. To test the transferability of the findings, the parameters can be investigated with other refugee camps in further studies.

One problem that appears after applying the best fitting parameters to the entire camp, is that some results are overlapping. Further considerations, such as the alignment of the individual shelters or moving single objects so that they do not overlap, are further research questions. Taking over and preparing the original data by simplification with other methods like the PEAK function or to combine it with the MBB would be an option that should be investigated to avoid overlapping results.

Another question with regard to the rapid development in satellite image analysis is the use of the approximation technique during Deep Learning in Convolutional Neural Networks processes. Such neural networks learn from a large amount of data and can perform a variety of tasks. With higher resolution images and improved dwelling extraction techniques, the outlines may be better. In combination with some in situ data or better reference data the extracted outlines may be improved for the RBF. Whether the result of this study can also be transferred to Deep Learning and Convolutional Neural Networks could be investigated in further research.

7. References

BANK WORLD. 2023. *Climate Change Knowledge Portal* [Online]. Available: https://climateknowledgeportal.worldbank.org/country/cameroon/climate-data-historical [Accessed 01.08.2023].

BENZ, S., PARK, H., LI, J., CRAWL, D., BLOCK, J., NGUYEN, M. & ALTINTAS, I. Understanding a rapidly expanding refugee camp using convolutional neural networks and satellite imagery. Proceedings - IEEE 15th International Conference on eScience, eScience 2019, 2019. 243-251.

BLASCHKE, T. 2010. Object based image analysis for remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing,* 65**,** 2-16.

BODANSKY, E. & GRIBOV, A. Approximation of Polylines with Circular Arcs. *In:* LLADÓS, J. & KWON, Y.-B., eds. Graphics Recognition. Recent Advances and Perspectives, 2004// 2004 Berlin, Heidelberg. Springer Berlin Heidelberg, 193-198.

BODANSKY, E., GRIBOV, A. & PILOUK, M. Smoothing and compression of lines obtained by raster-to-vector conversion. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2002. 256-265.

DE LIMA, R. P. & MARFURT, K. 2020. Convolutional neural network for remote-sensing scene classification: Transfer learning analysis. *Remote Sensing,* 12.

DOUGLAS, D. H. & PEUCKER, T. K. 1973. Algorithms for the Reduction of the Number of Points Required to Represent a Digitized Line or its Caricature. *Cartographica: The International Journal for Geographic Information and Geovisualization 10.2.*

ESA. 2023. Available: https://earth.esa.int/eogateway/missions/worldview-2 [Accessed 25.06.2023].

ESRI. 2023a. *How Smooth Line and Smooth Polygon work* [Online]. Available: https://pro.arcgis.com/en/pro-app/latest/tool-reference/cartography/how-smooth-line-and-smooth-polygon-work.htm [Accessed 28.07.2023].

ESRI. 2023b. *Regularize Building Footprint (3D Analyst)* [Online]. Available: https://pro.arcgis.com/en/pro-app/latest/tool-reference/3d-analyst/regularize-building-footprint.htm [Accessed].

ESRI. 2023c. *Simplify Polygon (Cartography)* [Online]. Available: https://pro.arcgis.com/en/pro-app/latest/tool-reference/cartography/simplify-polygon.htm [Accessed 04.07.2023].

FÜREDER, P., TIEDE, D., LÜTHJE, F. & LANG, S. 2014. Object-based dwelling extraction in refugee/IDP camp - challenges in an operational mode. *South-Eastern European Journal of Earth Observation and Geomatics,* 3**,** 539-543.

GAO, Y., LANG, S., TIEDE, D., GELLA, G. W. & WENDT, L. 2022. Comparing OBIA-Generated Labels and Manually Annotated Labels for Semantic Segmentation in Extracting Refugee-Dwelling Footprints. *Applied Sciences (Switzerland),* 12.

GARZELLI, A., NENCINI, F., ALPARONE, L., AIAZZI, B. & BARONTI, S. 2004. *Pan-sharpening of multispectral images: a critical review and comparison*.

GELLA, G. W., TIEDE, D., LANG, S., WENDIT, L. & GAO, Y. 2023. Spatially transferable dwelling extraction from Multi-Sensor imagery in IDP/Refugee Settlements: A meta-Learning approach. *International Journal of Applied Earth Observation and Geoinformation,* 117**,** 103210.

GEOHUM. 2023a. *GEOHUM* [Online]. Available: https://geohum.eu/ [Accessed 06.07.2023].

GEOHUM. 2023b. *GEOHUM About* [Online]. Available: https://geohum.eu/about/ [Accessed 07.07.2023].

GEOSAGE. 2023. *imagefusion* [Online]. Available: https://www.geosage.com/highview/imagefusion.html [Accessed 11.08.2023].

GRIBOV, A. Searching for a Compressed Polyline with a Minimum Number of Vertices (Discrete Solution). Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2018. 54-68.

GRIBOV, A. Optimal Compression of a Polyline while Aligning to Preferred Directions. 2019 International Conference on Document Analysis and Recognition Workshops, ICDARW 2019, 2019. 98-102.

GRIBOV, A. & BODANSKY, E. 2004. A new method of polyline approximation. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics),* 3138**,** 504-511.

GRIBOV, A. & BODANSKY, E. Reconstruction of orthogonal polygonal lines. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2006. 462-473.

KEMPER, T. & HEINZEL, J. 2014. Mapping and Monitoring of Refugees and Internally Displaced People Using EO Data.

KINDZEKA, M. E. 2022. Cameroon Says Cholera Hits Minawao, a Nigerian Refugee Camp.

KÜHNAPFEL, J. B. 2021. Scoring und Nutzwertanalysen. *Springer Books*.

LAND LIFE. 2022. *Minawao Refugee Camp – Cameroon* [Online]. Land Life Company. Available: https://landlifecompany.com/projects/minawao-refugee-camp-cameroon/ [Accessed 25.06.2023].

LANG, S. 2008. Object-based image analysis for remote sensing applications: Modeling reality – dealing with complexity. *Lecture Notes in Geoinformation and Cartography***,** 3-27.

LANG, S., FÜREDER, P., KRANZ, O., CARD, B., ROBERTS, S. & PAPP, A. 2015. Humanitarian Emergencies: Causes, Traits, and Impacts as Observed by Remote Sensing.

LANG, S., FÜREDER, P., RIEDLER, B., WENDT, L., BRAUN, A., TIEDE, D., SCHOEPFER, E., ZEIL, P., SPRÖHNLE, K., KULESSA, K., ROGENHOFER, E., BÄUERL, M., ÖZE, A., SCHWENDEMANN, G. & HOCHSCHILD, V. 2020. Earth observation tools and services to increase the effectiveness of humanitarian assistance. *European Journal of Remote Sensing,* 53**,** 67-85.

SESTER, M. 2001. Generalization Based On Least Squares Adjustment.

SHAMOS, M. I. 1978. *Computational geometry.*

TIEDE, D., FÜREDER, P., LANG, S., HÖLBLING, D. & ZEIL, P. 2013. Automated analysis of satellite imagery to provide information products for humanitarian relief operations in refugee camps -from scientific development towards operational services. *Photogrammetrie, Fernerkundung, Geoinformation,* 2013**,** 185-195.

TIEDE, D., KRAFFT, P., FÜREDER, P. & LANG, S. 2017. Stratified template matching to support refugee camp analysis in OBIA workflows. *Remote Sensing,* 9.

TIEDE, D., LANG, S., HÖLBLING, D. & FÜREDER, P. Transferability of OBIA rulesets for IDP camp analysis in DARFUR. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives, 2010.

TOUSSAINT, G. T. Solving Geometric Problems with the Rotating Calipers 1983.

UNHCR 2016. Shelter Design Catalgue. *In:* SECTION, U. S. A. S. (ed.).

UNHCR. 2022. *UNHCR besorgt über Cholera-Ausbruch in Minawao und fordert dringende Unterstützung* [Online]. Available: https://www.unhcr.org/africa/news/news-releases/unhcr-concerned-over-cholera-outbreak-minawao-calls-urgent-support [Accessed 05.06.2023].

UNHCR. 2023. *Refugee Data Finder* [Online]. Available: https://www.unhcr.org/refugee-statistics/ [Accessed 06.07.2023].

Z_GIS & MSF. 2015. *Refugee camp Minawao Cameroon Dwelling extraction 10 March 2015*.

ZHOU, W., NEWSAM, S., LI, C. & SHAO, Z. 2018. PatternNet: A benchmark dataset for performance evaluation of remote sensing image retrieval. *ISPRS Journal of Photogrammetry and Remote Sensing,* 145**,** 197-209.

9. Appendix

Appendix 1: Map of the generalized dwelling outlines