Image-based Target Detection and Tracking Using Image-assisted Robotic Total Stations

Volker SCHWIEGER, Otto LERKE and Gabriel KEREKES (Germany)

Key words: Robotic total station, kinematic positioning, reflectorless target detection, reflectorless target tracking, feature detection, photogrammetric technique

SUMMARY

Robotic total stations are modern geodetic multi-sensor systems measuring horizontal and vertical angles as well as distances using time-of-flight methods, thus delivering 3D-coordinates for static as well as moving objects. Automatic target detection (by rough and fine pointing techniques) and tracking are standard techniques if the objects are signalized with reflectors and the total station is motorized. Nowadays these instruments are additionally equipped with one or two cameras to generate images mainly for documentation purposes. This paves the way to detect and track objects that are not signalized by reflectors. Photogrammetric techniques such as SURF (Speeded-up Robust Feature) or SIFT (Scale Invariant Feature Transform) are applied for the detection of special, recognizable object features in the images. The pixel coordinates of these features result in vertical and horizontal angles if the parallaxes between the camera optical center and the total station origin are known or calibrated. If the features are extracted in a sequence of images the movement of any object can be tracked automatically. For the position determination reflectorless distance measurement from the total station to the object is required additionally. Until now this was realized only for static objects. In this contribution an example of a kinematic application is also shown. The quality of these tracking procedures may be verified by an instrument of higher accuracy. This has been done using a laser tracker.

Image-Based Target Detection and Tracking Using Image-Assisted Robotic Total Stations (9887)
Volker Schwieger, Otto Lerke and Gabriel Kerekes (Germany)

FIG Working Week 2019
Geospatial information for a smarter life and environmental resilience
Hanoi, Vietnam, April 22–26, 2019

Image-based Target Detection and Tracking Using Image-assisted Robotic Total Stations

Volker SCHWIEGER, Otto LERKE and Gabriel KEREKES (Germany)

1. INTRODUCTION

In the past 15 years, Image-assisted Robotic Total Stations (IATS) have enriched the application spectrum in the field of engineering geodesy through their combined 3D positioning capability and digital camera integration.

The theoretical background and the prerequisites of IATS for precise measurements can be consulted in Ehrhart and Lienhart (2017). For a historical insight of the development that leads to the current state-of-the-art, Wagner et al. (2014) is advisable. An overview of IATS producers to that time is given by Scherer and Lerma (2009).

In this paper, two commercially available IATS are used to exemplify applications in which objects are identified and tracked by means of images using image processing algorithms like SIFT and SURF. Object recognition as well as target tracking are realized by the digital cameras of the IATS.

2. IMAGE-BASED OBJECT RECOGNITION, POSITION DETERMINATION AND TRACKING

This section treats the usage of an image-assisted total station (IATS) in order to perform static object recognition, position determination, and tracking. Within this context, the potential of the integrated photo module of the IATS should be fully exploited and not be limited to documentation purposes only.

2.1 Image Processing Fundamentals

According to Luhmann (2010) the digital image processing can be subdivided into different steps: image capturing, pre-processing, segmentation, detection, clustering, configuration, and result assessment.

The pixel coordinate system has a fundamental function in the image processing task. It is defined as a left-handed 2-dimensional x-y coordinate system, described by rows and columns, whereby the x-axis pointing into the direction of rows and the y-axis into direction of columns (Sonka et al. 1994). In general, an object in the image covers multiple pixels in the image. These covered pixels are coherent and their grey values are rather similar.

To increase the efficiency of image processing, image pyramids are used. An image pyramid is a series of images, where the next following image is reduced in resolution and size by factor n, compared with its predecessor. Additionally, the reduced image is smoothed by filter. Thus with decreasing resolution, small image structures disappear because the informational content decreases (Luhmann 2010). This allows to firstly searching for rough features in images with decreased resolution. Afterwards the search can be focused on

Image-Based Target Detection and Tracking Using Image-Assisted Robotic Total Stations (9887)
Volker Schwieger, Otto Lerke and Gabriel Kerekes (Germany)

FIG Working Week 2019
Geospatial information for a smarter life and environmental resilience
Hanoi, Vietnam, April 22–26, 2019

previously found, interesting areas of the image, by the use of images of the pyramid with higher resolution.

Filtering can be realized by convolutions in the spatial domain and multiplications in the frequency domain. For frequency domain operations the image must be firstly transformed into the frequency domain e.g. by Fourier transformation (Sonka et al. 1994, Luhmann 2010). The convolution is realized by a stepwise "sliding" of the convolution kernel across the image. The pixel value beneath the kernel is multiplied by the corresponding filter kernel value. These products are subsequently summed and multiplied by the sum of the kernel values. The resulting value is then allocated to the pixel, which best corresponds to the mean value of the filter kernel (Luhmann 2010).

There are different smoothing filter used in digital image processing. The most import smoothing filters are the Box filter and the Gaussian filter. Detailed information on smoothing filters may be found in Luhmnann (2010). Besides smoothing filters, edge detector filters play a superior role in image processing. They are based on numerical derivations of grey value functions and are used to locate sharp changes in grey values, which in turn, indicate edges. Edges are pixels, where the first derivative of the grey value function changes abruptly (Sonka et al. 1994).

The simplest edge detector is the Roberts detector, based on the 1st derivative of the pixel plane in x- and y-direction (Girod 2013). The Sobel operator combines a derivation with a smoothing. This helps to counteract the amplification of noise induced by the derivation. To obtain further information about the edge curvature the Laplace operator can be applied. It is based on the 2nd derivative of the grey value function. Thereby the edges are represented by sign changes. On the other hand the 2nd order derivative negatively affects the noise sensitivity. To counteract this adverse effect the image can be smoothed by the Gaussian filter before derivation. The combination of smoothing and differentiation leads to the Laplacian of Gaussian operator. The main disadvantage of edge detectors is the instability of their position. Edges are stable in only one direction. In contrast, most image processing algorithms need features with stable positions. Therefore corners are more suitable because of their fix localization in both directions. One of the most common corner detectors is the Hessian-detector. The detector is based on the usage of the determinant of the Hessian matrix (Bay et al. 2006). The Hessian matrix is defined as follows (Merziger and Wirth 2010):

$$\mathcal{H}_f = \frac{\partial^2 f}{\partial x_i \partial x_j}(x) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1}(x) & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n}(x) \\ \vdots & & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1}(x) & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_n}(x) \end{pmatrix} \qquad (1)$$

The feature detection is accomplished by the analysis of the determinant. The feature is located at the position where the determinant indicates a maximum. According to Merziger and Wirth (2010) the determinant is defined as follows:

$$det\,\mathcal{H} = D_{xx} \cdot D_{yy} - (w \cdot D_{xy})^2 \qquad (2)$$

The elements D_{xx}, D_{yy} and D_{xy} are so called Blob-filter responses at image point x. They represent a 9 x 9 Box-filter. w represents the weighting factor. In the next section the usage of

Image-Based Target Detection and Tracking Using Image-Assisted Robotic Total Stations (9887)
Volker Schwieger, Otto Lerke and Gabriel Kerekes (Germany)

FIG Working Week 2019
Geospatial information for a smarter life and environmental resilience
Hanoi, Vietnam, April 22–26, 2019

the presented image processing tools will be illustrated within common image processing algorithms.

2.2 Image Processing Algorithms for Feature Extraction Fundamentals

There are many possibilities of image processing that originate from the field of computer vision (Shapiro and Stockman, 2000) and are used for such purposes, but these can be grouped in three classes: edge-based, template-based and point-based (Reiterer and Wagner, 2012). As the name already suggests, edge-based implies identifying edges of an object and then computing, if necessary, its geometrical center. In this way, if the geometry doesn't change and the background shows contrast, objects can be easily detected based on their edges. Template-based involves using a pre-known pattern that the algorithm recognizes. Therefore patterns are always compared with the reference pattern and if a match is found, the object is considered to be the searched one. Point-based implies finding certain features (points) that, similarly to the template-based, match a pre-known image. Two well-known point-based feature extraction algorithms will be detailed in the following sections. The process of feature extraction generally consists of two steps: detection and description.

2.2.1   SIFT (Scale Invariant Feature Transform) Algorithm

According to Lowe (1999) the requirements on image-based object recognition from real world is, that image objects, so called features, that are randomly arranged in space and partly covered, are identified and detected uniquely. The features shall be invariant with respect to translations, rotations, scaling and changes in illuminance. Furthermore, the features shall be unaffected by image distortions and noise. To fulfill the requirements the image features must possess characteristic shapes, in order to be identified uniquely. The SIFT algorithm decomposes the image into a finite number of objects, which are described by local descriptor vectors. The required processing step after the image capturing is the localization of the characteristic image features. Their positions, identified in the space domain, must fulfill the invariance properties. The requisite mathematical steps are comprised in the so called detector. The detector is based on two convolutions of the image with the Gaussian kernel and a subsequent forming of Gaussian difference in order to detect curvatures.
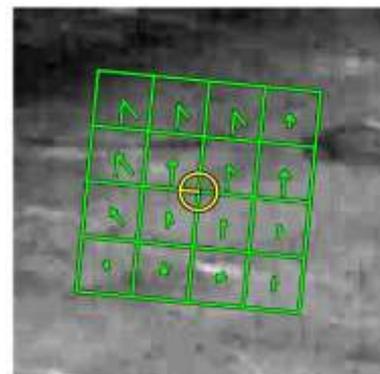

Figure 1. Feature descriptor; yellow: orientation R_(i,j); green: magnitude M_(i,j).

The Gaussian difference approximates the Laplacian of Gaussian in order to shorten processing time. After resampling the image by the use of bilinear interpolation (Luhmann 2010), the search for local maxima and minima within different levels of image pyramids is conducted, where the neighboring pixels of the image are compared with each other.

The detector step is followed by the descriptor step in order to characterize the image. For this purpose, the calculation of the feature gradient and orientation is conducted. This is done for each pixel $A_{i,j}$ by the calculation of the gradient magnitude $M_{i,j}$ and the orientation $R_{i,j}$. Figure 1 exemplarily depicts the descriptor.

Image-Based Target Detection and Tracking Using Image-Assisted Robotic Total Stations (9887)
Volker Schwieger, Otto Lerke and Gabriel Kerekes (Germany)

FIG Working Week 2019
Geospatial information for a smarter life and environmental resilience
Hanoi, Vietnam, April 22–26, 2019

$$M_{i,j} = \sqrt{\left(A_{i,j} - A_{i+1,j}\right)^2 + \left(A_{i,j} - A_{i,j+1}\right)^2} \qquad (3)$$

$$R_{i,j} = \arctan\left(\frac{A_{i,j} - A_{i+1,j}}{A_{i,j+1} - A_{i,j}}\right) \qquad (4)$$

The descriptor vector is created for each feature and is of dimension $n = 128$. It contains the feature's stable position in the image, the scale and the orientation. Optionally, the descriptor additionally might contain colors or textures. Detailed inside view into the SIFT algorithm is provided by Lowe (1999).

2.2.2 SURF (Speeded-Up Rubust Feature) Algorithm

According to Bay et al. (2008) the most important property of the detector is the repeatability. In this context, the repeatability means the reliability of the detector to identify and find the same physical object under changing visual conditions. Therefore the neighborhood of each relevant image point is described by the descriptor. The descriptor must be distinctive and robust towards noise, translations as well as geometric deformations and photogrammetric distortions. By the use of feature descriptors from two different images, these features and even the images (usage of multiple features) can be allocated to each other. The allocation is based e.g. on the Euclidean distance between the two descriptors. The dimension of the descriptor has a direct impact on the allocation time and thus on the computation duration. Hence, small dimensions of the descriptor vector are desirable on the one hand. On the other hand, small descriptor dimensions are less unique and thus less distinctive (Bay et al. 2008). SURF algorithm offers a good compromise between short processing times and sufficient descriptor dimension in order to ensure distinctiveness. SURF uses the scale and rotation invariants detectors and descriptors. No color information is used.

The detector is based on the usage of the determinant of the Hessian matrix (cf. Section 2.1). Therefore integral images are introduced. Integral images serve for fast computations of pixel sums within rectangular sections (Bay et al. 2008). The input of the integral image $I_\Sigma(\boldsymbol{x}) = \sum_{i=0}^{i \le x} \sum_{j=0}^{j \le y} I(i,j)$ at position $\boldsymbol{x}=(x,y)^T$ is the sum of all pixels of the input image I within a rectangular region, spanned between the image origin and image point $\boldsymbol{x}$.

Now the Blob-filter responses, which can be used to calculate maxima and minima, are stored in a Blob response map. The map represents the image scale space. The scale space is implemented by image pyramids and is divided into octaves. The octaves represent series of filter responses, determined by convolution. Each octave is subdivided into a constant number of scaling levels. Hence the detector contains the steps of suppression of the input image and the determination of the features by the described procedure, using the Hessian determinant. The use of integral images increases the computational speed and enhances the robustness (Bay et al. 2008).

The descriptor uses the intensity values to characterize the features. It is based on the distribution of the 1st order responses of the Haar-Wavelet in x and y direction. Details on Haar-Wavelets can be excluded from e.g. Talukder and Harada (2007). For the descriptor, based on the Haar-Wavelet, firstly a rectangular region around the point of interest is built.

Image-Based Target Detection and Tracking Using Image-Assisted Robotic Total Stations (9887)
Volker Schwieger, Otto Lerke and Gabriel Kerekes (Germany)

FIG Working Week 2019
Geospatial information for a smarter life and environmental resilience
Hanoi, Vietnam, April 22–26, 2019

The determination of orientation of each region is accomplished by the detector step. Then, the regions are subdivided into smaller rectangular regions. The Haar-Wavelet responses of these smaller regions are calculated, where dx is the response in x-direction and dy is the response in y-direction. Information about the polarity and change of intensity is obtained from $|dx|$ and $|dy|$. For the description of the intensity structure of each subpixel region a 4-dimensional descriptor vector, shaped as $v = (\sum dx, \sum dy, \sum|dx|, \sum|dy|)$ is then established. The local sub-region descriptor vector is then calculated for all sub-regions, which are 4 x 4 = 16 in total. Hence the descriptor vector for each feature is of dimension $n = 16 \times 4 = 64$. More detailed information on SURF algorithm may be extracted from Bay et al. (2008). A short summary of the SIFT and SURF algorithms is presented in table 1.

Table 1 *Comparison between SIFT and SURF*

	SIFT	**SURF**
Algorithm input	Greyscale images	Intensity images
Used filter	Original filter	Approximated filter
Structure of the scale space pyramid	Different resolutions of the image	Different resolutions of the filter
Base of the descriptor	Gradients	Haar-wavelet filter response
Descriptor dimension	128-dimensional	64-dimensional
General property	More reliable	Faster

2.3 Object Recognition and Matching

After the extraction of the features is accomplished by one of the pre-described algorithms of section 4.2, the next operation, in sequence of the image processing, is the object recognition. The procedure is based on a comparison between the reference image and the test image and is defined as matching step. In detail, the extracted key points (cf. Section 4.2) of the reference and test image are compared with each other. For the comparison, n-dimensional feature vector (SIFT $n = 128$, SURF $n = 64$) is defined as the position of the key points in n-dimensional space. The task to be solved is the finding of next neighbors between the two feature vectors (reference and test image) by the use of the Best-Bin-First algorithm, according to Muja and Lowe (2012). Best-Bin-First algorithm is suitable to efficiently find an approximate solution to the nearest neighbor search problem in very-high-dimensional spaces (Kybic and Vnucko 2010). The procedure is based on a binary coded description of the feature vector. It has to be stated that this solution provides an approximate solution only.

The key points, expressed by the feature vector created by SIFT or SURF, that are passed to the algorithm are subdivided in k clusters. Therefore k randomly chosen points form the cluster center. The remaining points are allocated to the particular cluster, to which their distance is minimal. If the cluster is larger than a predefined threshold, new cluster centers are chosen and the algorithm starts anew. This helps to prevent different cluster sizes, if unfavorable cluster centers have been chosen beforehand. Each cluster forms a k-dimensional tree, the so called k-d tree. The algorithm is advantageous in tree building and during search operations, because of the parallel and simultaneous processing of different trees. The search for the nearest neighbor in the tree is conducted from top to bottom, where at each branch the nearest node to the starting point is marked. Non-marked nodes are stored in a separate

Image-Based Target Detection and Tracking Using Image-Assisted Robotic Total Stations (9887)
Volker Schwieger, Otto Lerke and Gabriel Kerekes (Germany)

FIG Working Week 2019
Geospatial information for a smarter life and environmental resilience
Hanoi, Vietnam, April 22–26, 2019

priority list. After all trees are searched once, the search proceeds from the next point which is nearest to starting point. Now the next nearest neighbors of particular k-d trees from the priority list are compared. The points with the minimal distance are chosen. The number of chosen points defines the approximation grade. The higher the grade the more neighbors are found, though the processing time increases.

By the use of the recognized point pairs, the transformation matrix between the two images can be determined. The matrix allows transformations of pixels respectively points from the reference image into test image. This procedure is called pixel-to pixel transformation.

To avoid gross errors in the point cloud and to increase the robustness a filtering by the MSAC (M-estimator Sample Consensus) algorithm is applied. Detailed information on MSAC may be found in Torr and Zisserman (2000). The transformation itself can be expressed by an affine transformation according to Lowe (1999):

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} m_1 & m_2 \\ m_3 & m_4 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \end{bmatrix} \tag{5}$$

u, v:	test image point coordinates,
x, y:	reference image point coordinates,
t_x, t_y:	translation parameters,
m_1, m_2, m_3, m_4:	rotation and scale parameters.

Thus, two tasks have to be solved. The first task is the determination of the transformation parameters. These are determined by the least squares method according to e.g. Niemeier (2008). In order to estimate the 6 parameters, the minimum number of requisite matches between the reference and the test image must be three. In the second task all points of the reference image can be transformed into the test image by the estimated transformation parameters.

2.4 Object Position Determination

After the successful matching and identification of the object in the test image according to section 4.3, the next challenge is the position determination of the specific object.

In general, the determination of object positions in the target coordinate system, which is the tachymeter system in this specific case, requires horizontal and vertical telescope angles, as well as a distance measurement, obtained by the reflector-less distance measurement (EDM). Hence, the obtainment of the horizontal and vertical telescope angles from images is necessary and will be elaborated in the following. In each reference image key points are defined, of which the pixel coordinates in the reference image, as well as their coordinates in the object coordinate system are known. The geometry of the object is also fully known in the object coordinate system. By the use of the pre-described pixel-to-pixel transformation from section 4.3, the transformation parameters between the reference and the test image can be obtained. The next step is the transition from the image system into the tachymeter system. Therefore determination of the telescope's aiming direction, expressed by Hz- and V-angles, from the present pixel coordinate of the appropriate point must be determined. For this, the relation between pixel and angle is required. This relation, described by the transfer factor i, is

Image-Based Target Detection and Tracking Using Image-Assisted Robotic Total Stations (9887)
Volker Schwieger, Otto Lerke and Gabriel Kerekes (Germany)

FIG Working Week 2019
Geospatial information for a smarter life and environmental resilience
Hanoi, Vietnam, April 22–26, 2019

different for each instrument and is either given or must be determined by calibration. The relation describes the function between a specific telescope angle α and the induced shift p in the pixel system. Thus the transfer factor i can be expressed as follows:

$$i = \frac{p}{\alpha} \tag{6}$$

Reconsidering the correction terms, introduced for the eccentric camera-telescope layout, the calculation of the horizontal and vertical telescope angles from the present image point is carried out by equation (7-8).

$$Hz = i \cdot \left(hpix_g - (hpix_m + k_h) \right) \tag{7}$$

$$V = i \cdot \left(vpix_g - (vpix_m + k_v) \right) \tag{8}$$

$hpix_g, vpix_g$:	pixel coordinates (row and column) of the measured object in the image,
$hpix_m, vpix_m$:	pixel coordinates (row and column) of the image center,
Hz:	horizontal telescope angle,
V:	vertical telescope angle,
k_h:	horizontal correction term (valid for eccentric layout only),
k_v:	vertical correction term (valid for eccentric layout only).

By equation (7-8) the telescope directions of the key point can now be calculated. These directions can be adjusted by the tachymeter's actuators. By the additional use of the reflector-less distance measurement s, all required elements for the coordinate calculation are available. According to Torge (1980) the coordinates are calculated as follows:

$$x = s \cdot \cos Hz \cdot \sin V \tag{9}$$
$$y = s \cdot \sin Hz \cdot \sin V \tag{10}$$
$$z = s \cdot \cos V \tag{11}$$

2.5 Principles of Image-based Object Tracking

Basically, when using images a certain object of interest must be identified and some features need to be extracted for further processing. In sections 2.2.1 and 2.2.2, SIFT and SURF, which are point-based algorithms, were described in detail and further emphasis will not be put here. The main difference between different algorithms is the computation time, which plays an important role for the image-based tracking process.

Similar to a reflector tracking process, the difference between the crosshair point and object center needs to be constantly minimized. Therefore, if change or movement of the object is detected, the telescope is guided until the before mentioned difference is reduced to zero. Applying this in a continuous sequence creates the image-based tracking process. The quality of this process is mainly dictated by image resolution, optical zoom capacity, data transfer rate, processing speed, object speed and telescope rotations speed.

Image-Based Target Detection and Tracking Using Image-Assisted Robotic Total Stations (9887)
Volker Schwieger, Otto Lerke and Gabriel Kerekes (Germany)

FIG Working Week 2019
Geospatial information for a smarter life and environmental resilience
Hanoi, Vietnam, April 22–26, 2019

3. APPLICATIONS
3.1 Example of Static Object Recognition and Positioning

In this example the position of an unmanned ground vehicle (UGV) should be determined by the presented image-based method. The concerned object is model of a tracked loader at scale 1:14. The used instrument is the Trimble S7 robotic total station, that incorporates a digital camera which has a field of view of 20.3° x 15.2° (Trimble 2018).

The following presented steps are embedded within a control program, in order to automatically steer the total station. Most state-of-the art total stations can be steered externally by receiving commands from laptops or PCs via defined interfaces. This allows the user to create application and problem oriented programs. The steering possibilities concern almost every component group of the tachymeter. The interfaces also allow outsourcing image processing algorithms and others to external devices, in order to not overstrain the internal processor of the total station. In the current configuration, the steering program for the Trimble S7 total station is implemented in the programming language C#. The image processing algorithms are implemented in Matlab©. The superordinate control program, which coordinates and synchronizes the data flow between the individual programming components is realized in the graphical programming language LabView from National Instruments.

The implemented steering program lets the total station automatically move the telescope by pre-defined angles in vertical and horizontal direction, capture images and deploy reflector-less distance measurements.

The flowchart of the specific total station steering program for object recognition and positioning is depicted in figure 2.
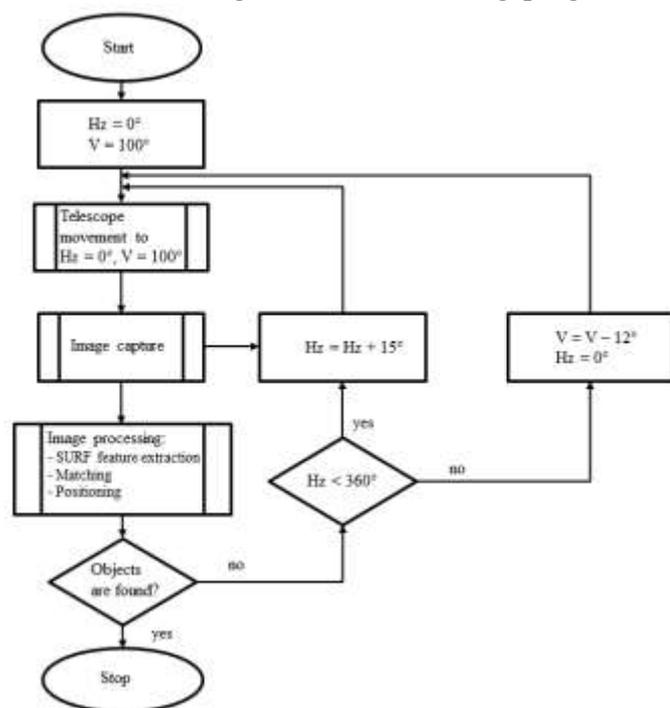

Figure2 Flowchart of static, image-based object recognition

After the image has been captured by the camera module, the feature extraction by the use of the SURF algorithm, according to section 2.2, is conducted. The extraction result in the reference image is depicted in figure 3.

Accordingly, figure 4 (right) shows the extraction result in the test image. The image was captured from another perspective than the reference image. This should underline the performance of the SURF algorithm, where test images might be taken from a different perspective, but the matching robustness is still given.

Image-Based Target Detection and Tracking Using Image-Assisted Robotic Total Stations (9887)
Volker Schwieger, Otto Lerke and Gabriel Kerekes (Germany)

FIG Working Week 2019
Geospatial information for a smarter life and environmental resilience
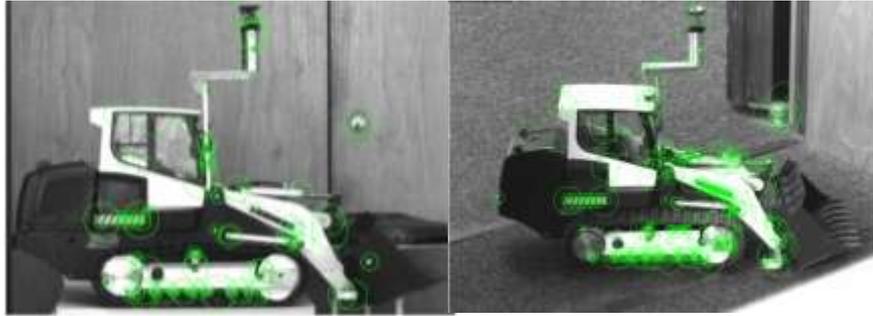Hanoi, Vietnam, April 22–26, 2019

Figure 3. Detected and extracted features in the reference image marked by green circles (left); Detected and extracted features in the test image marked by green circles (right)

Experiments showed that features, which are neither part of the object, nor part of the reference image, have been detected and extracted. Subsequently the matching step, according to section 2.3 is performed. At this point many features are matched faulty.

Therefore the MSAC algorithm is applied to the first matching result. After the MSAC filtering, fewer matches are left. These matches are unique and correct (figure 4).
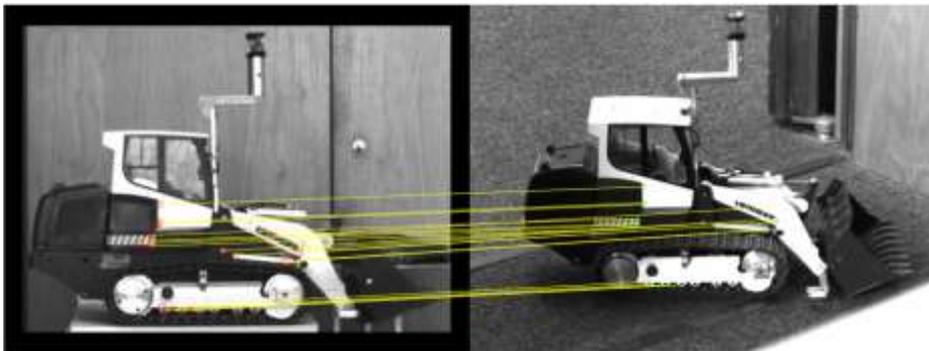

Figure 4. Matching result after MSAC filtering (left: reference image; right: test image)

Now the object recognition can follow up by the use of the pixel-to-pixel transformation, described in section 2.3. After this step, the object is uniquely identified in the test image. In the last step, telescope Hz and V angles are calculated from the pixel coordinates of the test image, accordingly to section 2.4. After the automatic aiming (setting of Hz and V angles by the servomotors), the reflector-less distance measurement is triggered and the position is obtained by equation (9-10).

3.2 Example of Kinematic Image-based Object Tracking

Tracking is generally understood in engineering geodesy as the process of following of moving objects with a certain sampling rate. The complete process is comprehensively explained in section 3 and further on, emphasis on the same process will be extended with the use of images. In contrast to using a reflector to signalize the object, using images offers versatility and flexibility of choosing which object should be tracked just by acquiring an image of it. This means that the object does not necessarily need to be accessible. First similar attempts and principle descriptions can be found in Bayer et al. (1989).

Image-Based Target Detection and Tracking Using Image-Assisted Robotic Total Stations (9887)
Volker Schwieger, Otto Lerke and Gabriel Kerekes (Germany)

FIG Working Week 2019
Geospatial information for a smarter life and environmental resilience
Hanoi, Vietnam, April 22–26, 2019

Specifically in the case of IATS used for object tracking, images are processed to constantly identify and track the desired object. Further on, this section will provide an insight into the image-based tracking principles that uses a SURF algorithm to identify the object in each frame (image) and then track it.

Recently, a system comprised of a Leica TS 16i IATS and the control software running under Matlab© was developed at the Institute of Engineering Geodesy. The TS16i is a high precision tachymeter that includes an overview camera with a 5 MPixel CMOS sensor. For object identification and tracking a SURF algorithm is used. The camera has a 15,5°x11,7° field of view and is capable of capturing up to 30 frames per second. Four optical zoom levels are available in this case and can be used for tracking at different distances.

Processing of the frames (images) takes place on an external computer that constantly receives and sends data to the IATS. The physical connection is realized through a wireless network and the developed program uses functions from the Image Processing Toolbox in Matlab. Examples of some of these functions are: image read, detect SURF features, extract features and match features.

Leica instruments can be controlled from an external source only with the use of special commands, sent as an ASCII message and defined by the GeoCOM Protocol (Leica 2018b). Depending on the hardware integrated into the IATS, only some commands are available. In the present, the CAM and MOT commands are used for controlling the camera and servomotors of the IATS.

In a first phase, the user needs to select the object that is going to be tracked. This can be done either by directly capturing an image of it and then cropping the area with the object or from a previously taken image. Once the object is selected, unique features (points) are identified on the object and will serve as basis for the tracking loop. Objects with a rich texture and varying geometry are the best suited in this case. This fact may be observed in figure 5, where the letters in the IIGS logo do not have the same amount of feature points as the building in the logo center.

Figure 5 Original image (left) and identified feature points (right)

The object center is then determined and from this point, the telescope is guided based on each processed frame. From the hardware point of view, this can happen at a rate of 30 fps or 30 Hz, but due to practical reasons of processing speed, a rate of 10 fps has been chosen. Additionally a distance measurement can be made (without reflector) and the absolute coordinates of the object are obtained. This whole tracking process is currently limited to a 0.5 Hz update rate, mainly caused by transfer and processing speed. Finally, the process can be summarized as seen in figure 6.
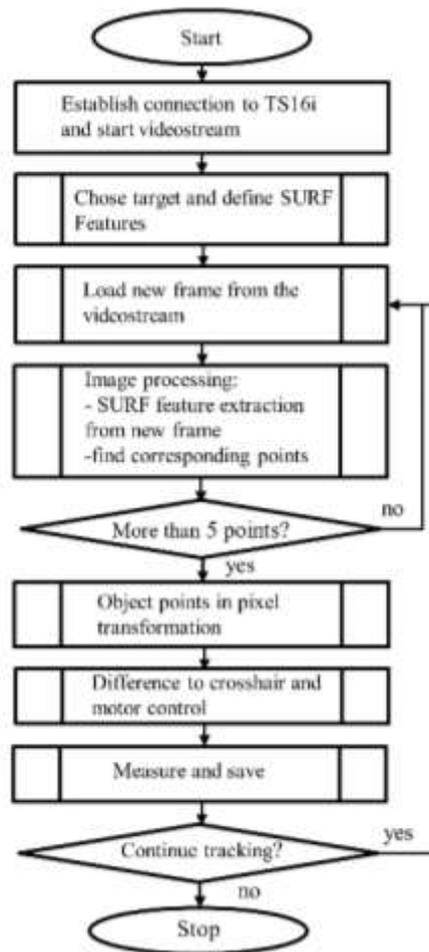
Image-Based Target Detection and Tracking Using Image-Assisted Robotic Total Stations (9887)
Volker Schwieger, Otto Lerke and Gabriel Kerekes (Germany)

FIG Working Week 2019
Geospatial information for a smarter life and environmental resilience
Hanoi, Vietnam, April 22–26, 2019

Figure 6 Tracking process flowchart

4. QUALITY CONTROL OF TOTAL STATIONS IN KINEMATIC MODE USING A LASER TRACKER

During measurements conducted by a total station it is not possible to evaluate the measurement quality, respectively the measurement accuracy, internally. Therefore external measurements by instruments with higher accuracy are required. Such an instrument is the laser tracker. Different manufacturers offer laser trackers systems, often in combination with additional accessories.

The laser tracker API Radian has been used for the following experiment examples. The distance measurement accuracy of the laser tracker is 250-times higher in kinematic mode and 500-times higher in static mode than that of the used robotic total station. The angle measurement accuracy is about 1.5-times better. Figure 7 depicts this instrument and the two IATS with their properties for a straightforward comparison.

Image-Based Target Detection and Tracking Using Image-Assisted Robotic Total Stations (9887)
Volker Schwieger, Otto Lerke and Gabriel Kerekes (Germany)

FIG Working Week 2019
Geospatial information for a smarter life and environmental resilience
Hanoi, Vietnam, April 22–26, 2019

Intrument/ Specification	Leica TS16 I	Trimble S7	API Radian™ Lasertracker
Measurement Accuracies	Angle: 0.3 *mgon* Distance static mode: 1 mm + 1,5 ppm Distance kinematic mode: 3 mm + 1,5 ppm	Angle:0.3 *mgon* Distance static mode: 1 mm + 2 ppm Distance kinematic mode: 4 mm + 2 ppm	Angle: 0.2 *mgon* Distance IFM (static only): 10 μm or 5 ppm Distance ADM: 10 ppm (Lock – On)
Image/Measuring Sampling Rate	≤ 30 Images/Second ≤ 10 Hz	≤ 5 Images/Second ≤ 20 Hz	- / IFM: continuously ADM: no details

Figure 7 Overview of the used measuring systems (Leica 2018a ; Trimble 2018; Automated Precision inc. 2018)

In the object tracking example, the regular laser tracker reflector SMR (Spherical Mounted Reflector) (Automated Precision inc. 2018) was used.

In order to test the system's performance, a target has been placed on a small trolley that is moving on a miniature railway and tracked during the movement. The reference is given by a laser tracker measurement of a reflector placed on the exact same axis of the trolley (Figure 8). For both kinematic measurements the object was manually shifted.

Figure 8. Moving trolley with laser tracker reflector (SMR) (left) and the target with feature points as identified by the IATS (right)

The results are further presented and the differences between the coordinates obtained from laser tracker and IATS can be observed in figure 9. A systematic deviation is firstly noticeable in comparison to the tracker measurement. After shifting the tracker coordinates by 5 mm in the *X* direction, which is firstly a manual correction, a plausible comparison may be conducted. Consequently the tracker coordinates were fitted to a 4th degree polynomial function and the individual distances from the IATS coordinates to this regression line were computed. A mean value of 0.6 mm for lateral deviation was achieved.
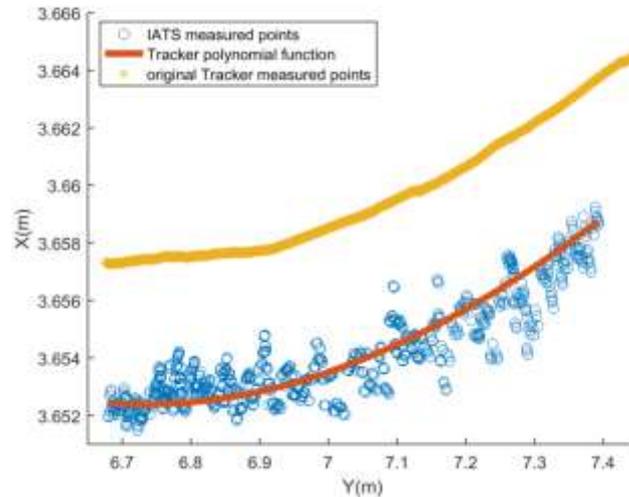
Image-Based Target Detection and Tracking Using Image-Assisted Robotic Total Stations (9887)
Volker Schwieger, Otto Lerke and Gabriel Kerekes (Germany)

FIG Working Week 2019
Geospatial information for a smarter life and environmental resilience
Hanoi, Vietnam, April 22–26, 2019

Figure 9 Differences between IATS and Laser Tracker measurement of the same reference line

Future improvements foresee the identification of this systematic effect and the usage of more efficient image processing tools in combination with a real-time industrial controller unit like the CompactRIO System from National Instruments that would help reduce latency time.

5. CONCLUSION

New features of IATS, the built-in cameras are highlighted in this contribution. These cameras open the way to image processing and object recognition. In this contribution the advances in object detection and tracking using image processing techniques like e.g. the well-known SURF and the SIFT algorithm are presented. The algorithms have been implemented on high-end total stations available on the market. The results are encouraging, especially with respect to the measurement accuracy.

The expected accuracy will be less than one mm in tracking mode if the systematic effects of the evaluation procedure are eliminated. The measurement accuracy is determined using a laser tracker delivering accuracies around 10 μm level for distance measurements. The tracking rate needs to be increased in the future to assure a continuous homogeneous real-time tracking. When the algorithms will run in real-time with a tracking rate of 10 to 20 Hz, any objects, even if they are not equipped with a reflector, may be tracked. A remaining challenge is the synchronization of the totals stations with other sensors within sensor fusion algorithms. The authors will further work on the open issues and see the results as a valuable input for the research within the DFG Research Cluster of Excellence "Integrative Computational Design and Construction for Architecture" that will start in 2019 at the University of Stuttgart.

REFERENCES

Image-Based Target Detection and Tracking Using Image-Assisted Robotic Total Stations (9887)
Volker Schwieger, Otto Lerke and Gabriel Kerekes (Germany)

FIG Working Week 2019
Geospatial information for a smarter life and environmental resilience
Hanoi, Vietnam, April 22–26, 2019

Automated Precision Inc. (2018): Laser Tracker Specifications. https://apisensor.com/download/radian-spec-sheet/. Last access: 10.11.2018.

Bay, H., Ess, A., Tuytelaars, T., Van Gool, L. (2006): Speeded-Up Robust Features. 9th European Conference on Computer Vision, 7-13th May 2006.

Bayer, G.; Heck, U.; Mönicke, H.-J. (1989): Einsatz einer CCD-Kamera bei der Objektführung mittels Motortheodolit, Allgemeine Vermessungsnachrichten, Vol. 96, no. 11-12, Karlsruhe

Ehrhart, M.; Lienhart, W. (2017): Accurate Measurements with Image-Assisted Total Stations and Their Prerequisites. Journal of Surveying Engineering, Vol. 143, no. 2, 2017

Girod, B (2013): Digital Image Processing: Edge Detection. https://web.stanford.edu/class/ee368/Handouts/Lectures/2016_Autumn/12-Edge-Detection_16x9.pdf. Letzter Zugriff: 12.10.2018.

Kybic, J., Vnucko, I. (2010): Approximate Best Bin First k-d Tree All Nearest Neighbor Search with Incremental Updates. Research Reports of CMP, Czech Technical University in Prague, No. 10, 2010. Center for Machine Perception, Department of Cybernetics, Faculty of Electrical Engineering, Czech Technical University. ISSN 121 3-2365.

Leica Geosystem AG (2018c): The MS60 GeoCOM Reference Manual, retrieved from www.myworld.leica-geosystems.com, last accessed on 11.2018.

Lowe, G. (1999): Object Recognition from Local Scale-Invariant Features. International Conference on Computer Vision, Corfu 1999, Proceedings.

Luhman, T. (2010): Nahbereichsphotogrammetrie. 3. Auflage, Wichmann, VDE Verlag GmbH, Berlin und Offenbach.

Merziger, G., Wirth, T. (2010): Repetitorium Höhere Mathematik. 6. Auflage, Binomi Verlag, Barsinghausen.

Muja, M., Lowe, D. (2012): Fast Matching of Binary Features. 2012 Ninth Conference on Computer and Robot Vision – CRV'12, IEEE Computer Society, Washington D.C., USA. Proceedings. ISBN 978-0-7695-4683-4.

Niemeier, W. (2008): Ausgleichsrechnung. Walter De Gruyter Verlag, Berlin, New York.

Reiterer, A.; Wagner, A. (2012): System Considerantions of an Image Assisted Total Station - Evaluation and Assessment, Allgemeine Vermessungsnachrichten, Vol. 3, p. 83-94, Berlin, 2012.

Scherer, M.; Lerma, J. L. (2009): From the Conventional Total Station to the Prospecive Image Asssited Photogrammetric Scanning Total Station: Comprehensive Review, Journal of Surveying Engineering, Vol. 135, no. 4

Shapiro, L., G.; Stockman, G.,C. (2001): Computer Vision, Prentice Hall, Upper Saddle River.

Sonka, M., Vaclav, H., Boyle, R. (1994): Image, Processing, Analysis and Machine Vision. Chapman & Hall Computing, London.

Image-Based Target Detection and Tracking Using Image-Assisted Robotic Total Stations (9887)
Volker Schwieger, Otto Lerke and Gabriel Kerekes (Germany)

FIG Working Week 2019
Geospatial information for a smarter life and environmental resilience
Hanoi, Vietnam, April 22–26, 2019

Talukder, K. H., Harada, K. (2007): Haar Wavelet Based Approach for Image Compression and Quality Assessment of Compressed Image. IAENG International Journal of Applied Mathematics, Volume 36, Issue1. ISSN 1992-9978.

Torge, W. (1980): Drei- und zweidimensionale Modellbildung. In „Geodätische Netze in Landes- und Ingenieurvermessung", Hrsg.: Hans Pelzer, Konrad Wittwer, Stuttgart.

Torr, P., Zisserman, A. (2000): MLESAC: A New Robust Estimator with Application to Estimating Image Geometry. Computer Vision and Image Understanding, Jahrgang 78. ISSN: 1077-3142.

Trimble Inc. (2018): Datasheet Trimble S7 Total Station. https://geospatial.trimble.com/products-and-solutions/trimble-s7#product-support. Last access: 11.11.2018.

Wagner, A.; Wasmeier, P.; Wunderlich, T.; Ingesand, H. (2014): Vom selbstzielenden Theodolit zur Image Assisted Total Station in Allgemeine Vermessungsnachrichten, Vol. 121, no. 5.

BIOGRAPHICAL NOTES

Prof. Dr.-Ing. habil. **Volker Schwieger**
1983 – 1989 Studies of Geodesy in Hannover
1989 Dipl.-Ing. Geodesy (University of Hannover)
1998 Dr.-Ing. Geodesy (University of Hannover)
2004 Habilitation (University of Stuttgart)
since 2010 Professor and Head of Institute of Engineering Geodesy, University of Stuttgart
2015 – 2018 Chair of FIG Commission 5 'Positioning and Measurement'
since 2016 Dean of Faculty of Aerospace Engineering and Geodesy, University of Stuttgart

Dipl.-Ing. **Otto Lerke**
2002-2008 University of Stuttgart, Studies of Geodesy and Geoinformatics
2008-2013 Klinger und Partner GmbH, Office for Civil and Environmental Engineering, Stuttgart
since 2013 University of Stuttgart, Institute of Engineering Geodesy (IIGS)

M. Sc. **Gabriel Kerekes**
2008 – 2012 Politechnica University of Timisoara, Romania – B.Sc. Land Surveying and Cadaster
2012 – 2014 Technical University of Bucharest, Romania – M.Sc. Geodesy and Geomatics
2014 – Master Thesis at the University of Stuttgart, Germany
2014 – 2017 intermetric GmbH, Limburg an der Lahn, Germany
since 2017 University of Stuttgart, Institute of Engineering Geodesy (IIGS) Stuttgart, Germany

CONTACTS

Prof. Dr. -Ing. habil. Volker Schwieger
Institute of Engineering Geodesy, University of Stuttgart

Image-Based Target Detection and Tracking Using Image-Assisted Robotic Total Stations (9887)
Volker Schwieger, Otto Lerke and Gabriel Kerekes (Germany)

FIG Working Week 2019
Geospatial information for a smarter life and environmental resilience
Hanoi, Vietnam, April 22–26, 2019

Email: volker.schwieger@iigs.uni-stuttgart.de

Dipl.-Ing. Otto Lerke
Institute of Engineering Geodesy, University of Stuttgart
Email: otto.lerke@iigs.uni-stuttgart.de

M. Sc. Gabriel Kerekes
Institute of Engineering Geodesy, University of Stuttgart
Email: gabriel.kerekes@iigs.uni-stuttgart.de

Image-Based Target Detection and Tracking Using Image-Assisted Robotic Total Stations (9887)
Volker Schwieger, Otto Lerke and Gabriel Kerekes (Germany)

FIG Working Week 2019
Geospatial information for a smarter life and environmental resilience
Hanoi, Vietnam, April 22–26, 2019