

Methods and Algorithms of an Imaging Total Station

Dissertation zur Erlangung des Doktorgrades der
Technischen Fakultät der
Albert-Ludwigs-Universität Freiburg im Breisgau
vorgelegt von Kira Eliza Zschiesche



Dekan: Prof. Dr. Roland Zengerle

Erstgutachter und Betreuer der Arbeit: Prof. Dr. habil Alexander Reiterer

Zweitgutachter: Prof. Dr. Andreas Eichhorn

Mentor: Prof. Dr. Martin Schlüter

Datum der mündlichen Prüfung: (19.10.2023)

Abstract

This work presents different methods and algorithms for the application of Image Assisted Total Stations (IATS). IATSs combine the accuracy of the angle measurement of a total station with photogrammetry and the associated possibility of image processing and analysis. The integration of cameras makes it possible to capture distinctive points on the object without the need for signalling. This offers the advantage that no access to the object is required, as distinctive points can be observed directly and without the use of geodetic targets such as prisms. By using methods such as template-based or feature-based matching, points are extracted and their movement detected. A disadvantage is the dependency on sufficient ambient lighting or illumination of the point to be observed.

First, the developments of various imaging total stations since the year 2000 are presented. A distinction is made between developments by research institutions and commercial versions. The various hardware components used as well as the image processing algorithms are presented in more detail. The focus is on the application in the field of Structural Health Monitoring (SHM). In particular, the photogrammetric methods for determining and tracking unsignalised points are discussed. Due to changing light conditions and atmospheric influences, automation is not trivial. Since deformation measurements are usually performed repeatedly, methods for the detection of discrete points are of particular interest. It is shown that the MoDiTa (Modular Digital Imaging Total Station) measurement system can be successfully used to determine deformations and dynamic characteristics. Using various measurement examples on bridge structures, it is shown how the vertical deflection and the vibration behaviour can be successfully recorded and how they contribute to the assessment of the structure condition.

In the following, a solution approach for the self-calibration of a MoDiTa is presented. This approach extends an existing Total Station modularly by a digital camera and enables calibration directly on site. This extension opens up further fields of application such as high-frequency target acquisition or tracking of moving targets. Calibration is performed using image processing and equalisation methods and is largely automated. The crosshair plane is captured in each image and thus provides identical points in both the camera image and the reference image. Due to the modular design and the associated non-exactly repeatable mounting of the camera, and in order to exclude any possible movement of the camera, the crosshair is continuously observed during the measurement. The methods and image processing algorithms required for this are explained in detail. Using an exemplary self-calibration, the parameters of the calculation are presented and it is shown that they also retain their validity for different distances to the object or focus positions.

In the third section, the use of artificial intelligence in combination with IATSs is considered. For narrowly defined tasks, machine learning methods, in this case deep learning, have proven to be an effective tool. The necessary data sets were self-generated and semi-automatically annotated for the essential training. The present approaches are limited to classification and object detection by means of supervised learning. Two different use cases are explained. The first approach is to minimise the manual intervention of the operator by classifying the captured images from the MoDiTa measurement system. The developed control software is supposed to check the captured image for further use, which is necessary for the crosshair recognition. In order to

determine the crosshair using classical image processing algorithms such as edge detection, an image must be used that is sufficiently exposed and free of disturbances. The crosshair should not be detected in this step, but only the background should be evaluated to see if it is appropriate for further use. So far, this decision is made by the user and offers potential for automation. The second use case involves finding geodetic targets in captured images using IATS and Deep Learning. It examines whether a supposed target (retroreflective and non-retroreflective) can be seen in the captured image and identifies and locates it (object detection). By means of a so-called bounding box, the target found is located in the image and the rough direction to the target in the superior system can be determined via the pixel information supplied. Already implemented applications for target identification are to be supported in this way and extended with additional information.

The various methods and algorithms presented are described in detail and evaluated for their practicality. The respective strengths and weaknesses are discussed and it is shown that IATSs have great potential, but that this potential is not being fully exploited at the present time. Especially the area of image processing and analysis has been neglected so far or the already existing product range of possible methods and algorithms has not been fully utilised.

Zusammenfassung

Diese Arbeit stellt verschiedene Methoden und Algorithmen für die Anwendung von Image Assisted Total Stations (IATS) vor. IATSS kombinieren die Genauigkeit der Winkelmessung eines Tachymeters mit der Photogrammetrie und der damit verbundenen Möglichkeit der Bildverarbeitung und -analyse. Die Integration von Kameras ermöglicht es markante Punkte am Objekt zu erfassen, ohne dass eine Signalisierung notwendig ist. Dies bietet den Vorteil, dass kein Zugang zu dem Objekt erforderlich ist, da markante Punkte direkt und ohne die Verwendung geodätischer Zielmarken wie Prismen, beobachtet werden können. Durch die Verwendung von Methoden wie vorlagenbasiertem oder merkmalsbasiertem Matching werden Punkte extrahiert und ihre Bewegung erfasst. Nachteilig ist die Abhängigkeit von genügend Umgebungsbeleuchtung oder einer Beleuchtung des zu beobachteten Punktes.

Zunächst werden die Entwicklungen von verschiedenen bildgebenden Total Stationen seit dem Jahr 2000 vorgestellt. Hierbei wird unterschieden in Entwicklungen von Forschungseinrichtungen und kommerzielle Ausführungen. Es werden die verschiedenen eingesetzten Hardwarekomponenten als auch die Bildverarbeitungsalgorithmen genauer vorgestellt. Der Fokus liegt bei der Anwendung im Bereich des Structural Health Monitoring (SHM). Im Besonderen werden die photogrammetrischen Methoden zur Bestimmung und Verfolgung von nicht signalisierten Punkten betrachtet. Auf Grund der wechselnden Lichtverhältnisse und atmosphärischen Einflüssen ist eine Automatisierung nicht trivial. Da Deformationsmessungen in der Regel wiederholt durchgeführt werden, sind Verfahren zur Erfassung von diskreten Punkten von besonderem Interesse. Es wird gezeigt, dass der Einsatz des Messsystems MoDiTa (Modular Digital Imaging Total Station) erfolgreich zur Bestimmung von Deformationen und dynamischer Charakteristika genutzt werden kann. An Hand von verschiedenen Messbeispielen an Brückenbauwerken, wird gezeigt wie die vertikale Auslenkung und das Schwingverhalten erfolgreich erfasst wird und zur Beurteilung des Bauwerkzustandes beitragen kann.

Im weiteren Verlauf dieser Arbeit wird ein Lösungsansatz für die Selbst-Kalibrierung einer modularen digitalen bildgebenden Total Station vorgestellt. Dieser Ansatz erweitert eine bestehende Total Station modular um eine digitale Kamera und ermöglicht das Kalibrieren direkt vor Ort. Durch diese Erweiterung eröffnen sich zusätzliche Anwendungsfelder wie eine hochfrequente Zielerfassung oder auch die Verfolgung bewegter Ziele. Die Kalibrierung erfolgt mittels Bildverarbeitungs- und Ausgleichsmethoden und ist weitestgehend automatisiert. Dabei wird in jedem Bild die Strichkreuzebene erfasst und liefert damit identische Punkte sowohl im Kamerabild als auch im Referenzbild. Auf Grund des modularen Aufbaus und der damit verbundenen nicht exakt wiederholbaren Montage der Kamera und um eine mögliche Bewegung der Kamera auszuschließen, wird das Strichkreuz während der Messung kontinuierlich beobachtet. Die dafür notwendigen Methoden und Bildverarbeitungsalgorithmen werden ausführlich erläutert. Anhand einer exemplarischen Selbstkalibrierung werden die Parameter der Berechnung dargestellt und es wird gezeigt, dass diese auch für unterschiedliche Entfernungen zum Objekt bzw. Fokusstellungen ihre Gültigkeit behalten.

Im Weiteren wird die Verwendung von Künstlicher Intelligenz in Kombination mit IATSS betrachtet. Für eng definierte Aufgaben haben sich Methoden des Maschinellen Lernens, in diesem Fall Deep Learning, als effektives Werkzeug erwiesen. Die dafür notwendigen

Datensätze wurden selbst erhoben und für das notwendige Training semi-automatisch annotiert. Die vorliegenden Ansätze beschränken sich dabei auf die Klassifikation und Objektdetektion mittels überwachtem Lernen. Zwei verschiedenen Anwendungsfälle werden erläutert. Der erste Ansatz soll den manuellen Eingriff des Bedieners minimieren, indem die erfassten Bilder des Messsystems MoDiTa klassifiziert werden. Die entwickelte Steuerungssoftware soll das für die Strichkreuzerkennung notwendige aufgenommene Bild für die weitere Verwendung prüfen. Um das Strichkreuz mittels klassischer Bildverarbeitungsalgorithmen wie Kantendetektion zu bestimmen, ist ein Bild zu verwenden welches ausreichend belichtet und frei von Störungen ist. Das Strichkreuz soll in diesem Schritt nicht detektiert werden, sondern lediglich der Hintergrund geprüft werden ob dieser für die weitere Verwendung angemessen ist. Diese Entscheidung wird bisher durch den Nutzer getroffen und bietet Potential zur Automatisierung. Der zweite Anwendungsfall beinhaltet das Auffinden von geodätischen Zielmarken mittels IATS und Deep Learning in erfassten Bildern. Es wird geprüft, ob ein vermeintliches Ziel (reflektierend und nicht reflektierend) in dem erfassten Bild zu sehen ist und dieses identifizieren und lokalisieren (Objektdetektion). Mittels einer sogenannten Bounding Box wird das gefundene Ziel im Bild verortet und es kann über die gelieferten Pixelinformationen wiederum die grobe Richtung zum Ziel im übergeordneten System bestimmt werden. Bereits implementierte Anwendungen zur Zielidentifikation sollen auf diese Weise unterstützt und um zusätzliche Informationen erweitert werden.

Die verschiedenen vorgestellten Methoden und Algorithmen werden detailliert beschrieben und auf ihre Praxistauglichkeit hin geprüft. Die jeweiligen Stärken und Schwächen werden diskutiert und es wird gezeigt, dass IATs ein großes Potential bieten, dieses aber zum jetzigen Zeitpunkt nicht vollständig genutzt wird. Gerade der Bereich der Bildverarbeitung und -analyse wird bisher vernachlässigt bzw. die bereits bestehende Produktpalette an möglichen Methoden und Algorithmen nicht vollständig genutzt.

Contents

1. Introduction	1
2. Fundamentals and State of the Art.....	4
2.1 Image Assisted Total Station for Structural Health Monitoring	4
2.2 Optical-mechanical Implementations for IATS.....	5
2.3 Background on Photogrammetric Distortion	6
2.4 Digital Image Analysis for Geodetic Applications	8
2.4.1 Filter Operations	9
2.4.2 Notes on Image Matching	10
2.5 Deep Learning for Classification and Object Detection	12
3. Articles in Peer-Reviewed Journals.....	16
3.A Image Assisted Total Stations for Structural Health Monitoring - A Review	17
3.A.1 Introduction	17
3.A.2 Materials and Methods.....	18
3.A.3 Image Assisted Total Station (IATS).....	19
3.A.4 Structural Health Monitoring with IATS	28
3.A.5 Summary and Conclusion	31
3.B Self-Calibration and Crosshair Tracking with Modular Digital Imaging Total Station	33
3.B.1 Introduction	33
3.B.2 Measurement System	35
3.B.3 Self-Calibration	36
3.B.4 Crosshair tracking	40
3.B.5 Practical Applications.....	46
3.B.6 Conclusion and Outlook.....	54
3.C Classification and Object Detection with Image Assisted Total Station and Machine Learning	55
3.C.1 Introduction	55
3.C.2 Classification.....	56
3.C.3 Object Detection	61
3.C.4 Conclusion	64
4. Outlook	66
References.....	75

1. Introduction

Geodesy has always been concerned with the recording, representation and monitoring of our environment. In the process, objects of the most varied characteristics are recorded by means of different sensor systems. One sub-area is contactless surveying. This feature can reduce the risk potential for the surveyor and lead to fewer or no installations on the object. Examples include terrestrial laser scanner, photogrammetric sensor systems or imaging total stations. Sensor fusion to form multi-sensor systems is becoming increasingly important. These optical measuring systems show great potential in their wide range of applications. They are used in large-scale applications such as the monitoring of natural objects like glaciers or landslides, but also in small-scale applications like the industrial measurement of components in production.

A special aspect of the development of new measuring systems is automation. Here, work packages are successfully carried out independently with consistent quality without human intervention. At the same time, these systems represent mobile instruments that are repeatedly exposed to different conditions at the place of use. The objects to be recorded are just as diverse. Thus, hardly any object to be observed nor its environmental conditions correspond to each other. Developments such as the automatic detection of retroreflective targets by means of target tracking show that there are individual steps that lend themselves to automation. Repeating functional sequences can therefore be taken over by special sensor technology and associated algorithms and methods. If we take a look at the development towards mobile robotics, we can see that within a measuring task there are changes in the environment and thus also a permanent change in georeferencing or localisation and orientation. Here, the recording and evaluation of the environment takes place, at least in part, in real time. Due to the change in the 3D position of the sensor system, the retrieval of identical points (points of interest) is not trivial and is the subject of research. In the field of image processing, various algorithms have been developed in recent years to find such points (see scale invariant feature transform (SIFT), speeded up robust feature (SURF) and blob analysis). To obtain absolute localisation, Global Navigation Satellite Systems (GNSS) or total stations are often used. However, GNSS requires a free horizon to provide the necessary views to the satellites. Total stations have the disadvantage that they require on retroreflective targets to detect them.

In this work, methods, algorithms and applications of an image assisted total station (IATS) are presented. In the following, this mainly concerns the use of the image information provided. This type of measuring instrument is already used successfully in the geodetic context. Among other things, they have the capability of precise automatic distance and direction measurements (also reflectorless), often automatic target tracking, built-in lasers for signalling and compensators. Furthermore, they are optionally motorised and are equipped on the software side with the most common geodetic calculation programmes. Basically, the trend is towards expanding existing proven instruments with additional sensor technology. Surveying instruments are generally becoming more powerful and user-friendly tools through automation and the use of new sensor technology.

IATS combine the geodetic precision of total stations with the area coverage of image acquisition. Usually, one or more cameras are integrated into the instrument. These serve as an overview camera by means of a wide-angle camera, or they are mounted on the

telescope and installed coaxially to the collimation axis. A distinction can also be made between a modular and a permanently installed design. At present, however, the cameras installed by the manufacturer mostly support the operation. Thus, the view through the telescope can be displayed on a screen and no longer has to be performed by the user. This makes the measurement more independent from the subjective perception of the observer's eye. In addition, standard field survey tasks can be supported by documentation or target acquisition. It is possible to make notes directly in the images for a better understanding of the measurements. The manufacturers provide the calibration parameters for permanently installed systems. However, they do not represent a complete photogrammetric calibration but serve the aspect of operability via the display. With modular systems, it is necessary to determine the calibration parameters for each use due to the design. By calibrating the camera, captured images are directly georeferenced (with additional determination of the distance to the target) and measurements in the image become possible. In this case, self-calibration is used to determine the parameters on site for the camera used. In *Section 3.B.3*, the self-calibration procedure is explained in detail using the example of Modular Digital Imaging Total Station (MoDiTa), a development of the i3mainz, Mainz University of Applied Sciences. The modular design makes it possible to select the camera depending on the application and to extend a common total station with an additional sensor. This makes it possible to use cameras that have better or different characteristics than the fixed ones. For example, the possible recording rate and storage of images with integrated cameras has so far been significantly too slow for many applications, such as in the field of structural health monitoring (SHM).

The great advantage of calibrated cameras is that digital image processing algorithms become applicable. Based on detection and localisation of distinctive discrete points, signalised and non-signalised structures can be used as targets. Often it is not required or possible to use artificial signalling. Applications for automatic detection and matching are given in *Section 3.A.3.3*. Proven methods for localisation, detection and tracking of targets can be used through calibration. This enables further automation for autonomous monitoring systems. Applications such as SHM become feasible. An overview of previous applications in SHM is given in *Section 3.A.4*.

The development of more powerful processing units offers new methods of automatic target detection and tracking compared to classical image processing algorithms. With the help of machine learning (ML), feature extraction, pattern recognition and self-learning decision algorithms can be used in practice. Various task areas have developed in the process. The main tasks of image recognition using ML include classification, object detection and segmentation. Classification is limited to the identification of certain classes or categories. The trained network should independently assign images to a class. Object detection involves classification and localisation of objects. These are marked in the image by means of bounding boxes and are assigned to a class. The segmentation assigns a certain class or category to each pixel and is used, for example, in autonomous driving to recognise objects on or near the road area. In the present work, only the methods of classification and object detection are used (see *Section 3.C*).

In the field of image recognition, Deep Learning (DL) has proven to be a useful tool. DL is a method of ML that uses artificial neural networks (ANN) [1]. Due to their layered architecture, which can be staggered to any depth, they are able to learn even very complex features. However, extensive data is necessary to train the artificial intelligence

on the precisely defined task. This is not the development of a new architecture but the application of DL in the geodetic context of target recognition. With the help of ML, the manual part of the human user should be minimised or procedures that have already been implemented should be supported and additional information should be provided. Using two examples, two different approaches are tested and evaluated with self-generated datasets.

The use of image information from IATS has not exhausted its full potential both in the field of classical image analysis and in the field of ML. In the field of ML, only a few applications have been investigated so far.

In the work at hand, the previously mentioned topics are examined in the context of image processing using IATS and further investigated in terms of their practical suitability. The focus is on the application-related automation of individual work packages. The aim is to minimise the user's manual effort, to provide additional information and to extend existing methods. The methods and case studies presented are intended to demonstrate the enormous potential of this measurement system. The following section first introduces the state of the art and then explains the methods and algorithms used in more detail.

2. Fundamentals and State of the Art

This chapter gives a brief overview of the basics presented in this thesis that are not described in detail in the published articles. Individual relevant topics are addressed and placed in relation to the current state of the art. First, the hardware development is discussed, followed by basics and methods.

2.1 Image Assisted Total Station for Structural Health Monitoring

Section 3.A.3 gives a brief overview of the evolution of IATS after the year 2000. The section refers mainly to applications of the SHM. In the field of commercially available IATS, there has been a trend towards further sensor fusion in one instrument. The IATS produced by manufacturers usually include electro-optical distance measurement (EDM), reflectorless EDM and data registration. Moreover, they are motorised, have image sensors and autofocus, and they have automated aiming, tracking and imaging. The built-in image sensors bring advantages to the user on site as they can act independently of the user's own optical sensors: the eye. Cameras are able to work longer and at higher frequencies, and they also capture a wider electromagnetic spectrum. Radiation that is harmful to the human eye can also be detected in this way, e.g. by laser pointers. The requirement for this is the right sensor. In the case of IATS produced by manufacturers, this cannot be adapted to the application and is therefore a disadvantage compared to modular IATS (with external cameras). The use of an IATS with the option to easily extend the existing system also opens up further fields of application for the end user. Nevertheless, the IATS mentioned are currently among the most advanced geodetic instruments in the world.

With the Leica Nova TM60, Leica Geosystems has defined a total station especially for monitoring tasks [2]. However, as these are not equipped with cameras as standard, but can be added as an option, this is not mentioned in *Section 3.A.3*. The manufacturer mentions permanent and campaign-related monitoring, monitoring of buildings and facilities as well as of transport and energy infrastructures, environmental monitoring and use in mining as fields of application. The image sensor technology corresponds to the TS60 [3]. The main difference between the TM60 and the TS60 is the automatic target recognition for long distances. The autofocus is also optional. These specifications illustrate the ambition to continue using the instrument for the already proven tasks in surveying. The image sensor technology does not open up any improvements or additional fields of application compared to the TS60.

In spring 2022, Topcon launched a new scanning robotic total station, the GTL-1200 [4]. Compared to its predecessor, the optimised workflow should save time. The GTL1200 also features a higher scanning resolution than the GTL-1000. The manufacturer specifies the two integrated cameras with 5 megapixels and a field of view of 60° (diagonal) for the telescope camera and 135° for the panorama camera. A maximum frame rate is not specified. The focus here is more on the use of the laser scanner and the georeferenced 3D scans for construction verification and documentation than on the use of the image information.

By using a scanning multistation with imaging sensors, the term image-assisted scanning total station (IASTS) becomes applicable. [5, 6] show the application of laser scanner data

and image data from an instrument to create a so-called RGB+D (Red, Green, Blue + Distance) image. Here, D stands for the distance determined by the laser scanner. The aim of the investigations was to determine the achievable accuracy under real conditions and to evaluate the applicability for the SHM. A Trimble SX10 [7] was used which achieves a scanning rate of up to 26.6 kHz. In order to compare discrete points from different epochs, different matching algorithms are tested. Based on these tests, the ORB algorithm is determined as optimal in this scenery. ORB stands for Oriented FAST and rotated BRIEF and is a feature detector and descriptor developed as an extended version of FAST (features from accelerated segment test [8]) and the binary descriptor BRIEF (Binary robust independent elementary features [9]) [10, 11]. It is characterised by high enough computing performance, the lowest processing times and the highest number of usable features in the region of interest (ROI) [6]. The investigations carried out show the applicability of the RGB+D method for SHM. It offers the possibility to observe more than one discrete point.

2.2 Optical-mechanical Implementations for IATS

In *Section 3.A.3* different realisations of IATS are presented. According to [12], three different types can be distinguished, depending on the arrangement of the sensor. In the first design (type 1), the image coming from the telescope is imaged again with a lens onto a image sensor (CCD or CMOS). The focus is set to infinity. This corresponds largely to the design of the IATS from the University of Zagreb [13]. In this setup, a camera is mounted on the ocular of the telescope. The MoDiTa measuring system is an extension of type 1 in which the crosshair plane is imaged on the sensor with a lens as imaging optical system. Thus, the crosshair is also imaged. The eyepiece installed by the manufacturer is removed. The industrial camera used in the MoDiTa measuring system records directly in the crosshair plane due to the adapter. There are no changes to the optical beam path in the telescope. The length of the eyepiece holder determines the magnification and how much of the crosshair is displayed on the sensor. The second design (type 2) involves the fixed sealed mounting of the sensor without objective on the unchanged telescope sight. The crosshair is thus no longer visible, but offers the advantage that the sensor is fixed closer to the eyepiece. The DAEDALUS measuring system is an extension of type 2. It does not require any additional optical components between the objective and the CCD chip, but for longer distances it does require an additional meniscus lens in front of the telescope [14]. The third design (type 3) requires the conversion of the scope. The eyepiece and crosshair plate are replaced by a CCD or CMOS sensor and mounted in the focal plane. Manufacturers of IATS often use a further development of this principle today. There, the light beam is deflected by a beam splitter and directed to the sensor. By using the same distance to the lens plane as the crosshair plane, the image is displayed sharply. The original imaging ratios remain largely unchanged. The fixed modification or extension of the components offers the advantage of a one-time calibration of the system with known parameters. *Table 2.2-1* shows a non-exhaustive overview of various implemented designs. Listed are the prototypical implementations specified in *Table A.3-1* and a commercial IATS.

Table 2.2-1 Different designs of IATS according to the position of the image sensor [12].

		Type 1		Type 2 (further development)	Type 3	
structure		The image coming from the telescope is imaged onto the sensor with a lens in focus position at infinity.	The crosshair plane is imaged onto the sensor with a lens as optical imaging system.	The image coming from the telescope plus front lens is imaged without an eyepiece onto a sensor mounted behind the crosshair plane.	The sensor is mounted in the focal plane of the telescope instead of the crosshair.	The sensor is mounted via a beam splitter at the same distance from the lens optics as the crosshair plane.
examples of implementations		University of Zagreb [13]	MoDiTa Mainz University of Applied Sciences [15]	DAEDALUS ETH Zurich [14]	IATS2 Technical University of Munich [16]	MS50 Manufacturer Leica Geosystems [17]
path to the sensor	front lenses	✗	✗	✓	✗	✗
	objective lenses	✓	✓	✓	✓	✓
	beam splitter	✗	✗	✗	✗	✓
	eyepiece	✓	✗	✗	✗	✗
	camera lenses	✓	✓	✗	✗	✗
visible crosshair		✓	✓	✗	✗	✗

2.3 Background on Photogrammetric Distortion

In the following, various mathematical approaches and algorithms are explained that are used in *Sections 3.A* and *3.B*, but are not described in detail.

In close-range photogrammetry, short focal length lenses with a large aperture angle are mainly used, which leads to short recording distances, greater object coverage and more favourable beam intersections. The main focus of close-range photogrammetry is on object sizes from approximately 1 m to about 200 m, with an acquisition distance less than approximately 300 m [11].

As a rule, larger aperture angles lead to greater radial-symmetrical distortions, which must be taken into account during calibration [11]. In the usual distortion model for normal angle lenses, this is not sufficiently taken into account for the image corners.

Many of the telescope lenses installed by the manufacturer have a format angle smaller than 25° and can be described as telephoto (small angle). They are often used to increase the magnification of distant objects. As a rough rule, the focal length of a standard lens corresponds to the diagonal of the image format [11]. Telephoto lenses, such as those used in IATS, usually have a longer focal length. By using a telephoto lens, only the rays close to the axis and thus with little distortion are imaged. This means that the determination of the distortion parameters, which usually act towards the edge of the image, can be omitted for the modelling. The eyepiece adapter for the MoDiTa measuring system shown in *Figure B.2-1* and its built-in between-lenses cause additional distortions.

Radial-asymmetrical and tangential distortions mainly result from the decentration and skewing of individual lenses [11]. Usually, it is assumed that constructionally centred lens systems are used. This means that all surfaces are rotationally symmetrical with respect to the common axis [18]. Due to the modular design of the MoDiTa measuring system, it is not possible to guarantee an absolutely identical and centric design. Therefore, depending on the application, it is necessary to determine all distortions during calibration. In photogrammetry, the individual optical aberrations (exception: radial-symmetric distortion) are not treated according to their causes or occurrence, but only their effects are determined [11].

The automatic target point detection and matching in the digital image offers the advantage that the target does not have to be aimed exactly in the centre, but that the field of view (FoV) can be exploited. As shown in *Section 3.B.5*, maximum values of the distortion parameters can be seen especially at close range (~ 15 m, depending on the camera used).

For IATS with fixed cameras, it is possible to determine the calibration parameters for each change to the optics before using them. These can then be applied automatically without the user having to actively support this. With modular models, the photogrammetric distortions should ideally be re-determined for each change to the optics since the cameras can never be reassembled exactly the same way. In this section, the distortion for the entire optics is determined during self-calibration.

A distinction can be made between radial, tangential and asymmetric distortion, as well as affine errors. These are applied to the measured image coordinates (x'_p and y'_p) in the form of correction functions $\Delta x'$ and $\Delta y'$. The correction values are determined iteratively in a bundle block adjustment with self-calibration and are calculated with the final image coordinates (x' and y').

Radial-symmetrical distortion (A_1, A_2, A_3) is caused by changes in refraction at the individual lenses. It represents the greatest influence below the mentioned aberrations. The parameter A_1 determines whether the distortion is barrel-shaped or pincushion-shaped. A_2 and A_3 are primarily used for modelling towards the edge of the image. The distortion depends on the image radius r' of an image point and increases with increasing distance from the image main point. As usual, r_0 is set at $2/3$ of the maximum image radius. x'_0 and y'_0 represent the principal point.

$$\Delta x'_{rad} = x' \frac{\Delta r'_{rad}}{r'} \quad (2.1)$$

$$\Delta y'_{rad} = y' \frac{\Delta r'_{rad}}{r'}$$

with

$$r' = \sqrt{(x'_p - x'_0)^2 + (y'_p - y'_0)^2} \quad (2.2)$$

$$\Delta r'_{rad} = A_1 r' (r'^2 - r_0^2) + A_2 r' (r'^4 - r_0^4) + A_3 r' (r'^6 - r_0^6) \quad (2.3)$$

Decentralised lenses achieve the radial-asymmetric and tangential distortion (B_1, B_2).

$$\Delta x'_{tan} = B_1 (r'^2 + 2x'^2) + 2B_2 x' y' \quad (2.4)$$

$$\Delta y'_{tan} = B_2 (r'^2 + 2y'^2) + 2B_1 x' y'$$

Affinity and shear (C_1, C_2) describe the deviations from orthogonality and non-uniform scales of the axes. These distortions refer to the sensor and image plane and are not related to the lens.

$$\Delta x'_{aff} = C_1 x' + C_2 y' \quad (2.5)$$

$$\Delta y'_{aff} = 0$$

From this follows for the summarised total correction (see also *Section 3.B.3*):

$$\Delta x' = \Delta x'_{rad} + \Delta x'_{tan} + \Delta x'_{aff} \quad (2.6)$$

$$\Delta y' = \Delta y'_{rad} + \Delta y'_{tan} + \Delta y'_{aff}$$

2.4 Digital Image Analysis for Geodetic Applications

Digital image analysis refers primarily to image acquisition, pre-processing, segmentation, recognition, classification and configuration. The goal is digital information extraction and processing for information extraction. In the case of measurement tasks, these should take place automatically with sufficient robustness and accuracy [11]. In the geodetic context, varieties of different methods have been developed to fulfil this task. The successful use of digital image analysis depends on factors such as the application, the sensors and technology used, as well as the external conditions. *Section 3.A.3.3* deals with various image processing algorithms that can be applied to automatically detect unsignalised points on an object to be monitored.

In the field of SHM using IATS, correlation methods are often used. These methods find application in [19 – 24]. 2022 [25] show accuracy and precision investigations under laboratory conditions. The telescope camera of a Leica Nova MS60 [26] is used in comparison to a DSLR Nikon D3200 [27]. The target is a metal plate with a random speckle pattern. This is moved perpendicularly to the axis of the cameras. COSI-Corr (Coregistration of Optically Sensed Images and Correlation; [28, 29]), an open-source plugin, is used to determine the accuracy and precision. The results show that from a distance of 20 m, the accuracy of the IATS decreases significantly and becomes worse than that of the Nikon D3200 DSLR. These results complement the studies on distance

dependence from *Section 3.B.5.1*. The smaller size of the optical sensor is mentioned as a possible explanation. Smaller sensors with lower resolution are more affected by longer distances [25].

[30] have also addressed the detection of geodetic targets using machine learning with a proof of concept. To merge point clouds from laser scanners and unmanned aerial vehicles (UAVs), there is currently no standardised target design for both techniques. A terrestrial laser scanner (TLS) target design developed by the Institute of Geodesy and Geoinformation at the University of Bonn (IGG) is used as target design [31]. Since this pattern is not automatically recognised by common software, it must be done manually. Deep Learning (DL) methods are used to increase the degree of automation of the work step of searching for the centre points of the targets used. This involves detection, followed by digital image processing to determine the centre points. Similar to *Section 3.C.3*, a pre-trained network is trained for the precisely defined task using transfer learning. After the network has detected the searched target and defined those using coordinates to the bounding boxes, the centre is determined using common image algorithms. The aim is to detect significant edges between the black and white areas and to determine the centre using these. The results show equivalent accuracy using DL and image processing methods compared to manual centre point determination. Overall, object recognition and localisation from camera-based UAV flights using Machine Learning is still the subject of research [32, 33].

2.4.1 Filter Operations

There are many different methods of digital filtering for the detection of features in images. The basis is usually the Fourier transformation, whose signals can be represented as a linear combination of trigonometric functions. Filters are used to highlight or reduce certain frequencies in the image. In the present work, various filter operations are used for the detection of the crosshair. These are largely implemented in MVTec's Halcon [34] image processing software and are applied in the control software developed for the IATS MoDiTa. In order to detect edges, strong changes of the grey values in the image are detected. Edges are extracted by emphasising high frequencies and suppressing low frequencies.

General characteristics for edge detection can be described according to [11]:

- complete extraction of all relevant edges (robustness);
- simple parametrization (preferably without interactive input);
- high sub-pixel accuracy;
- minimum computational effort.

The 2D-Gaussian smoothing filter (2.7) has optimal smoothing properties and leads to smearing effects of edges, depending on the size of the filter. Smoothing filters are often used to reduce grey noise.

$$f(x, y) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right) \quad (2.7)$$

By calculating the second derivative of a discrete two-dimensional function, an edge appears as the zero crossing of the resulting function. The convolution mask is called the

Laplace operator. The operator detects edges in all directions (invariant to rotations) but is very sensitive to noise. Since it increases the noise, it is often combined with the Gaussian filter (2.7). The order of convolution is irrelevant. The second derivative of the Gaussian function is considered the optimal edge filter, which combines the smoothing properties with edge extrusion.

With the second derivation of equation (2.7) follows

$$f''(r, \sigma) = \nabla^2 GAUSS = -\frac{1}{\pi\sigma^4} \left(1 - \frac{r^2}{2\sigma^2}\right) \exp\left(-\frac{r^2}{2\sigma^2}\right) \quad (2.8)$$

$$r^2 = x^2 + y^2$$

This filter is called Lapacian of Gaussian (LoG) and is used in *Section 3.A*.

2.4.2 Notes on Image Matching

Basically, image matching is used to find identical object features in images. This procedure is based on the concept of calculating similarities between the template and the image. The similarity can be defined in different ways. The grey values of the template or the closeness of template edges can serve as a basis [35]. If there are changes in the size of the searched object, similarity transformations become necessary. In the process, the task of matching is to be done automatically. This correspondence analysis must deal with the challenge that there may be no solution at all, that it is not distinct, or that it is too unstable or sensitive to noise.

In the following, three different methods are explained. These find applications in the control software for the MoDiTa measuring system (see *Section 3.A.3.3*, *3.B.3* and *3.B.4*). The application cases using the MoDiTa measuring system are limited to finding only one object at a time in the searched image. It is a priori known that only one object is being searched for.

The normalised correlation coefficient (NCC) is used in *Section 3.A* as well as in *Section 3.B* and is therefore explained again here. Applications are found, among others, in the use of the MoDiTa measuring system. For crosshair matching, this method is used to detect the movements of the crosshair or the camera relative to the crosshair. The current centre of the crosshair in the image is compared with the last crosshair detected [36]. It is suitable for calculating the similarity between two image areas. For this, they must not differ too much geometrically. The best match is calculated between the pattern image $f(x, y)$, referred to as template in the following, and a search image $g(x, y)$. For pattern recognition, the template is successively moved over the search image and the correlation coefficient is calculated at each position. A template is considered to have been successfully detected if the calculated value is greater than its threshold value t . It is calculated from the standard deviations and the covariance.

The correlation coefficient is calculated as follows

$$\rho_{fg} = \frac{\sigma_{fg}}{\sigma_f * \sigma_g} \quad (2.9)$$

with m_f, m_g as arithmetic means of grey values, f_i, g_i as grey values in template and search image with n pixels, $i = 1 \dots n$ and

$$\sigma_{fg} = \frac{\sum[(f_i - m_f)(g_i - m_g)]}{n} \quad (2.10)$$

$$\sigma_f = \sqrt{\frac{\sum(f_i - m_f)^2}{n}} \quad \sigma_g = \sqrt{\frac{\sum(g_i - m_g)^2}{n}}$$

The least squares matching (LSM) method is based on minimising the sum of squares of the grey value differences between the template $f(x, y)$ and the search image $g(x, y)$. The template and the search image differ in the proportion $e(x, y)$ caused by the noise. This method was already developed in the 80s for stereo image evaluation. It can be used for point determination, edge extraction and line tracing in two-dimensional image analysis [11].

$$f(x, y) - e(x, y) = g(x, y) \quad (2.11)$$

The model of the affine transformation with the parameters a_0 and b_0 is used to describe the geometries. For the radiometric transformation, a linear grey value stretching is usually used by introducing two additional parameters r_0 and r_1 [37]. The solution cannot be solved directly because it is a non-linear function. This results in the improvement equation linearised on approximate values to (2.12).

$$\begin{aligned} & f(x, y) - e(x, y) \\ &= g(x', y') + r_1 g_x da_0 + r_1 g_x x da_1 + r_1 g_x y da_2 + r_1 g_y db_0 \\ &+ r_1 g_y x db_1 + r_1 g_y y db_2 + dr_0 + dr_1 g(x', y') \end{aligned} \quad (2.12)$$

The partial derivatives (2.13) are formed by the grey value gradients. For the choice of approximate values $a_0^\circ = a_2^\circ = b_0^\circ = b_1^\circ = r_0^\circ = 0$ and $a_1^\circ = b_2^\circ = r_1^\circ = 1$ can be assumed.

$$g_{x'} = \frac{\partial g(x', y')}{\partial x'} \quad g_{y'} = \frac{\partial g(x', y')}{\partial y'} \quad (2.13)$$

After setting up the adjustment system, it can be solved iteratively.

Separated edges are useful for a variety of geodetic applications and are successfully used in Section 3.B.3 for detecting the crosshair. Shape-based matching is based on linking edges together to form a contour. Instead of grey values, features are extracted along a contour and then used for matching. This makes the methods independent of changes in illumination and variations in grey values. It is more robust against missing object parts, blur, noise, defocus and slightly deformed models [38]. In this method, for each image point $p_i = (x_i, y_i)^T$ of the model, a direction vector $d_i = (t_i, u_i)^T$ is determined. The same applies to the entire search image $e_{x,y} = (v_{x,y}, w_{x,y})^T$. As a similarity measure, the non-normalised dot product of the direction vectors of the model and the image is summed over all points to calculate a match value at a given point of the image [35]. Further on, the affine translational part is separated from the linear one. From this follows:

$$p'_i = Ap_i \quad d'_i = (A^{-1})^T d_i \quad (2.14)$$

$$s = \frac{1}{n} \sum_{i=1}^n \frac{t'_i v_{x+x'_i, y+y'_i} + u'_i w_{x+x'_i, y+y'_i}}{\sqrt{t_i'^2 + u_i'^2} \sqrt{v_{x+x'_i, y+y'_i}^2 + w_{x+x'_i, y+y'_i}^2}} \quad (2.15)$$

If there is a perfect match between the template and the search image, a score of one is achieved. An intuitive threshold value can be set to define a successful recognition.

2.5 Deep Learning for Classification and Object Detection

Object detection, localisation and tracking have long been a part of geodetic work. The applications are becoming more and more numerous and a large number of successful networks have been developed in the field of computer vision. At the same time, object identification and tracking remains a difficult research topic. In the following, the evaluation metrics used in 3.C.2 and 3.C.3 are explained in more detail. The list of evaluation metrics is not exhaustive and includes only the most common or used ones.

To train a network for a specific task, a function is used. This compares the prediction with the given information. By changing the weights, the target function is adjusted and the error is minimised. To evaluate this training, it helps to visualise the loss function. In an ideal case, the curve moves closer and closer to the horizontal axis. If the value remains high, there is an underfitting. This means that the model fails for the type of data given. A possible solution would be to increase the complexity of the network, e.g. by adding more layers or neurons. Overfitting occurs when the model starts to learn by memorising instead of learning general rules for the task. This is illustrated when the graph on the training dataset improves but the graph on the validation dataset decreases. In this work, an already pre-trained network was used whose architecture was used unchanged. In order to influence the course of the function, various hyperparameters were evaluated (see 3.C.2.1 and 3.C.3.1).

The most commonly used metrics are based on the elements of the confusion matrix (*Table A.1-2*). This determines the degree to which the classification was successfully carried out in different classes. In a binary classification, the following types of states can result:

- True Positive (TP), correctly recognised
- False Positive (FP), incorrectly recognised
- True Negative (TN), correctly unrecognised
- False Negative (FN), incorrectly unrecognised

It is worth mentioning that in the case of object detection, a true negative (TN) result is never the case, as there are almost infinite possibilities of not detecting the object in the image [39].

Table A.1-2 Structure of confusion matrix.

		predicted class	
		Positive	Negative
true class	Positive	True Positive (TP)	False Positive (FP)
	Negative	False Negative (FN)	True Negative (TN)

From this matrix, further indicators can be determined, some of them are also mentioned in *Section 3.C.2* and *3.C.3*.

The precision is the ratio between the numbers of positive samples correctly classified to the total number all positive classified samples. It measured the model's accuracy in positive classifying.

$$precision = \frac{TP}{TP + FP} = \frac{TP}{all\ Detections} \quad (2.16)$$

The recall (also called sensitivity) measures the model's ability to detect positive samples. It cares about how the positive samples are classified. If the recall is high, it means it classifies correctly positive. Thus, the model can be trusted in its ability to detect positive samples.

$$recall = \frac{TP}{TP + FN} = \frac{TP}{all\ ground\ Truth} \quad (2.17)$$

The specificity evaluates the model's ability to predict actual negatives. It determines the prediction skill, if samples does not belong to a class.

$$specificity = \frac{TN}{TN + FP} \quad (2.18)$$

Recall and specificity are often used in bio- and medical applications and in studies involved image and visual data [40].

The accuracy describes how the model perform across all classes. It is the ratio between the numbers of correct predictions to the total number of predictions. For a correct representation of this value, the data must be balanced. In practical terms, unbalanced data are common, so that the asymmetric misclassifications must taken into account [41].

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (2.19)$$

The error rate is the opposite of accuracy. It shows how often the model classified incorrectly. It is calculated as the number of all incorrect predictions divided by the total number of the dataset. A value of 0 is ideal, whereas the worst is 1.

$$\text{error rate} = \frac{FP + FN}{TP + TN + FP + FN} \quad (2.20)$$

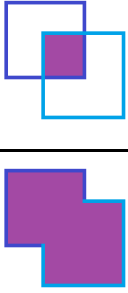
The so-called F-Score can be calculated from precision (2.16) and recall (2.17). This connects the two parameters in a context. By using the harmonic mean, extreme values of one of the two variables are weighted much more strongly. A value of 1 is ideal.

$$F_{\beta} = \frac{(1 + \beta^2)(\text{precision} * \text{recall})}{(\beta^2 * \text{precision} + \text{recall})} \quad (2.21)$$

β is commonly 0.5, 1 or 2. The F-Score is balanced when $\beta = 1$. If β is less than 1, it favours recall. Otherwise, it favours precision [40]. In this thesis, β is assumed to be 1. From this follows

$$F - \text{Score} = 2 * \frac{\text{precision} * \text{recall}}{\text{precision} + \text{recall}} \quad (2.22)$$

For object detection, a common metric is the Intersection over Union (IoU). It is a measure of the correctness of localisation. This metric is based on the Jaccard index, which is a similarity coefficient of two datasets. In terms of object detection, this means that the overlap area of the predicted bounding box and the ground truth bounding box is set in relation to the united area [39]. In this work, the enclosing rectangles are defined exclusively parallel to the axes.



$$IoU = \frac{\text{Area of Overlap}}{\text{Area of Union}} \quad (2.23)$$

By setting a threshold value t , we can classify a detection as true or false. An Intersection over Union score > 0.5 is normally considered a “good” prediction. It applies:

$$\text{correct considered: } IoU \geq t \quad (2.24)$$

$$\text{incorrect considered: } IoU < t \quad (2.25)$$

The Average Precision (AP) value is an average of maximum precisions at different recall values [42]. It is the weighted sum of precisions at each threshold. The weight is defined as the increase in recall. n stands for the number of thresholds.

$$AP = \sum_{i=1}^n [recall_i - recall_{i-1}] * precision_i \quad (2.26)$$

The mean Average Precision (mAP) measures the accuracy of object detection over all classes. It is also commonly used in segmentation and is based on the confusion matrix (*Table A.1-2*), Intersection over Union (2.23), recall (2.17) and precision (2.16). AP_i is the average precision in the i -th class and N is the number of classes [39].

$$mAP = \frac{1}{N} \sum_{i=1}^N AP_i \quad (2.27)$$

In the present application, the mAP is not used as a measure of how well the target centre has been accurately determined, but is only to be understood as a guideline value.

3. Articles in Peer-Reviewed Journals

In this chapter, the three publications in peer-reviewed journals are presented in the *Sections 3.A to 3.C*. The context of the articles is discussed in *Section 2*.

A. Image Assisted Total Stations for Structural Health Monitoring - A Review

Autor: Kira Zschiesche

Journal: *Geomatics* 2022, 2(1), 1–16, <https://doi.org/10.3390/geomatics2010001>

Received: 19. November 2021

Accepted: 21. December 2021

Published: 23. December 2021

B. Self-Calibration and Crosshair Tracking with Modular Digital Imaging Total Station

Autors: Kira Zschiesche, Marvin Fitzke and Martin Schlüter

Journal: *PFG – Journal of Photogrammetry, Remote Sensing and Geoinformation Science*, volume 90, 543–557, <https://doi.org/10.1007/s41064-022-00220-0>

Received: 01. March 2022

Accepted: 07. September 2022

Published: 26. September 2022

C. Classification and Object Detection with Image Assisted Total Station and Machine Learning

Autors: Kira Zschiesche and Martin Schlüter

Journal: *Journal of Applied Geodesy*, 1–9, <https://doi.org/10.1515/jag-2023-0011>

Received: 01. March 2023

Accepted: 24. April 2023

Published: 09. May 2023

3.A Image Assisted Total Stations for Structural Health Monitoring - A Review

Abstract

Measuring structures and its documentation is one of the tasks of engineering geodesy. Structural health monitoring (SHM) is defined as a periodic or continuous method to provide information about the condition of the construction through the determination of measurement data and their analysis. In SHM, wide varieties of sensors are used for data acquisition. In the following, the focus is on the application of image assisted total stations (IATS). The combination of tacheometry and photogrammetric measurement offers high flexibility and precision. Different approaches of automated detecting and matching whose applications have been tested in practice are briefly explained. A distinction is made between built-in cameras (commercial) and external camera systems (prototypes). Various successful applications of IATS in the field of SHM are presented and explained.

3.A.1 Introduction

Structural health monitoring (SHM) deals with the systematic measurement of structures and their characteristics, such as factory chimneys, bridges, wind turbines or dams. These are periodically examined in order to gain possible conclusions about the structural health of the object [43, 44].

Due to environmental impacts and permanent strain, civil structures in particular are affected by damaging influences such as cracking in the concrete, spalling, corrosion of the steel and even failure of the structure. For the years 2017 to 2020, the German government provided EUR 2.9 billion in budgetary funds for the repair and renovation of bridge structures [45]. In addition to the economic costs, safety concerns are coming into focus. As the Ponte Morandi motorway bridge [46] in Italy or the Florida International University pedestrian bridge [47] in the United States have shown, there is an enormous risk for users or pedestrians if structures fail or partially fail. SHM is an efficient and reliable method to monitor the condition of structures and is necessary to detect damage to structures. The most common method is visual inspection, carried out by experienced inspectors. Often this is very time-consuming and leads to traffic obstructions or temporary closure of the structure.

Another approach is the concept of the "smart bridge". Already during the construction phase, bridges are equipped with adaptive sensors to enable a complete evaluation of the bridge. First existing bridges are comprehensively evaluated and then expanded appropriately with sensor technology [48].

The instruments commonly used include accelerometers [43, 49], extensometers [50, 51], terrestrial laser scanners (TLS) [52, 53] or tachymeters [54, 55]. Of particular interest are measurements to determine the vibration and displacement behaviour of the structures. It is possible to analyse the monitored vibrations and calculate the natural frequencies. Changes in this natural frequency indicate possible structural damages and require further investigation by qualified experts. It is also possible to compare the parameters determined by the measured values with the calculation from a finite element model (FEM) [13, 56]. By this way, measured values can be compared with calculated values and analysed.

Another complementary measuring system to the SHM is the use of camera sensors, e.g., in form of so-called image assisted total station (IATS). IATS are characterised by the extension of a total station by using one or more cameras. The accuracy of the angle measurements of the total station is combined with photogrammetry and the associated simple target mark definition. By using telescope lenses, the distance to the structure is increased; depending on the situation, the danger zone does not have to be entered. Signalling on the object can be omitted due to the photogrammetric evaluation. However, this depends on the specific monitoring object. If there is not enough texture or contrast on the structure, signalisation may be necessary. Numerous examples of applications with IATS and signalling can be found in the literature [13, 14, 19, 57, 58].

Deformations are evaluated by capturing image information from a camera and natural frequencies can be calculated. Without continuous distance measurements, a two-dimensional movement parallel to the camera sensor can be measured. The measurement field is limited by the field of view (FoV). For the measurement of frequencies, the maximum frames per second (fps) are essential. Only by an adequately frequent sampling rate the natural frequency can be determined from the measured values.

In this review, we introduce the IATS (*Section 3.A.3*), classified into commercial instruments and research prototypes (with external cameras). The focus will be on the IATS extending the total station with an external camera to substitute the camera integrated by the manufacturer. We will present the image processing methods used to determine targets in *Section 3.A.3.3*. Following this, various field tests of SHM with IATS on structures and the resulting outcome will be presented (*Section 3.A.4*). Finally, the results are discussed, and the practicability of the system is evaluated.

3.A.2 Materials and Methods

The current systematic review follows the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) [59, 60]. The research performed in June 2021 to find the relevant papers was conducted by browsing “Image Assisted Total Station*” AND “Structural Health Monitoring” and “Video tachymetry” AND “Structural Health Monitoring” on Google Scholar and EBESCO Discovery Service, defining the period from 2000 to 2021. The first step excludes the duplicates and all papers that are not in English or German. The next step involves the assessing of the collected papers based on irrelevant abstracts or content. The research resulted in identifying 39 relevant papers. The papers were classified into three topics: camera-only system, built-in and external camera system for total stations. Camera-only systems are not discussed in the following because of the lack of a total station and therefore these seven papers were also excluded.

Fifteen additional papers were found by reviewing citations of the remaining 32 relevant papers whereby 10 of these were classified as suitable. This results in the total number of 42 relevant papers.

The following *Figure A.2-1* presents the processing steps through the different phases of finding relevant papers for this systematic review.

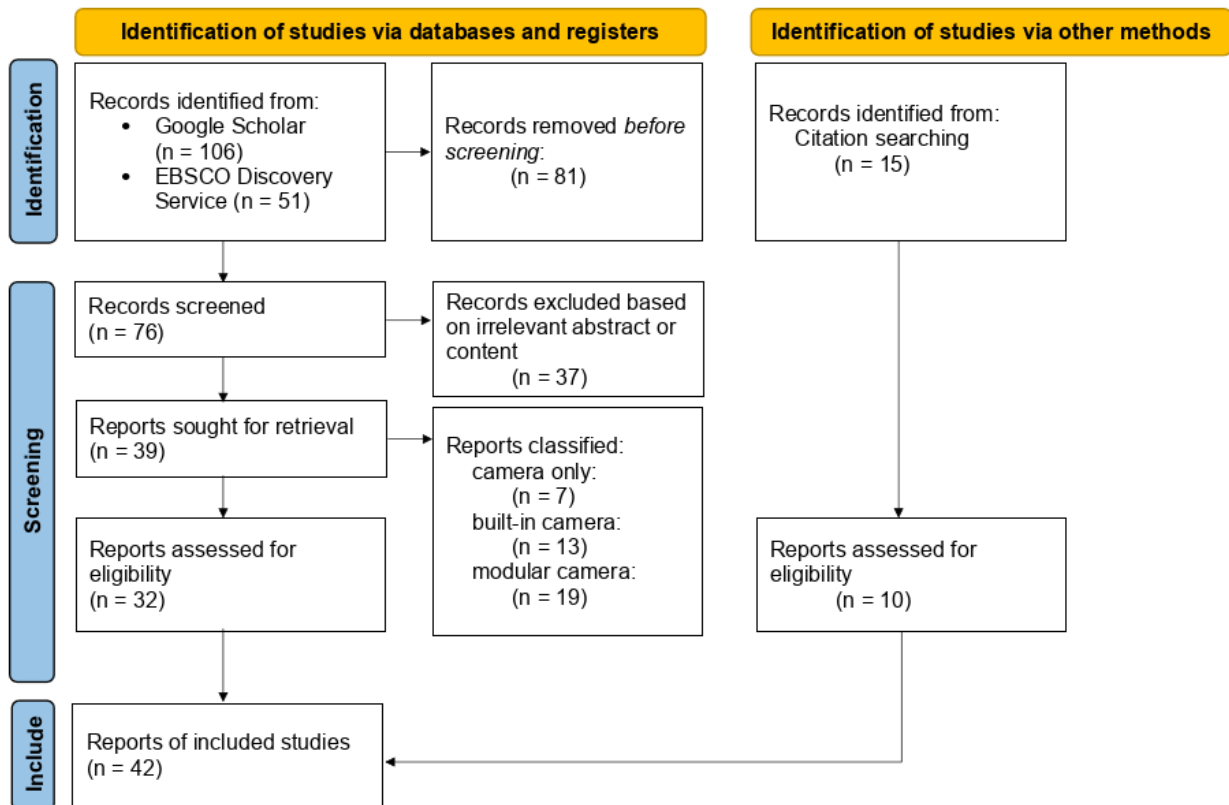


Figure A.2-1 Chart for identifying relevant papers [59, 60].

3.A.3 Image Assisted Total Station (IATS)

Since the beginning of 2000, the development of IATS has continued. For many manufacturers, the integration of one or more cameras into a total station has become standard. The integration of cameras enables to record the relevant scene and offers additional potential for the evaluation by the image-based recording. The combination of angular measurement accuracy from tacheometry in conjunction with image processing of captured image data has expanded the applicability also in the field of SHM. At IATS's, a camera is installed in the line of sight and the visual field is captured by the image. Limitations are the FoV of the camera used and the need for light (at least for the observed object). The automatic acquisition of images removes the need for human aiming and thus the potential impact of error. A major advantage is that there is no need for access to the structures, as prominent points on the structure can be observed directly without the use of reflectors. Measurement points can be flexibly selected over the entire structure, if a clear view of it is accessible. The focus here is on the mobile use of IATS. Due to the advantage of non-contact measurement, this measurement method can also be used spontaneously.

Using image-processing techniques such as template matching or feature matching, distinctive points can be extracted and their movement detected [11, 61]. Various algorithms have been tested for their practical suitability and are explained in more detail in this paper.

To achieve the desired accuracy, a calibration of the different sensors is necessary. For fixed commercial systems, the calibration parameters are provided by the manufacturer. Some of these calibrations are carried out under the aspect of using the display, e.g., to

rotate the instrument, and does not represent a complete photogrammetric calibration. For modular systems or external cameras, the parameters have to be determined in advance. By using a calibration, the captured images are directly georeferenced and measurements in the images become possible.

A basic distinction is made between commercial total stations, where the manufacturer integrates the camera as a sensor, and research prototypes, where an external camera is attached to the total station instead of the ocular. Advantages and disadvantages are explained and discussed.

An overview of the entire historical development of photo and video theodolites up to today's IATS is described in [13].

3.A.3.1 Commercially Available Systems

The technical development up to today's state of the art reaches into the last century. Based on the technical development of tacheometers, these were successively expanded by further software, sensors and technology up to today's IATS. According to [13], total stations can be divided into 4 different types (total station, robotic total station, image-assisted total station and image-assisted scanning total station) based on the integrated sensors and functions. Accordingly, IATS includes data registration, electrical distance measurement (EDM), reflectorless EDM, motorization, image sensor, automated aiming, tracking and imaging. By adding a scanning function, the term image-assisted scanning total station (IASTS) becomes applicable.

Today's instruments usually have the possibility to store additional information of the measured points by the captured images and videos. Images are digitally linked to surveyed points and stored externally for documentation purposes via memory card or interface. The collected georeferenced image data are available for photogrammetrical evaluation. The development of autofocus and motorisation of the instruments allows for a point of interest to be selected in the image (with the help of the display) and the instrument aligns automatically. This way of data acquisition optimises the previous measurement process. In the event that the instrument is equipped with a scan function, the scan area can be defined in the display using the image-based selection option.

Various manufacturers of IATS have established their products on the market; in the following, some instruments are briefly introduced.

By integrating two VGA cameras into a total station, Topcon created a new generation of IATS in 2005 [13]. A wide-angle camera with fixed focus (FoV $28^\circ \times 22^\circ$), was mounted next to the telescope enabling it to rotate accordingly. The second coaxial camera (FoV $1^\circ 30' \times 1^\circ 30'$) records the image section through the telescope with variable focus. In 2019, Topcon launched the GTL-1000, an instrument that fits under the IASTS designation, equipped with a full dome laser scanner integrated into a robotic total station. According to the manufacturer, a 5 megapixel CMOS sensor is installed [62] but there is no definition for the possible framerate because GTL only allows capturing a single image. The GTL-1000 has no function to provide images continuously.

The Leica Nova TS60, Leica Geosystems offers an IATS for permanent or campaign monitoring for bridges, buildings and steel structures [3]. The motorised IATS has two integrated cameras. The overview camera with a FoV of 19.4° and a telescope camera

with 1.5° . It is equipped with a 5 megapixel CMOS sensor with a resolution of 2560×1920 pixels. The achievable frame rate is up to 20 frames per second.

With the Trimble S9/S9 HP, Trimble labelled a total station for monitoring applications in 2015. The integrated camera (not available in all models) has a FoV of $16.5^\circ \times 12.3^\circ$ with a resolution of 2048×1536 pixels. The sensor specification is only a colour digital image sensor [63]. The instrument has a simple scan mode. The Trimble SX 12 [64], which was introduced at the beginning of 2021, includes an overview and a primary camera, which are both positioned parallel to the measurement axis, a telescope camera with coaxial position and a plummet camera for stationing. The cameras integrated in the telescope are already calibrated. The resolution of the chips is 3296×2472 pixels. With the telescope camera, one pixel corresponds to 0.69 mm at a distance of 50 m. The Trimble SX 12 has no eyepiece hence controlling and aiming is done via the field computer. Since this instrument already has advanced scanning functions it can be classified as an IASTS similar to the Topcon GTL-1000.

Despite the large number of possible instruments, according to [65, 66] the full potential of commercial IASTS has not yet been fully realised.

3.A.3.2 Research Prototypes

In the development of prototypes, different types of construction have emerged since the year 2000. On the one hand, there are external implementations making it possible to mount the camera on the ocular or to replace it. These are used in combination with commercial total stations or tacheometers, and can be converted and adapted to the monitoring conditions. The technical development of the automated focus in the total station made the enormously simplified combination of external cameras possible. Examples of such a modular system would be DAEDALUS [14, 58, 67, 68] of ETH Zurich, MoDiTa [24, 69, 70] of i3mainz or the applications of the University of Zagreb [13, 20] using a GoPro5.

In 2010, [14] presented DAEDALUS, a measurement system originally for automatic online astro-geodetic observations, and also for the use in SHM. The setup includes replacing the eyepiece with a CCD chip. Since no further optical components are installed, the image is no longer displayed exactly in the plane of the crosshair instead the image is now displayed in the plane of the CCD chip. This can be compensated for distances up to 13 m by changing the focus position. For longer distances, an additional lens is required on the telescope. The camera used is a monochrome Guppy F-080C from Allied Vision Technologies (AVT). The CCD sensor has a resolution of 1024×768 pixels with a pixel size of $4.65 \mu\text{m}$ and a frame rate of 30 Hz (full frame) up to 60 Hz (reduced field of view) [68]. At that point in time, no automated focusing was available for tacheometers. Therefore, for the necessary focusing, a special mechanism was developed enabling to autofocus by means of a small stepper motor without modification of the total station. When using newer generations of total stations or multi-stations with coaxial cameras already installed, the focus mechanism is already available and the special mechanics for focusing is no longer needed. The compatibility is given for models of the TCA, TPS, TS and MS series from the manufacturer Leica Geosystems [58]. For applications with required software pulses, control by means of a GNSS receiver can be implemented for

precise control of exposure start and exposure time. The optical system has a resolution of approximately four arcseconds/pixel (1.1 mgon/pixel).

The prototype MoDiTa developed by i3mainz (University of Applied Sciences Mainz) is based on modularity. Via bayonet ring the camera replaces the standard eyepiece of the tacheometer. The camera used is mounted on the eyepiece using an adapter. By this way the crosshair plane is also captured. There are no changes at the optical beam path in the telescope. This allows a quick exchange of the camera or a quick change to the classical application of the total station. Based on the modular design both, camera and total station, are replaceable. Attached to the telescope is a counterweight to compensate for the weight of the camera. Due to the modular design of the external cameras, active crosshair tracking is required during the measurement. Because of longer pauses between measurements or after changing the telescope position, the crosshair position may change. In [24], a motorised theodolite TM5100 and TM5100A from Leica Geosystems is extended by means of a CCD camera and used for collimation and autocollimation measurements. As an example, in [69] a monochromatic CMOS sensor camera from the manufacturer IDS with a resolution of 1280 x 1024 pixels is used for the measuring. The camera achieves a frame rate of 60 Hz for full images, up to 2200 Hz for image sections and up to 62500 lines per second in single line mode. Similar to [14], a precise time signal can be produced using a GNSS receiver. The camera and the total station are controlled via own software development and GeoCom or USB ports. The necessary calibration can be carried out in situ and is described in more detail in [70].





Another prototype was presented in 2017 by combining a Leica TPS1201 and a GoPro5 Hero camera [20]. This camera offers different video recording modes. It is possible to choose between different FoV (narrow, linear, medium, wide and super-view), resolutions (720 x 400, 1280 x 720, 1920 x 1080, 2560 x 1440, 2704 x 1520, 3840 x 2160 pixel) and recording speeds (30 - 240 Hz). The field of view and the recording mode are directly related. An adapter is required for the camera mounting on the TPS1201. This was manufactured in a 3D printing process and offers the possibility to attach the camera to the eyepiece of the telescope. According to [13], there is no vertical movement of the mounted camera on the telescope after stability tests in the laboratory. The instrument is controlled via interfaces through a laptop. The camera is managed by a smartphone application. For the evaluation of the video sequences, the video has to be converted into images. These images are orthogonally projected using the known azimuth and zenith angles from the total station and converted from an RGB image to a binary black and white image. After these processing steps, the pixel coordinates of the searched target can be calculated by means of image processing.

The second design offers the advantage of a fixed camera with the instrument. This provides constant calibration parameters as opposed to the modular version which requires calibration after reconfiguration. An early prototype is mentioned 2004 in [71] and the prototype series IATS2 from the manufacturer Leica Geosystems in [16, 57, 66, 72 – 74]. Based on the knowledge gained from [71], a small series of instruments was produced in 2007, based on the TCRA1200 series. The eyepiece and crosshair were replaced by a CMOS colour chip with a resolution of 2560 x 1920 pixels at 2.2 μm pixel size. One pixel on the image sensor corresponds to an angular value of 0.61 mgon. In [72] a possible frame rate of 5 to 200 Hz is specified, depending on the measuring mode. By connecting the focusing ring to a servomotor, automated focusing is possible. Camera and total

station are addressed via GeoCom or USB interfaces. Since this prototype no longer has a crosshair, this must be specified in the image for the necessary calibration. In [16], the chip centre is defined as reference point and the calibration is carried out according to the principle of the virtual control point field. A detailed description of the calibration can be found in [71, 74]. The measurement accuracy of a target detection that can be achieved shows subpixel accurate determinations with an angular accuracy of > 0.6 mgon under laboratory conditions and thus corresponds to the usual accuracy of such algorithms.

In *Table A.3-1* the different prototypes are compared with their main specifications and characteristics described in the papers mentioned.

Table A.3-1 Main specifications and set-up of IATS prototypes mentioned in [13, 14, 16, 75].

	DAEDALUS ETH Zurich [14]	MoDiTa i3mainz - Mainz University of Applied Sciences [75]	University of Zagreb [13]	IATS2 Technical University of Munich [16]
Tacheometer/ total station				
	TCA 1800, TCA 2003 and TDA 5005 ²	MS 50 ²	TPS1201	based on the TCRA1200 series
camera	Guppy F-080C	UI-3250 ML ¹	GoPro Hero 5	fixed camera from manufacturer
resolution [pixel]	1024 x 768	1600 x 1200	up to 3840 x 2160 ³	2560 x 1920
pixel size [μm]	4.65	4.5	1.5 ⁴	2.2
mgon/pixel	1.1	0.4 – 1.0 ⁵	–	0.61
frame rate	30 Hz/60 Hz (full frame/ reduced FoV)	60 Hz/2500 Hz (full frame/ reduced FoV)	30 – 240 Hz	5 – 200 Hz
sensor	CCD	CMOS	CMOS	CMOS

¹ Based on modularity the camera is changeable.

² Compatible to TCA, TPS, TS and MS series from the manufacturer Leica.

³ Multiple resolutions possible. Resolution mentioned in [13]: 1920 x 1080 pixel.

⁴ Estimated by the author.

⁵ Based on modularity this value is variable.

3.A.3.3 Automatic Detection and Matching

The great advantage of using an IATS is that even non-signalized points can be monitored, provided they are distinctive enough. Due to changing light conditions and atmospheric impacts, automation is not trivial.

It is important that the observed point can be clearly identified within the captured images. By calibrating the camera and the total station, the pixel coordinates can be transformed into directions. By additional distance measurements, with a previously calculated orientation, 3D coordinates of a global reference system of the observed point may also be calculated. Alone by the high number of images, which are seized with a vibration observation for example, an automatic detection and retrieval of the prominent point is essential.

In photogrammetry, numerous methods have been developed for these tasks. As with all optical methods, the existing light conditions play an important role. The accuracy that can be achieved depends on various aspects such as object size, resolution and correct focusing [39]. Since deformation measurements are to be performed repeatedly, methods detecting discrete single points are of particular interest, but there are also area-based approaches.

According to [71, 76] three main categories of matching are classified. Feature based matching extracts distinctive features in images by an operator in the first step and detects these features again in the various images to be analysed in the second step. In image processing, a feature is understood as information about the content of the image, such as points, edges or objects. By introducing additional information in the form of knowledge or rules, the search space can be reduced and thus mismatches can be minimized [11]. The features are usually stable to illumination changes, but these methods are sensitive to noise and performance declines on images with a lot of texture. Common methods include scale invariant feature transform (SIFT), speeded up robust feature (SURF) and blob analysis.

There are four steps in the SIFT operator to determine distinctive points and a corresponding vector. The first step involves determining edges in image pyramids and their extrema, followed by locating feature points. Afterwards, the main directions of the gradients are determined. The last step is the derivation of the descriptor [11]. This method finds application in [21, 77].

The SURF operator was developed at the ETH Zurich. The Algorithm is similar to the SIFT operator but needs less computation time. The operator computes the maxima of the determinant of the Hessian and the second derivations of the grey value distribution to find points of interest [11]. This detection of points of interest method is used in [21, 72, 77 – 79]. [72] shows a comparison of different detection methods with the conclusion that SURF delivers results differing by a factor of 2. Likewise, unfocused images show a significant decrease in accuracy.

Blob analysis is a special case of image analysis that segments blob-like patterns. Threshold operations extract features from connected pixels from the background based on different grey values. These operators are fast and robust [21]. In [66] the target on a bridge structure is easily extracted using blob analysis and the pixel position of the centroid is successfully computed.

Raster or area based matching is based on the matching of grey values. The algorithms try to find mapping functions in the image divided into raster by directly comparing the grey values or functions of the grey values of the images. This approach is quite fast and can still give good results even with changes in illumination. These include least squares matching (LSM), correlation based matching, optical flow and the use of geometrical primitive as additional information to increase the accuracy and reliability of the methods.

LSM uses an iterative geometric and radiometric transformation between the reference image and the search image. In this process, the sum of the grey level differences between the two images is minimized [11]. A previously defined template is used as the reference image for matching. In [14, 67, 68] LSM is successfully used for the calibration of the measurement system. Here, the result is considered successful if the empirical standard deviation of the unknown translations is a value smaller than 0.5 pixels. LSM is used by [80] in combination with an ellipse fit to detect the position of circular targets in the image. The workflow here involves the use of an image section, which is converted to an 8-bit grayscale and to a binary image. The image coordinates of the searched contour are computed by border following. Then, using the estimated approximate values, the ellipse is fitted according to the Gauss-Helmert model. The iterative calculation stops as soon as the coordinate changes are less than 0.01 pixel.

Correlation based matching is an image processing method which calculates a similarity measure between a reference pattern and a target image section. It is assumed that the position with the highest match corresponds to the position of the searched pattern in the image [11]. Among other things, this method finds application in [19 – 24] for determining the crosshairs in the measurement image, detecting natural targets, or as an aid to approximate values in a barcode reading. [22] shows that the searched circular target marks can be detected with a standard deviation of 0.1 mgon, which is a significant increase compared to the ATR measurement of current Leica Geosystems total stations in prism tracking mode (0.3 mgon).

Another common method of image processing is the optical flow method. The apparent motion of an imaged object in the field of view is calculated by the intensity changes of an image sequence. It is assumed that it does not change the grey values of identical object points in directly sequential images, provided that there is no occlusion or illumination change [11]. The optical flow methods used in [19, 22, 23] show comparable results to other matching methods for the matching of natural targets and the evaluation of resulting frequencies.

Relational or structural image matching is not based on the similarity of grey levels or point distributions. This matching method is based on the similarity of topological relations. These are stored in feature adjacency graphs. An advantage of this method is the use of topologies as image features, which are not changed during the perspective transformation [71]. For this matching method, no applications for SHM with IATS were found in the research.

Nevertheless, the software developed by i3mainz to control the prototype MoDiTa offers shape based matching. The software makes use of commercial image processing. This method uses the shapes of contours to describe the pattern image, and is robust against occlusions, blurs and illumination changes and is often used in industry [81]. The pattern image is specified by a region of interest (ROI) and defined by points with associated

direction vectors. The points and vectors are determined by edge extraction within the ROI and calculation of the grey value gradients at the points. As a similarity measure in this approach, the pattern image is compared with the search image in which the vector product of the normalised direction vectors is calculated. A score value between 0 and 1 and the transformation parameters are provided as a result for all instances of the object found in the image.

Table A.3-2 shows the comparison between correlation and shaped based matching using the control software for the prototype MoDiTa. An IDS UI-3240CP-M-GL camera with a pixel size of 5.3 μm was used. For this case study setup, one pixel corresponds to a value of 0.50 mgon. In both examples, the scores of matchings are between 0.71 and 0.76. They can be classified as roughly equivalent. The lighting conditions were identical. With the help of image processing a disturbance of the visual field by a wooden branch was simulated. The results show a difference in the localisation of the centre of the target of several pixels with the correlation based method. In both examples, the deviations compared to the undisturbed image are greater with the correlation based method. The target centre deviates by up to 2.91 px in the y-direction and up to 1.74 px in the x-direction. In contrast, the deviations with the shape based method are only 0.14 px in the y-direction and only 0.02 px in the x-direction.

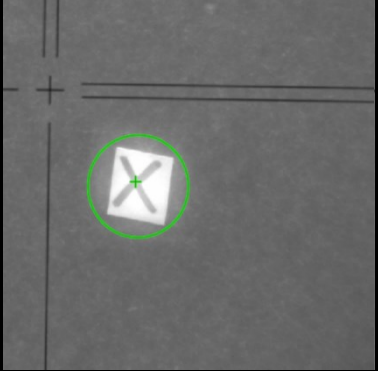
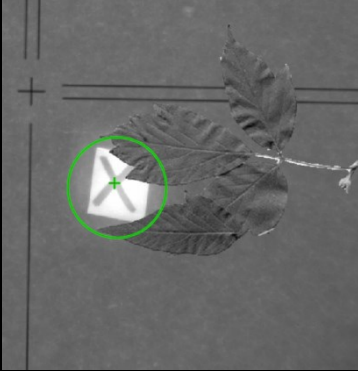
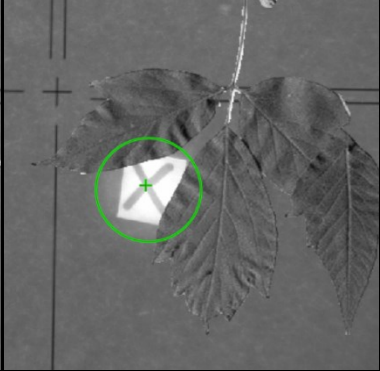
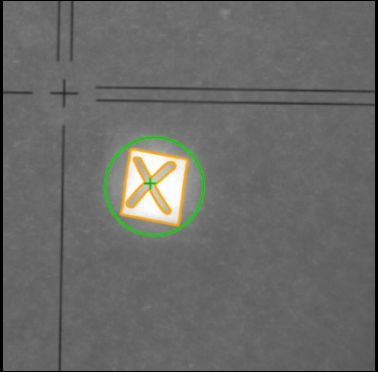
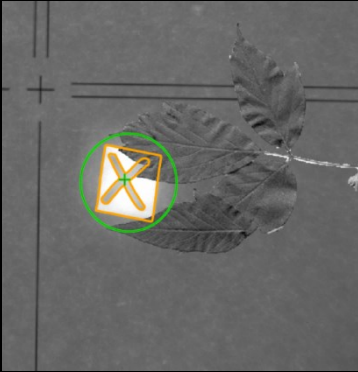
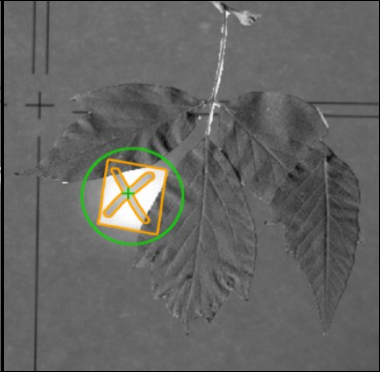
Typically, correlation based matching is preferred for SHM applications, because even targets with low contrast can be detected successfully. For measurements where it is expected that the target will be interferenced or partially covered, e.g., by passers-by, vehicles or moving vegetation, shape based matching offers advantages.

Recent approaches consider the use of artificial intelligence (AI) for detection and matching. [81] compares a deep learning (DL) approach with the classical shaped-based matching algorithm. The outcome of the investigation shows better results for the shaped-based approach as long as the necessary requirements are fulfilled. Shaped-based matching shows its advantages through its high accuracy and precision, high speed and low memory consumption. On the other hand, the DL approach offers a valid alternative for the simultaneous detection of multiple objects or objects of the same shape, or for objects that are not fixed. A detailed survey of classical and modern image matching methods can be found in [82].

Nevertheless, an approach using AI is developed and tested in [83]. The aim was to use AI in the selection of a suitable interest operator and in the classification of façade types. This would provide an automatic tool that reliably supports decision-making and thus enables successful monitoring of building façades.

However, AI methods have not yet been used for single target point recognition by means of IATS. Since these methods usually need large datasets to be able to perform the necessary training sufficiently, the use of correlation or shapes based matching to create such datasets would be conceivable. The user would then have to interact less and the measurements would have a higher degree of automation by means of AI.

Table A.3-2 Comparison of the results for correlation based and shaped based matching using the MoDiTa prototype and its software. The differences (Δy , Δx) shown represent the change in pixel position relative to the template definition.

Matching Method	Undisturbed Target Position	Example 1	Example 2
Correlation based			
	Deviation from the target position compared to the undisturbed target point:	$\Delta y = -2.91 \text{ px}$ $\Delta x = 1.27 \text{ px}$	$\Delta y = -1.22 \text{ px}$ $\Delta x = 1.74 \text{ px}$
Shape based			
	Deviation from the target position compared to the undisturbed target point:	$\Delta y = 0.14 \text{ px}$ $\Delta x = 0.02 \text{ px}$	$\Delta y = 0.24 \text{ px}$ $\Delta x = 0.02 \text{ px}$

In addition to the methods of target detection explained so far, there is also research into automatic target recognition (ATR or ATRplus [84]) by the manufacturer Leica. Other manufacturers, such as Topcon, also offer these methods (here X-pointing [85]), but since the previously mentioned publications mainly use instruments from Leica, the term ATR will be used in the following. In this technique, the reflected laser beam is projected onto a CCD or CMOS sensor chip. The ATR technique requires the use of prisms for high-precision measurements [14, 68], which cannot always be installed on the measurement object. Another disadvantage is that an ATR zero offset cannot be eliminated by measuring in two layers [14]. The advantage here is that the detection of the beam detected on the sensor without focusing is very simple and robust [14]. [22, 65, 86] show the use of an image-based analysis by means of IATS can achieve the same or even higher accuracies than with conventional robotic total stations in conjunction with ATR.

3.A.4 Structural Health Monitoring with IATS

In the field of SHM using IATS, mainly deformations and vibrations of the structure are measured. However, other applications are also being investigated. A basic distinction can be made between static and dynamic monitoring [23, 58, 87]. Static deformation measurements are usually carried out in the form of set measurements. For example, dynamic measurements are carried out in the SHM on bridges or wind turbines in order to map the frequency behaviour of the construction objects. Changes in the frequency response can be used to determine damage to the structure [88]. For dynamic monitoring, the maximum image frequency is important. According to Nyquist, the recording frequency must be at least twice the frequency of the oscillation. In practice, it turns out that a multiple of the frequency to be determined makes sense. Since bridges usually have a natural frequency range of 0.1 to 25 Hz, this has to be taken into account, especially for commercial IATS with much more limited recording frequencies than the prototypes [23, 89]. In addition, to be considered is that the vibration amplitude must be higher than the measurement resolution. The amplitude would otherwise not be identifiable in the measurement noise [23].

By observing a significant point or section on the structure, movements are measured transverse to the direction of view. For a 3D measurement, the use of the EDM or several IATS is necessary. Thus, a forward section is used to calculate the change. Distance measurements are time-consuming compared to image-based and angle measurements. It is not always possible to carry out angle and distance measurements simultaneously with image acquisition.

In the following paragraphs, applications with commercial IATS are explained first, followed by investigations using prototypes.

The application of commercial IATS for bridge monitoring is shown in [19, 22, 23, 80, 90, 91] using the example of the Augarten footbridge (steel construction, 74 m span width) in Graz, Austria. For the measurements, an MS50 [17] from the manufacturer Leica Geosystems was used with different sensors. The IATS has a maximum frame rate of 10 fps. For video recording, a frame rate of 20 fps is possible, but this cannot be transmitted at this speed via the interface. The bridge was stimulated by pedestrians. Both natural targets and prisms were observed. To compare the results, parallel measurements were taken with an acceleration sensor. The study also investigated to what degree the capturing frequency can be increased, e.g., by omitting the distance measurement. In this case, only displacements orthogonal to the viewing direction can be detected. The image-based measurements partly achieve a higher angular resolution than the measurements on a prism. Overall, the results show the successful use of a commercial IATS for dynamic SHM.

Another use of commercial IATS for monitoring bridges is shown in [89]. The stimulation of the 27 m pedestrian bridge was uncontrolled by pedestrians. The measured results could be validated by calculated results using an FEM. For the natural frequencies, only the 3.642 Hz could be verified due to the low sampling rate of the Leica Nova MS50 [17]. The further natural frequency of 13.294 Hz is not detectable due to the limited sampling frequency. [92] show in a proof-of-concept study the use of a MEMS accelerometer and an IATS. Here, the estimated model parameters from the IATS are supposed to counteract

the degradation of accuracy over time for the accelerometer due to coordinated updates. This method shows significantly improved results as well as sub-millimetre accuracy for the displacement and better than 0.1 Hz for the frequencies.

The combination of laser scan data with image data (both from a so-called IASTS Leica Nova MS60 [26]) for use in SHM show [78, 93]. The scan centre and the projection centre of the camera (almost) correspond in this instrument. This means that the same perspective is captured. The acquired data have the advantage of being in the same coordinate system. For evaluation, the images are combined to form a spherical panoramic image (RGB). These are then supplemented with a channel D (depth), i.e., the interpolated distance information from the scans. By means of a congruence model, a deformation analysis becomes practicable. The successful evaluation of the different epochs is shown by the displacement vectors using the SIFT operator.

For the application of façade monitoring, [94] use a IATS-prototype developed by Leica Geosystems in combination with a knowledge-based systems (KBS) and later also with a terrestrial laser scanner (TLS) [95, 96]. One of the main goals of this work was to extend the deformation analysis of single points and to use a point set based method instead. With the help of the KBS, an automated image-based online measurement system for the rapid feature extrusion of prominent façade points is presented. Here, image acquisition, image pre-processing and point detection are automated by the KBS using image feature extraction for deformation analysis. The results show a significant reduction of the necessary decisions by the user and provide a fast performance of the feature extraction.

The use of the prototype DAEDALUS for SHM with optical target recognition (OTR) is shown in [14]. A torch was mounted on the middle of a steel bridge and monitored at a distance of 40 metres. The stimulation was provided by the passage of a truck. The images were released during the day with a measuring rate of 15 Hz. Using a fast Fourier transform (FFT), the dominant natural frequency of 2.59 Hz was successfully determined. In another study [58] show the use of several synchronised QDaedalus for the determination of 3D positions at high frequencies (60 Hz). For this purpose, an innovative hybrid structure made of hardwood by the Institute of Structural Engineering of ETH Zurich was observed. For comparison, acceleration sensors with a recording rate of 1 kHz were attached to the object. The results show that the prototype has a better sensitivity than the acceleration sensor used in the measurements of displacements and accelerations up to a frequency of 3 Hz [58].

SHM measurements using IATS2 (manufactured by Leica Geosystems as a case study series) at the Fatih Sultan Mehmet Bridge in Istanbul show [57, 66, 77]. For this purpose, an LED target with a distance of around 128 m was observed on the object. The bridge was stimulated by normal traffic crossing the bridge. By reducing the field of view, the capturing rate was set to 25 Hz. The results were compared with calculations from a FEM and the determined frequencies can be confirmed. At three frequencies, deviations from previous investigations were found, which could indicate possible structural damage.

In conducting field tests on the Kloštar railway bridge in Croatia, [13] demonstrated the use of the prototype of the Department of Applied Geodesy, Faculty of Geodesy, University of Zagreb. Since no significant natural targets can be observed on the bridge, predefined photo marks were attached to the bridge. At a distance of about 28.5 m,

vertically to the longitudinal axis of the bridge, the target was observed using 60 fps. Train movements at different speeds stimulated the bridge. The comparison of the calculated natural frequencies from an FEM with the frequencies measured by means of IATS shows the successful application.

Measurements of the prototype MoDiTa, also carried out on a railway bridge, show promising results [97]. Natural targets on an arched bar bridge in Lahnstein (Germany) were observed at a distance of 30 to 70 m, stimulated by passenger and cargo traffic. The capturing speed was 500 fps. At the same time, measurements were taken with an accelerometer. Since a vertically oriented view of the structure was not feasible, the measured oscillation was converted to a balanced plane as an approximation of the structure plane. The software offers this solution automatically. The calculated natural frequency could be confirmed by both measurement techniques.

Further investigations with the MoDiTa prototype are concerned with the deformation of bridges that are loaded by large volume or heavy transports. The Wildgraben bridge in Mainz, Germany, was observed for this purpose. It is situated on a federal road in the city area and is approximately 95 m long. The measurement was carried out at night and in cooperation with the police. The bridge was illuminated externally for the measurement. Both previously installed targets and natural targets in the middle of the bridge were monitored. The deformation was caused by the passage of a 40 m long and 90-ton heavy load transport including support vehicles. In *Figure A.4-1*, both the heavy transport and the support vehicles can be seen in the displacement. Also clearly visible are the loads caused by the axles of the transporter. The different strong deformations of the axles of the truck indicate an uneven distribution of the load. Since the observed bridge section is only 35 m long, the entire heavy load transport was never on this section during the crossing. The observed deformation in the vertical direction is significantly smaller than the values calculated from an FEM, which can be attributed to the partial loading of the bridge. As this is a concrete bridge and therefore has a high damping coefficient, a frequency analysis at this point is difficult and as such was not examined further.

One of the main goals of these investigations is to check whether and in which way the bridge is suitable for the planned transport by means of prior measurements with a reference vehicle, e.g., trucks or buses. There is great potential in saving personnel costs for transport escorts, as it can be clarified in advance whether a closure of the opposite lane is necessary. Future research work will continue to focus on this topic.

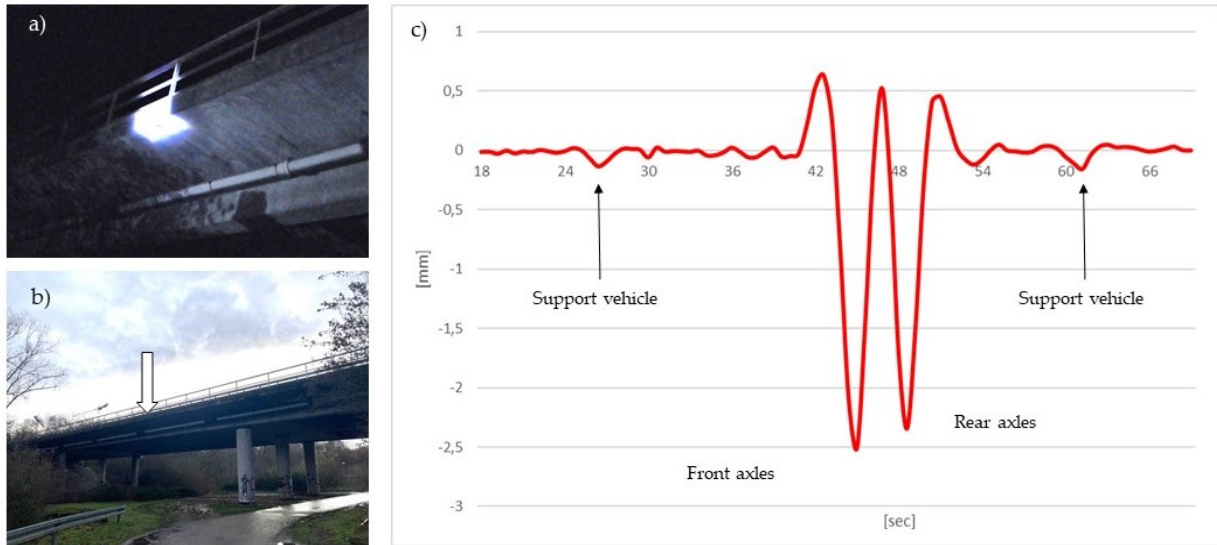


Figure A.4-1 Observed point and measuring results. Illumination a) and position b) of the measured target in the middle of the bridge. c) Representation of the vertical displacement of a target point during a heavy transport on a concrete bridge in Mainz (Germany) using prototype MoDiTa at night [98].

3.A.5 Summary and Conclusion

In this review, an overview of imaging total stations, so-called IATS, was provided. A distinction was made between commercial and research instruments. The image processing algorithms described here show the various approaches that have been implemented to enable further automation. The field tests carried out so far also show the successful use of this measurement technology to obtain helpful information about the structure in the field of SHM for existing structures.

The sensor technology based on image information is convincing due to its contactless detection, its mobility and its short-term usability. The explained set-ups of the prototypes can easily reverse the extension of a total station to an IATS. This means that the total station can be used again for the usual applications. Thus, the field of SHM would also be possible as an area of application for smaller surveying offices. In addition, the achievable accuracy and precision is comparable or even higher than with conventional robotic total stations or GNSS measurements [13]. It should be noted that the georeferenced results are limited by the accuracy of the used total station. In comparison to terrestrial laser scanners, which capture unstructured point clouds, discrete and distinctive points on the object are observed. The great advantage here is the fast, contactless mobile deployment from the situation.

Structures do not need to be equipped with sensors or the necessary power supply. A distinctive point on the existing structure is sufficient. This means that even points that are difficult or impossible to access can be observed. However, the necessary illumination of the target point to be observed is a limiting factor. Night measurements can only be implemented with active target lighting [23]. If the lighting conditions change, the matching must be adjusted depending on the operator used. Various detection and matching algorithms have already been used successfully. However, mainly feature based, raster or area based matching methods are tested. Correlation based matching has been used

most frequently. Nevertheless, other applications have also proven their worth depending on the situation. Previous research in the field of artificial intelligence leads us to expect further applications.

The temporal synchronisation of several IATS makes a contactless 3D determination of target points possible. This requires significantly more hardware. Due to the tendency of long focal lengths, measurements at greater distances from the object are possible, but atmospheric influences must be taken into account. By combining the system with a total station, it is possible to obtain directly georeferenced results. The images are taken in an oriented manner. Furthermore, the stability of the camera can be monitored and corrected by the sensors built into the total station. Commercial IATS are less capable of vibration measurements than the prototypes due to the limited speed at which the images can be captured. However, they can successfully measure them at frequencies below 15 Hz. A higher frame rate and transmission rate of the captured images would be desirable. The ability to transmit the readings wirelessly would add flexibility. Similarly, the commercial IATS offer only limited options for camera settings. Here, the images often show strong noise [93]. By omitting an eyepiece, such as the previously mentioned SX12 from Trimble, more light can reach the camera chip, improving the effectiveness of the following image analysis [65]. It is shown that the focus of image-based observation is primarily on the automation or partial automation of the geodetic monitoring, such as SHM.

Despite the advantages and disadvantages of the measurement technique shown, the results of such SHM can be used by structural engineers to detect damaged areas or to adjust their FEM using actual measured values [78].

Overall, the functionalities provided by IATS are not yet fully utilised in practice. However, the developments on the part of the manufacturers and the research community lead us to expect further application opportunities in the field of SHM.

3.B Self-Calibration and Crosshair Tracking with Modular Digital Imaging Total Station

Abstract

The combination of a geodetic total station with a digital camera opens up the possibilities of digital image analysis of the captured images together with angle measurement. In general, such a combination is called Image Assisted Total Station (IATS). The prototype of an IATS called MoDiTa (Modular Digital Imaging Total Station) developed at i3mainz is designed in such a way that an existing total station or a tachymeter can be extended by an industrial camera in a few simple steps. The ad hoc conversion of the measuring system opens up further areas of application for existing commercial measuring systems, such as high-frequency aiming, autocollimation tasks or tracking of moving targets.

MoDiTa is calibrated directly on site using image processing and adjustment methods. The crosshair plane is captured for each image and provides identical points in the camera image as well as in the reference image. However, since the camera is not precisely coaxially mounted and movement of the camera cannot be ruled out, the camera is continuously observed during the entire measurement. Various image-processing algorithms determine the crosshairs in the image and compare the results to detect movement. In the following we explain the self-calibration and the methods of crosshair detection as well as the necessary matching. We use exemplary results to show to what extent the parameters of self-calibration remain valid even if the distance and thus the focus between instrument and target object changes. Through this, one calibration is applicable for different distances and eliminates the need for repeated, time-consuming calibrations during typical applications.

3.B.1 Introduction

Modular Digital Imaging Total Stations show a wide range of experimental applications in the fields of engineering surveying and metrology [58, 70, 73]. Recent reviews are presented by [13, 99]. In short, the development of surveying instruments into more powerful and user-friendly tools is taking place through automation and the addition of new sensor technology. This can be seen, for example, in automatic target recognition (ATR) or the autofocus of so-called total stations. By extending a total station with one or more cameras, the possibilities of image processing also become available. In addition, a further advantage is the users' independence due to the subjective sense of the observer's eye. However, the cameras used so far by instrument manufacturers primarily serve to improve interactive user workflows, but they currently do not provide a real-time interface for user-specific image analysis or deep learning applications in the measurement process. Furthermore, the highest possible frame rate is significantly lower than the frame rate achievable by industrial cameras. For example, a multistation MS50 from the manufacturer Leica/Hexagon achieves 20 frames per second from a 5 MP coaxial camera and thus allows smooth user interaction. Looking into detail, the framerate of 20 Hz is only achieved with respect to VGA resolution of the display, 640 x 480 pixels [100]. Saving a full 2560 x 1920 pixel image to an SD-card usually takes more than 2 seconds with JPEG compression and even more than 6 seconds in raw format. In order to be able to target applications that we believe require frame rates of around 1 Hz to 1000 Hz, we have

decided to continue the concept of external cameras to achieve these framerates [69], while integrating the motorised focus support of the multistations. We consider it an advantage that the external camera does not disturb the thermal design of the multistation even at high pixel clock rates.

During the process of prototype development, different types of construction emerged. External implementations make it possible to mount the camera on the ocular or to replace it. These are used in combination with commercial total stations or tacheometers and can be converted and adapted to the particular conditions and requirements. One example of such a modular system is DAEDALUS of ETH Zurich [14, 58, 67, 68]. In this concept, a CCD chip replaces the eyepiece. The camera does not capture the crosshairs in the image. No additional optical component is added in between. This makes it necessary to attach a meniscus lens to the front of the telescope for distances of 13 m or more. This front lens shifts the focal plane to the image sensor for focused images. Similar to [101], the calibration is carried out by virtual control points, where the central projection is expressed by an affine approach. Instrument errors are not taken into account.

The developed application of the University of Zagreb attached a GoPro5 directly to the ocular [13, 20]. Another setup where the camera is directly attached to the eyepiece can also be found in [24]. For the measurement with the IATS at the University of Zagreb, videos are recorded which are later split into images. Here, photo targets with predefined circles of known diameter and distance between the circle centres are used. This enables the evaluation of the image data. For the frequency analysis raw image coordinates are used and no camera calibration is required. However, the photo target must be attached to the object to be observed.

The second design offers the advantage of a fixed camera with the instrument like commercial IATS. Commercial IATS often have too low speed of image acquisition for kinematic measurements (e.g. for frequency analysis in Structural Health Monitoring). The fixed camera provides constant calibration parameters as opposed to the modular version which requires calibration after reconfiguration. An early prototype is mentioned in [71] and the prototype series IATS2 from the manufacturer Leica Geosystems in [16, 57, 66, 72]. [71] describes the camera with an affine chip model and uses a combined approach to take camera and instrument errors into account. [74] shows a comparison of different methods.

The measuring system MoDiTa developed at i3mainz extends an existing instrument modularly by an external industrial camera. The self-calibration based on the photogrammetric camera model fully integrates the external camera into the measurement process. By permanently tracking the crosshair, the accuracy characteristics of the total station are maintained. In the following we explain the measurement system and the calibration. The necessary image-based acquisition of the crosshair for the calibration and the further measurement process will be discussed in the following in more detail. The approach used here shows how a calibration can be calculated flexibly on site using software and various cameras and total stations (Compatible to TCA, TPS, TS and MS series from the manufacturer Leica) without any additional equipment.

3.B.2 Measurement System

The Modular Digital Imaging Total Station (MoDiTa) combines a highend industrial camera with a digital total station in a modular and flexible way and is currently on prototype level (*Figure B.2-1*). As described in [24], the standard eyepiece of the total station is replaced by an industrial camera via a bayonet ring. To balance the weight of the camera, we attached a counterweight to the telescope. The cameras can be mounted in any rotation around the target axis by means of a simple clamping screw. By means of a corresponding adapter, the eyepiece camera used takes images directly from the crosshair plane. The crosshair is thus captured in every image. Among other things, this enables automatic, image-based targeting, which is within the accuracies of the total station (standard deviation according to [102]). With the help of template matching, non-signalled distinctive features are captured without contact. The use of the total station's motorised autofocus is advantageous because, among other things, it enables simple self-calibration. After self-calibration, we calculate the corresponding horizontal or vertical angle for each point of interest in the image.

Due to the modular design of the measuring system, a camera can be selected depending on the respective project requirements. Project requirements might include:

- a monochrome, NIR or RGB (Bayer pattern) sensor,
- low light suitability (usually by large pixel pitch) or high resolution,
- a global or rolling shutter,
- availability of a hardware trigger,
- availability of line scan modes,
- high frame rate (frames per second).

The industry standard C-mount used makes it easy to replace components. Depending on the industrial camera used, images can be captured in different modes. By selecting an Area of Interest (Aoi), the range of captured lines and columns can be defined. In line wise mode, only one line is captured over the width of the image. The data to be transmitted can thus be reduced, enabling a higher image capture frequency. A more detailed description can be found in [24].

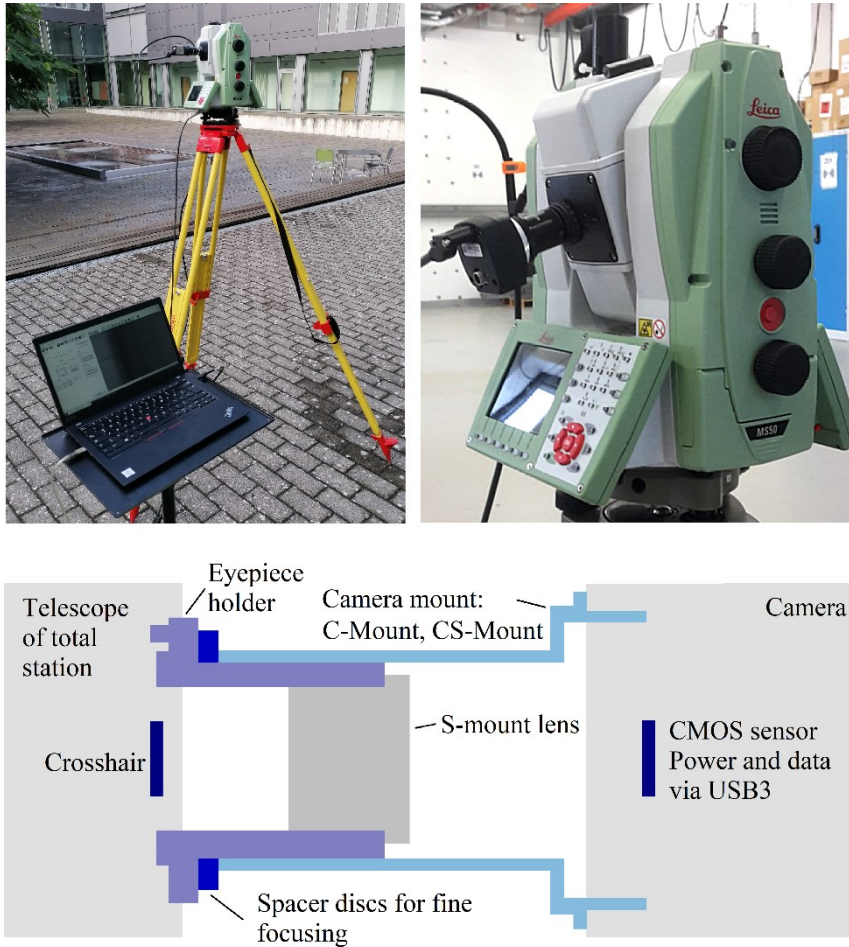


Figure B.2-1 MoDiTa. The upper pictures show the ready-to-measure system MoDiTa in combination with a multistation MS50. The picture below shows the schematic structure of the eyepiece adapter for attaching the digital camera with the optics. The optics are attached to the eyepiece holder via an S-mount connection. This holder is connected to the total station via a bayonet connection for the eyepiece. The length of the eyepiece holder determines the magnification and thus how much of the crosshair is imaged onto the sensor. The digital camera is attached to the camera mount via a C-mount or CS-mount connection.

3.B.3 Self-Calibration

In order to obtain measurement results within the measurement accuracy of the total station, calibration of the entire system is required. This is done by self-calibration directly on site. Given the speed of the self-calibration process, we do not intend to achieve repeatability of the calibration parameters of the camera in different set-ups. The aim is rather to be able to use the measuring system quickly and in an application-oriented manner. The determination of interpolable parameters for a particular combination of camera and total station was never attempted.

The user installs or replaces the camera on site and the measurement can be continued after calibration. Due to the simple mounting of the camera and the modular design, it is near impossible to recreate an identical setup. As a result, there are minimal differences in the optical path for each setup. Differences of several pixels in the image are possible.

Calibration is mainly carried out automatically and only needs to be operated manually by the user at the beginning. Before calibration it is necessary to detect the crosshair to provide a reference image of the crosshair. The crosshair reference image ensures consistency of visual aiming through the eyepiece to camera-based aiming. Furthermore, the reference image of the crosshairs is used to correct any camera movements computationally, cf. *Section 3.B.4*. The telescope is moved relative to a fixed target point in such a way that the target point is imaged at favourably distributed locations on the image plane [24]. This allows for the collection of data for an overdetermined linear system of equations. The software provides for different patterns with different distributions of the observation points in the image. The selection of patterns makes it possible to open up new fields of application in an applied, scientific environment by means of an inexpensive measuring system. For example, a comprehensive high-precision calibration with up to 36 measurements can be carried out for an investigation into atmospheric refraction. The implemented maximum number of observation points in the image is set to 9 points per quadrant (4 quadrants x 9 observation points = 36). It is possible to define fewer observation points, thus reducing the over determination. For a simpler example, see *Figure B.5-2 a*). After the measurement, we have 12 images, each with the target at different positions in the image. In this case we use a black and white laserscanning target with a checkerboard pattern. For the automatic, rough approach of these target directions, the knowledge of a rough start transformation is sufficient, which only includes the camera constant and the rotation of the camera coordinate system around the target axis of the total station. We merely tilt the telescope to the side by a small, fixed amount to determine the start transformation. During the measurement the software continuously observes the crosshair; the so-called matching. Due to the simple mounting of the camera, the crosshair is not in the centre of the image. It is also possible to rotate it around the optical axis.

In the context of self-calibration, we calculate the parameters via parameter estimation based on the least squares method. The functional model is based on the mapping relations between sensor space and object space [71].

Optical distortion and a possible tilt of the camera are compensated by the distortion approach according to [11]. With c as the camera constant for the entire optics, the unknown angles \tilde{H} and \tilde{V} for the target. κ , c and the photogrammetric radial, tangential and asymmetric distortions ($A_1, A_2, A_3, B_1, B_2, C_1, C_2$) are obtained. We describe the illustration of the camera chip by a 2D transformation using a photogrammetric distortion model. $\Delta x'$ and $\Delta y'$ represent the parameters of the distortion, x'_0 and y'_0 represent the principal point respectively the detected crosshair. The angle readings to the fixed target are not measured directly but result from the pixel coordinates of the reference crosshair. The index P represents the measured values to the target point.

The unit vector to the searched target point is formed from the total station readings and the pixel position of the image point of one measurement:

$$\begin{bmatrix} x' \\ y' \\ c \end{bmatrix} = \begin{bmatrix} x'_P - x'_0 - \Delta x' \\ y'_P - y'_0 - \Delta y' \\ c \end{bmatrix} \quad (3.28)$$

The corresponding point in object space is

$$\begin{bmatrix} \bar{X}_P \\ \bar{Y}_P \\ \bar{Z}_P \end{bmatrix} = R_{IP_P} R_{IA_P} R_{IP_P}^T R_{H_P} R_{V_P} R_K \frac{1}{|[\dots]|} \begin{bmatrix} x' \\ y' \\ c \end{bmatrix} \quad (3.29)$$

with

$$|[\dots]| = \sqrt{x'^2 + y'^2 + c^2} \quad (3.30)$$

The total vector in image space is normalised to unity, which is indicated by the division by $|[\dots]|$. The spatial distance D is actually not required for the calibration. D prolongs the unit vector to the object point.

The matrix R_K describes the rotation of the camera sensor around the optical axis.

$$R_K = \begin{bmatrix} \cos \kappa & -\sin \kappa & 0 \\ \sin \kappa & \cos \kappa & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (3.31)$$

R_{H_P} and R_{V_P} follow from the graduated circle reading. The required values are supplied by the total station. The matrices describe the necessary rotations to transform the direction vector from the system of the total station into a coordinate system of its ancestries. Thus, when the instrument is previously stationed, the coordinates are converted into the used system directly.

$$R_{H_P} = \begin{bmatrix} \cos H_P & -\sin H_P & 0 \\ \sin H_P & \cos H_P & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (3.32)$$

$$R_{V_P} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \sin V_P & \cos V_P \\ 0 & \cos V_P & -\sin V_P \end{bmatrix} \quad (3.33)$$

R_{IP_P} and R_{IA_P} describe the rotation of non-compensator corrected total station readings into compensator corrected ones (I_P and I_A are calculated from the inclination in the direction of the target axis and transversely to the direction of the target axis). \bar{X}_P , \bar{Y}_P and \bar{Z}_P are calculated and denotes the normalised direction vector to the target point.

By using (3.34) and (3.35) the angles \bar{H}_P and \bar{V}_P can be calculated.

$$\bar{H}_P = \text{atan} \left(\frac{\bar{X}_P}{\bar{Y}_P} \right) \quad (3.34)$$

$$\bar{V}_P = \text{atan} \left(\frac{\bar{X}_P / \sin \bar{H}_P}{\bar{Z}_P} \right) \text{ for } |\sin \bar{H}_P| > |\cos \bar{H}_P|$$

else

$$\bar{V}_P = \text{atan} \left(\frac{\bar{Y}_P / \cos \bar{H}_P}{\bar{Z}_P} \right) \quad (3.35)$$

The different telescope positions H_P , V_P result in the corresponding image point x'_P , y'_P . From equation (3.29) follows \bar{H}_P , \bar{V}_P to the target point. We use this concept for self-

calibration. This means that the target point does not have to be aimed directly but is introduced as an unknown \tilde{H} and \tilde{V} .

$$\bar{H}_P - \tilde{H} = 0 + v_{\bar{H}_P} \quad (3.36)$$

$$\bar{V}_P - \tilde{V} = 0 + v_{\bar{V}_P} \quad (3.37)$$

We calculate residuals through a summary modelling of all stochastic influences. Due to practical reasons, the stochastic portions of image coordinates, circle readings and compensator readings are not modelled separately from each other. Atmospheric flicker can be reduced by grouped multiple exposures suitably.

Here, the corrected tachymeter readings to the target point correspond to the direct measurement to the target. We did not differentiate between total station and camera-related corrections.

x'_{ch} and y'_{ch} represent the pixel position of the crosshair. Distortion and κ have no effect on the principal point.

$$R_{\tilde{H}} R_{\tilde{V}} R_K \frac{1}{|[\dots]|} \begin{bmatrix} x'_{ch} - x'_0 - 0 \\ y'_{ch} - y'_0 - 0 \\ c \end{bmatrix} = R_{\tilde{H}} R_{\tilde{V}} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \quad (3.38)$$

We consider measurements independently and equally accurate. The weight matrix $P=I$ is defined with ones on the main diagonal or zeros if the measurement is not to be included in the equation as an error.

The Cholesky factorisation according to [61] is used to solve the system of normal equations. The normal equation matrix is split into an upper and lower triangular matrix C and CT . We solve the system of normal equations by subsequent forward and backward substitution. This saves computing time because instead of the entire normal equation matrix N only one triangular matrix needs to be inverted [11, 61].

From the linear dependent residuals of the unknowns follows the estimation of the unknowns as

$$\hat{x} = (A^T P l)^{-1} A^T P l \quad (3.39)$$

The unknowns and the termination criterion are calculated per iteration. Termination occurs after the limit has been reached.

$$x^T A^T P l = l^T P l - v^T P v < 0.00000001 \quad (3.40)$$

As a result, the compensated direction angles to the target are provided. A transformation into Cartesian coordinates can be done afterwards by a distance measurement, if required. For this purpose, the determined target point is directly entered by the total station and a reflectorless measurement is carried out. The measured distance is used to extend the unit vector to the target point.

3.B.4 Crosshair tracking

Based on the modular adapter for mounting a camera, it is possible to capture the crosshair. The crosshair is a geodetic crosshair that is not located in the exact centre of the image. We distinguish between detecting and matching. Detection of the reference crosshair should take place as soon as possible after the camera is mounted. As with manual eyepiece adjustment, a monotone image background is preferred for this step, e.g. a sky or grossly out-of-focus image, so as to get an even background. The position and orientation of the crosshair in the pixel coordinate system of the camera is determined with the help of further crosses, which we refer to as (virtual) *réseau* crosses in the following. The software continuously observes the position of the crosshair during further measurement. Any image coordinate is transformed to the reference crosshair using 2D transformation including two translations and one rotation. Smaller deviations are recognised and taken into account by matching. If the change is too large, a new detection is necessary. *Figure B.4-2* provides a simplified overview of the single steps. In the following, we distinguish between the inner and the outer crosshair. The two inner lines mark the centre of the crosshair. The outer crosshair is composed of the six outer lines of the geodetic crosshair (*Figure B.4-1*). The elaborate modelling of the reference crosshair makes it possible to ensure a largely continuous tracking later on, even if line elements are only recognisable in parts.

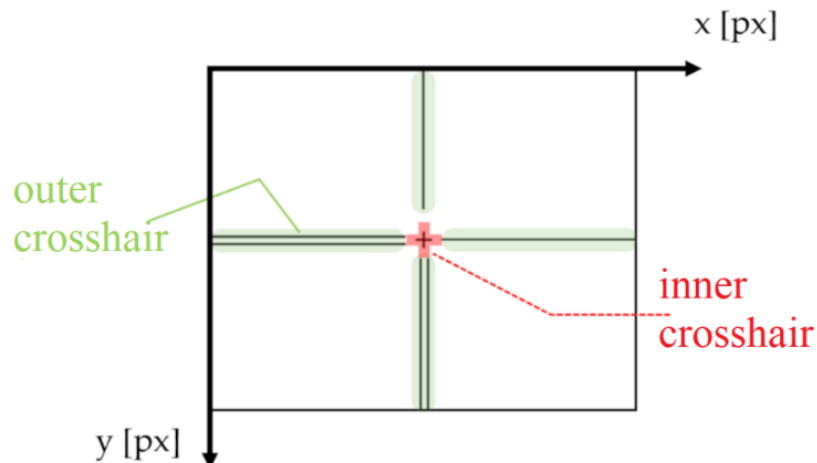


Figure B.4-1 Geodetic crosshair with pixel coordinate system. Also shown is the distinction between inner (red) and outer (green) crosshair.

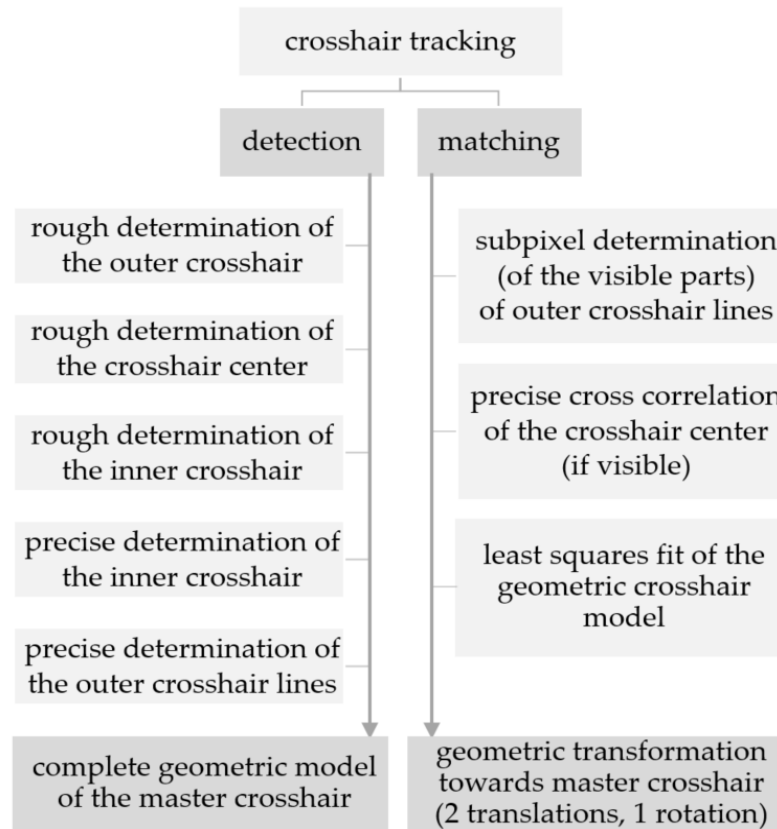


Figure B.4-2 provides a simplified overview of the single steps. In the following, we distinguish between the inner and the outer crosshair. The two inner lines mark the centre of the crosshair. The outer crosshair is composed of the six outer lines of the geodetic crosshair (*Figure B.4-1*).

3.B.4.1 Crosshair detection

During crosshair detection, first the outer crosshair with its six lines will be roughly determined. According to [103 – 105], a Gaussian smoothing filter in combination with its partial directional derivatives is applied. According to [106], the smoothing and the threshold value are determined. If the value of the second partial derivative of an image point exceeds the upper threshold in a pixel, this is detected as a line point with sub-pixel accuracy. If the second partial derivative is smaller than the lower threshold, we dismiss the pixel. If the value is between both thresholds, the pixel is only used if it can be connected by detected line points. As a result, we obtain several line segments. These are then examined for false detections according to [107] or [108]. By calculating a regression line through the image points of the lines, the mean distance of the individual points to the line can be calculated. We reject points with greater distance than the mean. The regression line is then calculated again. By defining a limit value for the direction difference and the distance of the end points of neighbouring lines, these are merged if necessary (*Figure B.4-3*). The regression line is then calculated again. This is repeated until the maximum values for the direction difference and the distance of the line end points are no longer undercut. The longest six lines of the calculations correspond to the outer geodetic crosshair. These are still in the form of polylines. They are calculated individually by adjustment according to [107, 108] as a straight line equation. In addition, the line width is determined for later use in the precise determination of the crosshair.

Simplified, we calculate the width of the longest line by edge detection. Starting from a point on the line, we search the perpendicular distance on both sides up to the edge. The length of the perpendicular is determined for each pixel on the vector line. The mean value then gives the line width. The contour width can differ depending on the camera and the total stations crosshairs, but it must be at least one pixel wide in order to be detectable.

The results are the start and end points, the straight line equation of the six crosshair lines and the line width.

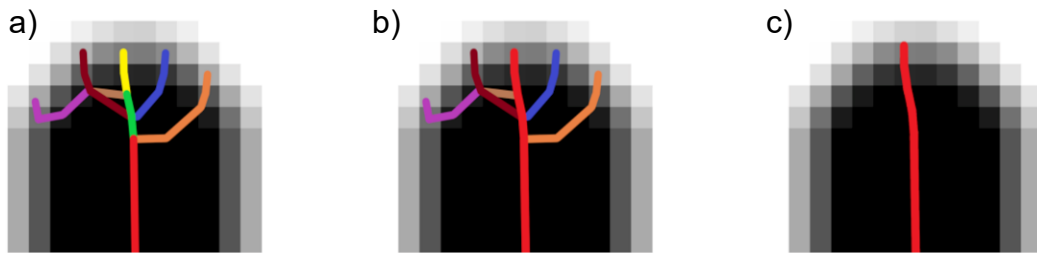


Figure B.4-3 Detection of the outer crosshair according to [103 – 106, 108]. Solving the ambiguities using an example of an end of a line. a) Line detection. b) Merging lines based on limits. c) Discarding falsely detected lines

We use the six compensated lines of the outer crosshair to determine the rough crosshair centre. All intersections of the straight lines are formed. We eliminate negative coordinates. By forming the median of the nine intersections, the rough centre is calculated. The maximum distance from the rough centre to the intersection points and the approach of an isosceles triangle are used to calculate the distance between two parallel crosshair lines.

For the definition of the inner crosshair, we define a circle with the centre equal to the roughly determined centre and the radius equal to the distance between two parallel crosshair lines. According to [11], one-dimensional grey value profiles are formed vertically to the circumference.

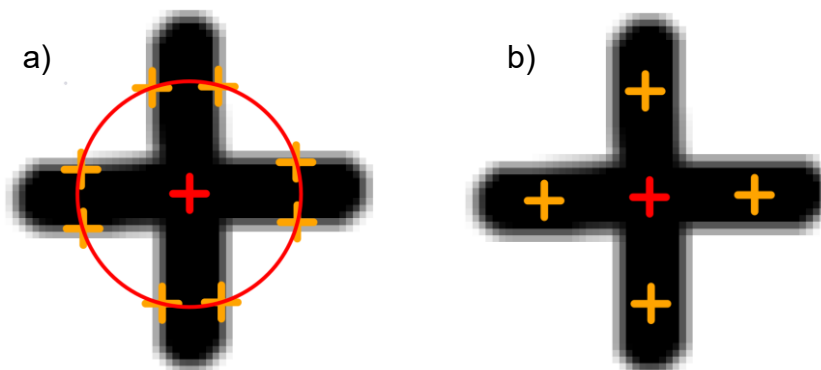


Figure B.4-4 Rough detection of the inner crosshair. a) Definition of an intersection circle around the rough centre. Edge detection following the circle. b) Mean of the edge points.

These are obtained by averaging all existing grey values of a line that lies vertically to the circular ring. Using a Gaussian smoothing filter and a Laplace operator, we are able to calculate the edge positions with sub-pixel accuracy. We compare the edge amplitudes with a previously defined threshold value. If the amplitude is greater than the threshold value, an edge is present at the corresponding image position. A total of eight points are

detected on the inner crosshair, two points per line (*Figure B.4-4 a*). If more than four lines intersect with the circular ring, only the best four edges are used. The selection is made via the calculated edge amplitude. The greater the amplitude, the higher the contrast of the contour at the image position. The start and end points of the edges per line are calculated as a mean so that there is one point for each inner line of the crosshair (*Figure B.4-4 b*). For the orientation of the cross, the direction angle from the rough centre to the point with the largest edge amplitude is used.

For the precise determination of the crosshair, we consider the lines individually and points are determined at equal distances on each line. With the help of the direction angles, we form sectors of circles around the rough crosshairs. Within two defined circles with different radii, the edges per pixel of the lines are detected again according to [11] (*Figure B.4-5 a*). As a result, the edge beginnings and ends of the inner crosshair are available for each line as coordinates between the two circles. These are averaged again so that the centre points of the line contours are available. Points with the same direction on the inner crosshair are combined. This results in two lines: one horizontal and one vertical. These are equalised according to the same principle of [107, 108] (*Figure B.4-5 c*). The exact image coordinates of the crosshair centre are now available with subpixel accuracy via the point of intersection.

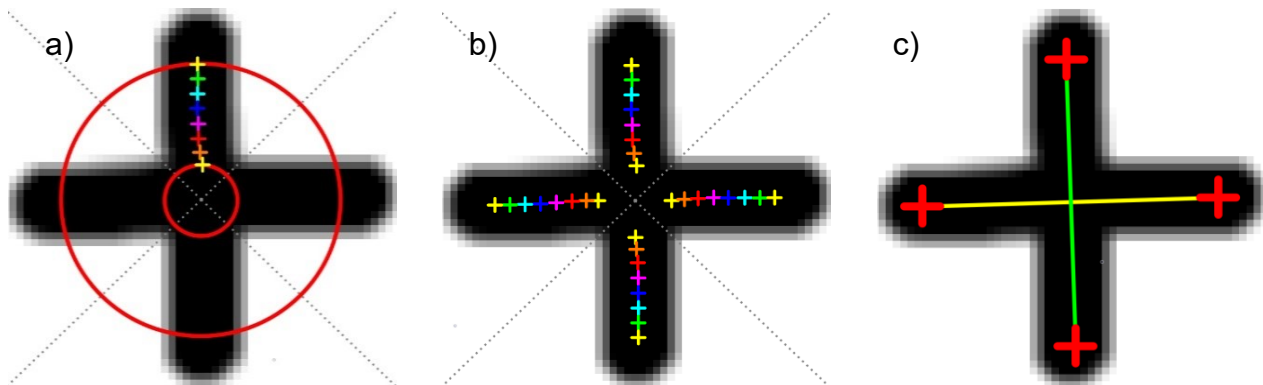


Figure B.4-5 Precise detection of the inner crosshair. a) Definition of circles around the rough centre. Creating sectors. b) Mean of the edge points. c) Compensated lines with end points.

We calculate *réseau* crosses for the alignment of the precise crosshair. This is done by calculating two points per line. By defining two circles with different radii (small circle: 3 times the distance of the parallel crosshair lines, large circle: shortest distance from the centre to the edge of the image reduced by 10%). The resulting intersections with the outer cross lines are all within the image area.

The smaller circle is intersected with the precise inner crosshair lines (*Figure B.4-6 a*). Similar to the procedure already described for the rough determination of the inner crosshair, the edge contours are determined perpendicular to the line from the crosshair centre according to [11] (*Figure B.4-6 b*). The detection of the contours is again carried out via grey value profiles and the start and end points are then averaged. We repeat the steps for the outer radius so that two points are available for each outer crosshair line. Finally, we sort the detected points separately for the inner and outer circle in the correct quadrant. The twelve points on the outer cross lines are determined precisely and are available in the correct position as a pair of points per line. One hundred further points are

then determined between two points of a line according to the procedure shown in *Figure B.4-5*. We defined the number for which a large number of points is available for the definition of the straight line and the following adjustment. A different definition is also possible. The one hundred points to be detected are distributed at equal intervals along the length between the start and end point of a line. As in the previous step, the detection for each point of the line contour perpendicular to the line is performed. The points on the opposite crosshair lines are combined so that the final result is a horizontal and a vertical line (*Figure B.4-7*). The adjustment calculation is carried out according to the least squares method. We detect and eliminate outliers before the adjustment.

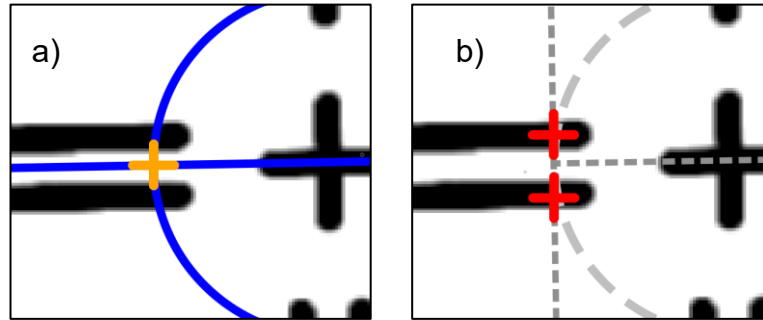


Figure B.4-6 Point detection of the outer crosshair using the example of a double line. a) Intersection of the inner circle with an inner line. b) Line points are formed via edge detection perpendicular to the intersection point.

The calculation of the straight lines is based on the procedure according to [109], method 3. We transferred the model of the adjustment calculation mentioned to the model of the crosshair with two straight lines and adapted to its special features. The double lines of the crosshair result in an additional unknown so that the functional model in coordinate form is as follows:

$$a_{1,2} * x_i + b_{1,2} * y_i + (d_{1,2} \pm \delta_{1,2}) = 0 \quad (3.41)$$

The variable δ corresponds to the half parallel distance of a double line to the adjusted straight line. These are chosen with different signs for the double lines and should be defined the same for both straight lines. The parameter δ is omitted for the single line. The equation is set up independently for the two lines, so that the parameters must be determined separately for each equation. The eight parameters of the two lines to be determined are listed in the unknown vector X .

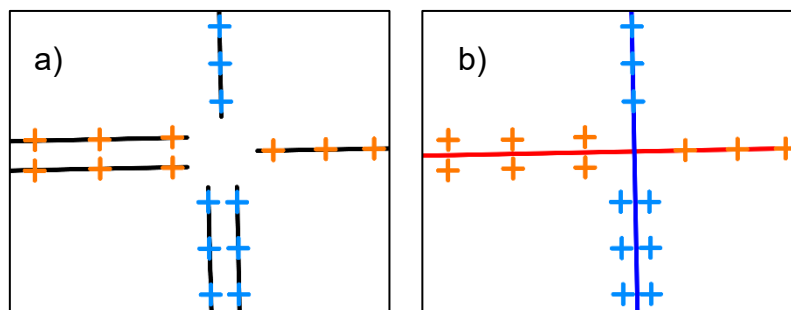


Figure B.4-7 Principle of adjustment (simplified example). a) Before the adjustment. b) After the adjustment.

The adjustment of both straight lines is done in one calculation. The approximate values for the variables $d_{1/2}$ and $\delta_{1,2}$ are determined empirically. b_1 and a_1 are given the value zero and are thus defined as parallel straight lines to the image coordinate system. Since the

underlying equation of the functional model is linear, the partial derivatives according to the parameters are equal to the observations used. Simplified, the condition equation can be regarded as an observation equation for the software solution, but with a significantly higher weight than the observations. The weight for all observations equals 1, and the conditional equation is given the weight 10^6 [109]. As previously described in the section on calibration, the solution of the system of normal equations is carried out by means of Cholesky factorisation according to [11] or [61]. The adjustment is iterative until the termination criterion is reached. For the second iteration the parameter estimates are updated to the extent that the unknown X are defined as parameter estimates X_0 for following iterations. For the determination of the two straight lines of the outer crosshair in the subpixel area with one decimal place, a few iterations are already sufficient. We choose the termination criterion in such a way that in normal cases only a few iterations are necessary. The three opposite lines with approximately the same orientation are balanced to form a straight line (*Figure B.4-7*). The outer line cross has thus been determined precisely so that the *réseau* crosses can then be determined as described. These are located on the straight lines determined at this point and, together with the crosshair centre, define the precise position and orientation of the crosshair in the image coordinate system.

3.B.4.2 Crosshair matching

Due to the possibility that the crosshair position changes after a longer period of time and when the telescope position changes, this must be determined continuously [70]. In the case of smaller movements of the crosshair, this can be matched by calculating a normalised correlation coefficient (NCC). The current centre of the crosshair in the image is compared with the last detected crosshair. The current position of the crosshair centre must be within a generated model in order to be matched. If the difference in position is too large, the crosshairs must be detected again. Matching is again carried out using the NCC procedure according to [11]. This procedure corresponds to a simplified detection of the precise crosshair centre since no homogeneous background is required throughout. The current crosshair centre should be able to be continuously tracked in the image during a measurement.

For the model image, we define a circle with 0.75 times the distance between the parallel lines around the precise centre of the crosshair. Within this image area, the similarity comparison is carried out by means of normalised cross-correlation. The model image, in this case the pixels of the circle area, is generated on several image pyramids in different planes and rotations. We generate the image pyramids until the top level still provides enough information about the image. The processing effort is higher than with other correlation methods due to the large number of images generated. However, for the selected image area and due to today's technology standards, this is not a disadvantage [11]. Finally, the origin of the model image is set to the precise crosshair centre.

For the crosshair matching the NCC model of the last detected crosshair centre together with the current crosshair is required. We define the search area for the inner crosshair by a circle around the last detected crosshair. We define the radius with 1.5 times of the distance between the parallel lines. This saves unnecessary computing time since the centre of the crosshair cannot be located at the edge of the image. Then, according to [11], all defined instances of the crosshair are detected from within the image section of

the current crosshair. We only use the best instance for the crosshair since it is unique in the image. The best instance is characterised as the highest value of the correlation coefficient, whereby this can take values between zero and one. The calculation of the precise crosshairs together with the resulting *réseau* crosses is then carried out according to the procedure already described. The software saves the coordinates again in a local file. However, the inner crosshair can be determined mathematically by detecting the outer crosshair (*Figure B.4-8*). By determining the crosshair lines and the geometric reference to the target axis in subpixel accuracy, the intersection of the crosshair lines does not have to be determined repeatedly in every image, but, for example, may also be overlapped. In the user interface of the control software, coordinate differences between the last detected and the matched crosshair centre are displayed to the user.

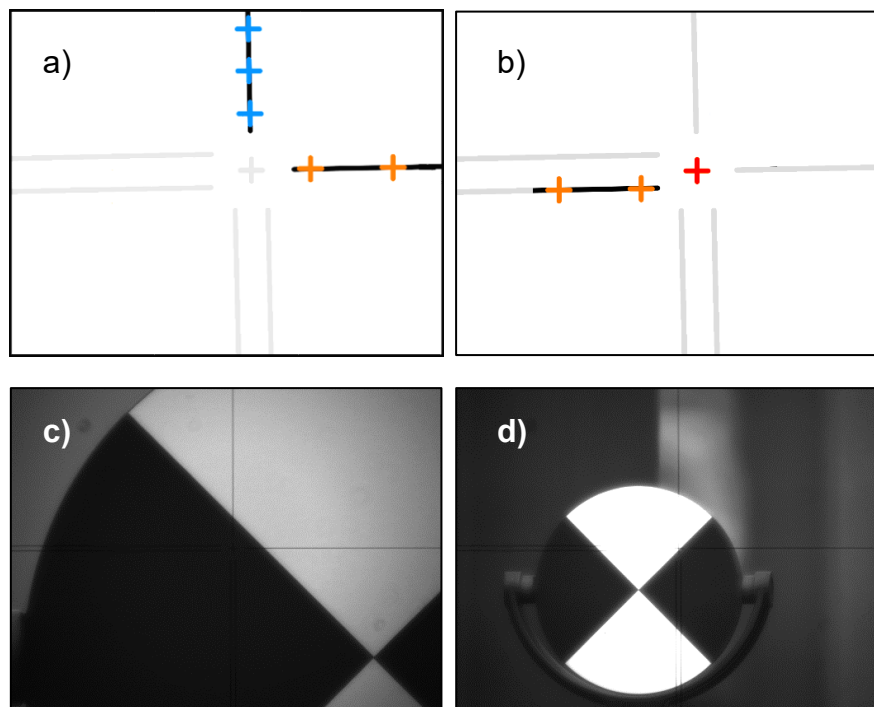


Figure B.4-8 Crosshair matching. a) and b) Goal of crosshair matching (simplified example). Matched points on at least two lines in two directions a) or on one line in combination with a successful matching of the inner crosshair ensure tracking of camera motions. c) and d) examples of an unmatchable inner crosshair due to dark background c) and overexposure d). Nevertheless, the inner crosshair is calculable by the detection of the outer crosshair.

3.B.5 Practical Applications

In the following, we cover exemplary studies on different applications of MoDiTa in the Structural Health Monitoring (SHM) of existing structures.

3.B.5.1 Studies on Distance Independence

A practical application for IATS is the SHM of structures, such as factory chimneys, dams or bridges [13, 99]. What all these structures have in common is that they usually have elongated dimensions. Often it is impossible to stand directly perpendicular to the structure or to measure the entire structure from the same distance due to environmental

conditions, such as rivers or railway lines (*Figure B.5-1*). Changes in the distance to the measured object lead to the refocusing of the optics and thus also to changes in the distortions. In order to discuss this aspect in more detail we have carried out measurements from different distances to the instrument. For the measurement, we used a TS30 [110] and an industrial camera UI-3250 ML-M [111] with 1.92 MPixel.

The measurement took place on 3 February 2022 in the courtyard of Mainz University of Applied Sciences between 10 am and 1 pm (CET).

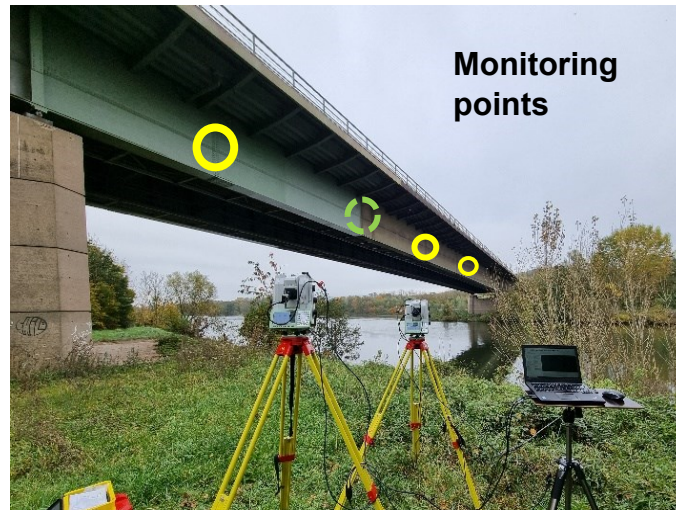


Figure B.5-1 Exemplary view of two IATS (MoDiTa) on an elongated structure with different forced distances. Here, the observation of a reference point (green dashed line) and simultaneous observation of other monitoring points on the bridge (yellow) are shown.

We limited the distance between 3.5 m and 100 m. Over this distance calibration measurements were carried out with MoDiTa with a sample of 12 measurements for one distance. These measurements were taken over the entire image area. For comparison, we applied the calibration of the measurement with 20 m distance also to the measurements with shorter or longer distance. We compared the results in the form of residuals or deviations in *Figure B.5-2*. The different positions that IATS moves to are visible.

Figure B.5-2 b shows residuals of the 20 m measurement calculated with the 20 m calibration in the range of -0.3 to 0.2 mgon for the horizontal and zenith angle, which is within the expected angular accuracy of the measurement system. We assume that the adjustment modulates the system successfully. In comparison, even at a longer distance to the object, deviations appear only slightly. However, the use of the same calibration shows that at a greater distance significantly larger deviations occur in the marginal area of the measurement image (*Figure B.5-2 c*). Close to the crosshair, in this case also close to the centre of the image, the deviations have values of max. 0.3 mgon; in the outer range of -0.6 to 0.9 mgon horizontally and -0.1 to 0.5 mgon zenith distance. Therefore, the middle part is still modelled within the accuracy of measurement but systematics can already be identified in the outer area. From a distance of approximately 7 m, the deviations increase significantly. For clarification, *Figure B.5-2 e* shows the measurement with 4.5 m distance. The range of the residuals is from -0.7 to 1.1 mgon for the horizontal and zenith angle. Here again, the values are smaller towards the centre but on average

well above 0.3 mgon. The systematic deviations can be traced back to an unsuitable model of the adjustment. [71] and [16] achieved similar results with an integrated coaxial camera and thus correspond to our expectations. We suspect the influence of a scale factor, possibly the optical system's focal length. *Figure B.5-2 d* and *f* show the residuals of the self-calibrations with the corresponding measurements. In both cases, the residuals remain within the measurement accuracy of the total station and do not exceed 0.3 mgon.

Figure B.5-3 shows the estimated distortion parameters over the entire distance. The estimated nine parameters are presented with their respective accuracies as described in chapter 3. The calculation is made for the entire optics, i.e. for the telescope of the total station and the adapter used (*Figure B.2-1*). In the parameter estimation, we calculated a separate calibration for each distance and autofocus setting.

In addition to the individual distances, we gave the respective stepper motor positions since these reflect the respective required mechanical movements of the focusing optics, which are responsible for the changes in the distortion parameters [74].

For the template matching with the cross-correlation method, a circular section with a radius of less than 40 pixels was used in each case. The maximum mean error of unit of weight from the adjustment is 0.32 mgon. The smallest value is reached during measurement at a distance of 20 m, where it is 0.13 mgon. All measured values of the 12 measurements were used, and no measured values were eliminated during the adjustment.

Table B.5-1 shows the maximum and minimum estimated parameter values and empirical standard deviations corresponding to *Figure B.5-3 a-h*.

All symmetric radial distortions show their minima and maxima at close range *Figure B.5-3 a-c*. In total, the calculated empirical standard deviation of the tangential distortions B_1 and B_2 (*Figure B.5-3 d-e*) show similar orders of magnitude and are more consistent than those of the symmetric radial distortions. Both B_1 and B_2 show their greatest value at a long distance but fluctuate more in the near range.

Table B.5-1 Results of the adjustment

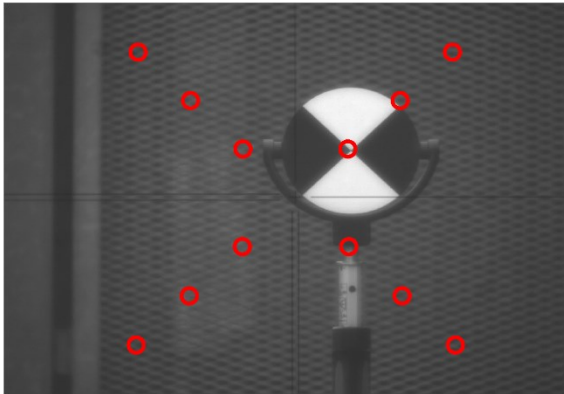
	estimated results		empirical standard deviation	
	min	max	min	max
κ [rad]	2.05827×10^{-5}	1.14836×10^{-3}	1.75569×10^{-4}	4.33541×10^{-4}
c [mm]	456.5227	461.1159	0.3116	1.1808
A_1 [1/mm ²]	-5.99941×10^{-3}	6.70138×10^{-4}	1.55959×10^{-3}	5.80336×10^{-3}
A_2 [1/mm ⁴]	-5.65061×10^{-5}	1.82210×10^{-3}	4.60688×10^{-4}	1.78219×10^{-3}
A_3 [1/mm ⁶]	-1.36531×10^{-4}	-5.42970×10^{-7}	3.48835×10^{-5}	1.36997×10^{-4}
B_1 [1/mm ²]	-2.70265×10^{-4}	2.59471×10^{-2}	8.47366×10^{-3}	2.01747×10^{-2}
B_2 [1/mm ²]	-9.56102×10^{-4}	3.31246×10^{-2}	2.89327×10^{-2}	1.02284×10^{-2}
C_1 [1/mm]	-8.08512×10^{-1}	-9.74693×10^{-2}	5.33810×10^{-2}	1.32803×10^{-1}
C_2 [1/mm]	3.14872×10^{-2}	2.84365×10^{-1}	5.72376×10^{-2}	1.39879×10^{-1}

The results show slightly strong changes of the parameters in the near range. This is in line with our expectations: With larger movements of the focusing optics, we anticipate correspondingly larger changes in the calibration parameters.

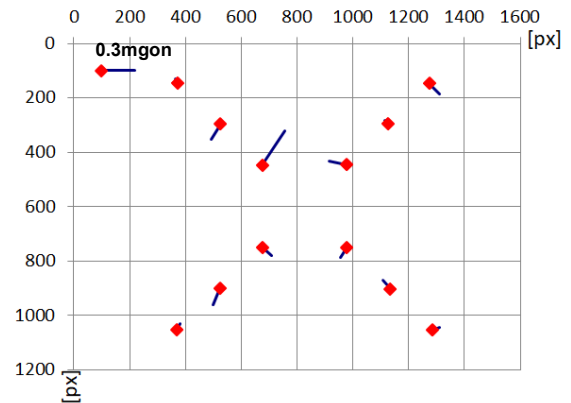
Overall, our studies show that micromovements of 0.5 mm/100 m can be resolved. The achievable accuracy depends on the atmospheric conditions and the illumination situation and decreases with increasing distance.

The investigations indicate that in the case of typical distances to structures for SHM of 10 m or greater, it is sufficient to use a single self-calibration for a complete project.

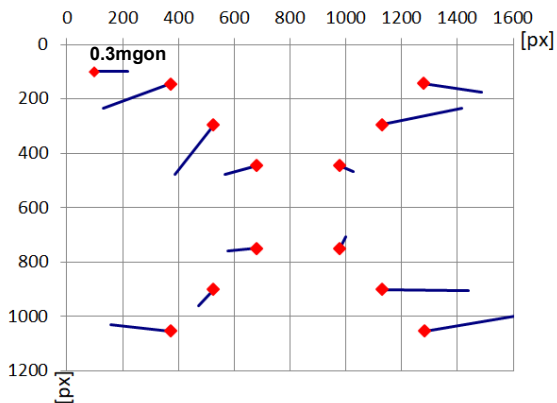
a) Measurement pattern



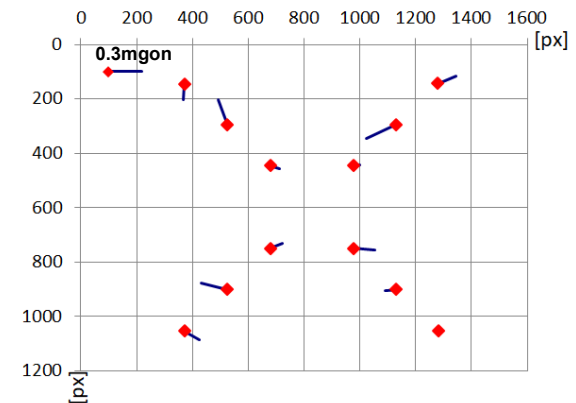
b) Self-calibration at 20 m distance



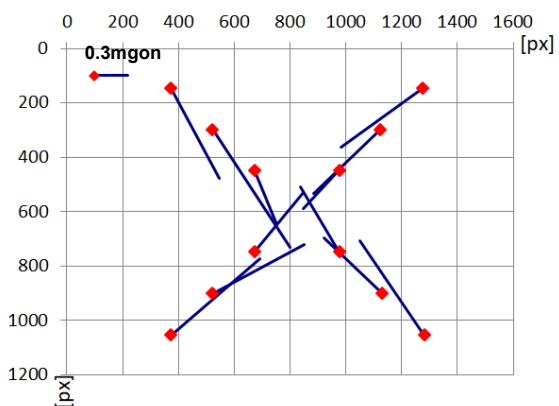
c) 20 m calibration with 70 m measurement



d) Self-calibration at 70 m distance



e) 20 m calibration with 4.5 m measurement



f) Self-calibration at 4.5 m distance

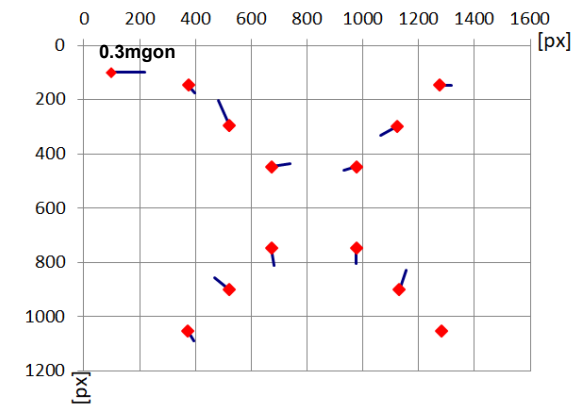
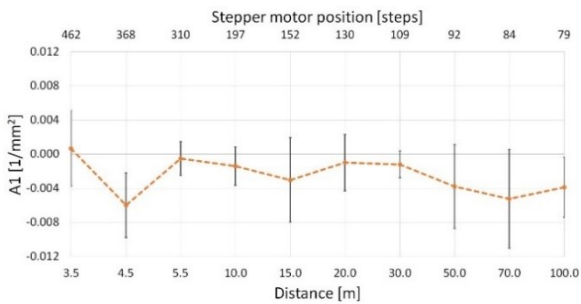
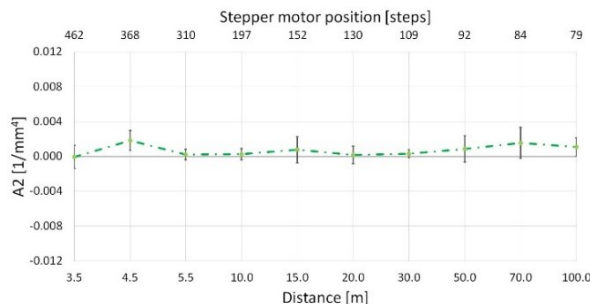


Figure B.5-2 Results of the self-calibration. a) shows a measurement image with displayed positions for self-calibration. The defined target is located at the positions marked in red. In this way, different positions (in this case 12) are approached across the image for adjustment. b) shows the resulting residuals of a measurement at 20 m distance to the instrument. c) shows the resulting deviations of a measurement at 70 m distance calculated with the self-calibration at 20 m distance to the instrument. d) shows the resulting residuals of a measurement at 70 m distance to the instrument e) shows the resulting deviations of a measurement at 4.5 m distance calculated with the self-calibration at 20 m distance to the instrument. Clearly visible are the larger deviations compared to (b) and (c). f) shows the resulting residuals of a measurement at 4.5 m distance.

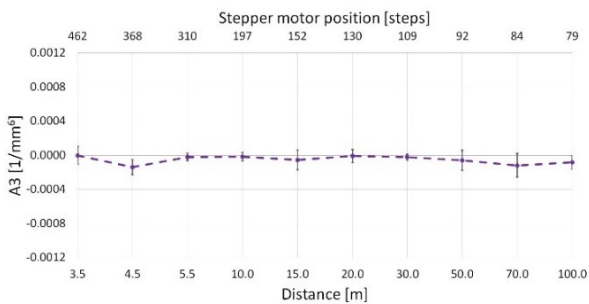
a) Symmetric radial distortion A_1



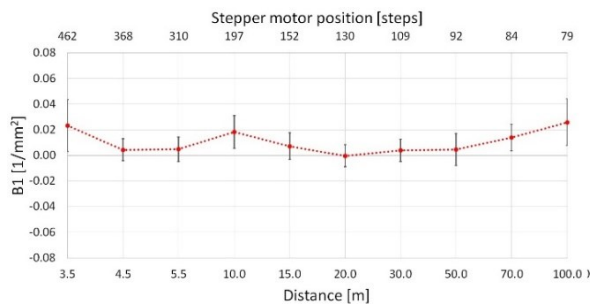
b) Symmetric radial distortion A_2



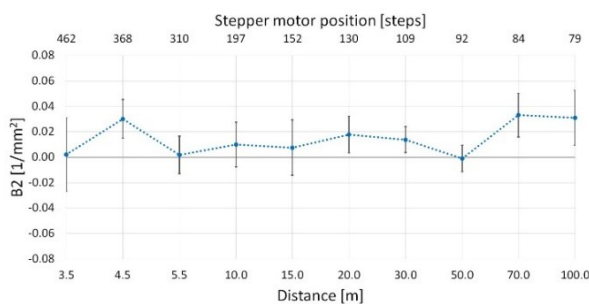
c) Symmetric radial distortion A_3



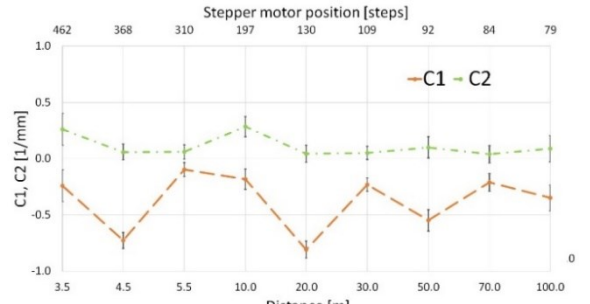
d) Tangential distortion B_1



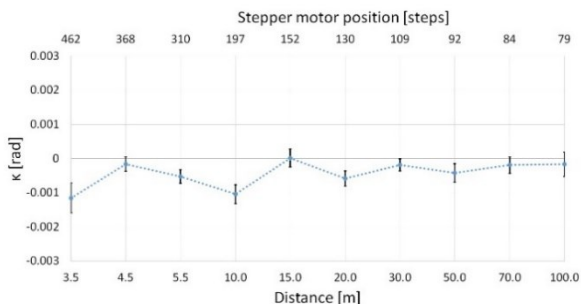
e) Tangential distortion B_2



f) Affinity and shear C_1, C_2



g) Angle κ



h) Camera constant c

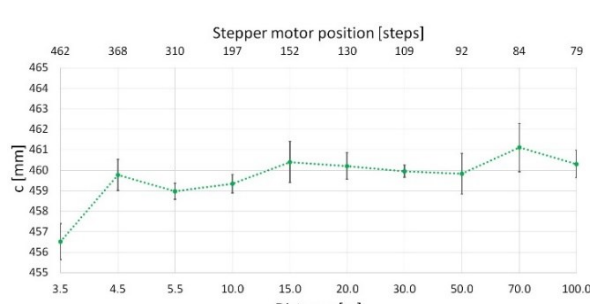


Figure B.5-3 Plot of estimated parameters and a-posteriori/empirical standard deviation.

3.B.5.2 Deformation Measurement on a Steel Bridge

The identification of dynamic structural characteristics is an important aspect of SHM. It enables the analysis of monitored vibrations and the calculation of natural frequencies. Changes in natural frequency indicate possible structural damages. Conventionally, acceleration sensors are attached to the structure with a high labour input. MoDiTa enables high-frequency recordings of the movement behaviour without having to enter the structure. For the measurement of frequencies, the maximum frames per second (fps) are essential. Only at an adequately frequent sampling rate can the natural frequency be determined from the measured values and aliasing ruled out. [112, 113] do not recommend determining the damping from the excitation of a train crossing, because the selection of the ambient window has too great an influence on the result and therefore the results scatter too much.

To capture the deformation behaviour of a bridge during a crossing, we observed a steel bridge using MoDiTa (*Figure B.5-4 a and b*). We used a Leica TS60 [3] in combination with an industrial camera UI3080 CP-M with 5.04 MPixel [114]. The distance to the bridge was approximately 10 m. The recording frequency was 500 Hz with an exposure time of 0.002 s. To achieve this high frequency, we only observed an area of interest. The observed point is located in the upper centre of the western bridge. We attached no targets to the structure. *Figure B.5-4 c* shows the recorded deformation over time. For further analysis, we used the ambient window (shown in *Figure B.5-4 c* by a green box). The ambient window captures the oscillation behaviour of the bridge without additional mass from the train.

Figure B.5-4 d shows the results of the Fast Fourier Transform (FFT) of the ambient window to determine the natural frequencies. Only the results of the low frequencies are shown. A clear peak is visible at 3.9 Hz.

In order to calculate the damped oscillation using the least squares fit, we evaluated the measured values from second 19.0 to 20.0 (*Figure B.5-4 e*). The result is a calculated natural frequency of 3.79 Hz. The mean error of unit of weight is estimated at 0.046 mm/10 m. This corresponds to the expected resolution and accuracy, cf. 3.B.5.1.

Both evaluations come to almost the same result and confirm each other.

a) Railway crossing structure



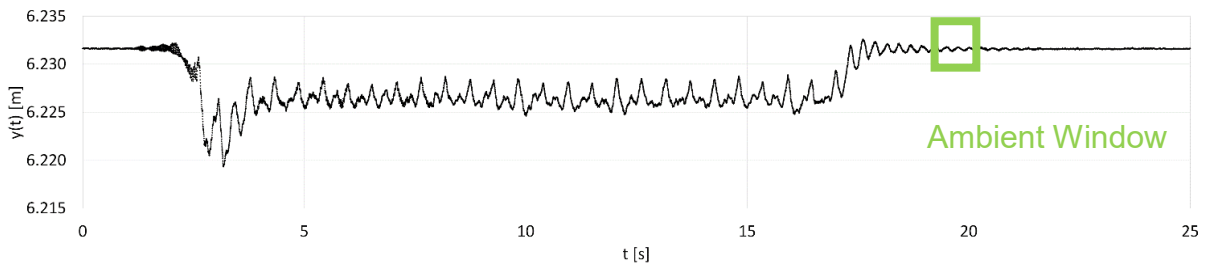
Lat: 49.893556
 Lon: 8.637458
 Date: June 17,
 11:16 CET

b) Measuring system MoDiTa

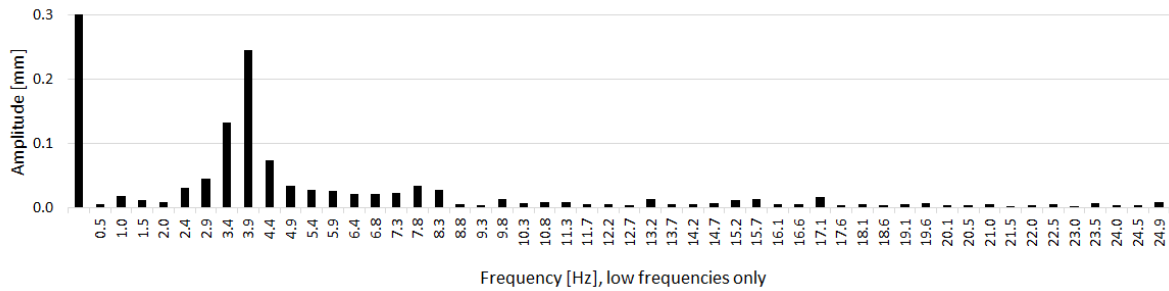


Total Station: Leica TS60
 Camera: UI3080 CP-M with
 5.04 MPixel
 Exposure time: 0.002 s
 Recording frequency:
 500 Hz
 Distance: 10 m

c) Movement of the target in vertical direction



d) FFT, ambient window only



e) Least square fit for damped oscillation, ambient window only

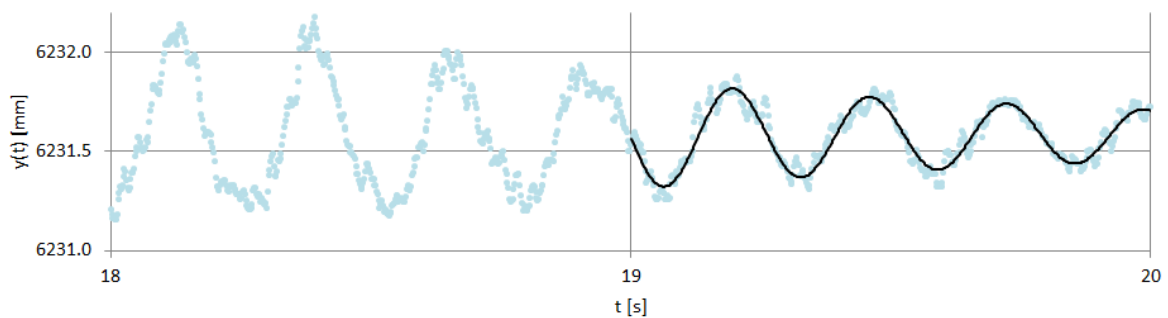


Figure B.5-4 Deformation measurement on a steel bridge. a) and b) shows a steel bridge (30m span) for freight and passenger traffic. c) shows the recorded measured values in vertical direction. 25 seconds of the recording are shown. For this period, 12 500 observations are shown. For the evaluation of the vibration behaviour we evaluated the ambient window. d) shows the result of the Fast Fourier Transform (FFT) with a main peak at 3.9 Hz (approximate value) e) shows a best-fit evaluation of the ambient window with a natural frequency at 3.8 Hz (balanced value).

3.B.6 Conclusion and Outlook

In this article we explain how the general setup of the measuring system in combination with the developed software detects the crosshair and thus performs self-calibration. The self-calibration achieves accuracies within the measuring accuracy of the total station. We made assumptions such as the lines of the crosshairs are parallel and six precise lines represent the outer crosshair. We have not investigated either assumption further.

The algorithms explained track the crosshair position and orientation correctly and with sub-pixel accuracy, although the inner crosshair cannot be found by means of cross-correlation. Due to the determination of the crosshair lines and the sub-pixel accurate geometric reference to the target axis, the intersection point of the crosshair lines does not have to be determined again in every image but may also, for example, be covered. This offers the measuring system more flexibility.

In the explained practical application we show results of several calibration calculations. We have shown that it is not necessary to calculate a separate photogrammetric calibration for each distance to the object. For the combination of camera and total station used, areas can be defined which, for example, meet the requirements of the SHM and can be carried out with one calibration of the optical system. As the distance to the calibration distance increases, so do the angular deviations. Generally, the deviations towards the edge of the measurement image are larger. This shows that the distortion approach varies significantly, especially in a close-up range. We assume the influence of the changed camera constant c is responsible for the systematic deviations. The use of one calibration can save a lot of time during the measurement, as it is no longer necessary to measure the calibration pattern again and calculate it.

Furthermore, we carried out an exemplary deformation measurement on a steel bridge and demonstrated the successful use for the determination of natural frequencies. Due to the high recording frequency, it is possible to record the vibration behaviour. Further research is needed in this area, such as the comparison to other sensor systems.

3.C Classification and Object Detection with Image Assisted Total Station and Machine Learning

Abstract

This paper deals with applications of digital imaging total stations in a geodetic context using artificial intelligence (AI). We present two different use cases. The first is to minimise manual intervention by the operator by classifying images with different backgrounds. We use a developed software to control a total station extended by an industrial camera, which is used for the in-situ calibration of the camera. We show that the AI successfully tests the captured image for its suitability for further use and under which circumstances the AI fails. The second case is the detection of different geodetic targets (reflective and non-reflective). Captured images of an imaging total station are automatically checked to see whether a supposed target is shown in the image, identify it and localise it in the image. Already implemented applications for target identification are to be supported in this way and extended by further information.

3.C.1 Introduction

In the following we evaluate the use and applicability of captured images from an Image Assisted Total Station (IATS). The cameras built into IATSs are often used to provide a live overview on the display. The user is thus offered an additional possibility to overlook the field of view without having to actively look through the telescope. At the same time, motorised total stations can be controlled via the touch field, and the instrument can be aligned. For documentation purposes, images can be saved and manually supplemented with field notes. Measurement data can be checked in real time via the display and additional points in the images can be determined in the post-process when images are taken from different directions. Up to now, traditional image processing algorithms have been used for this purpose. The built-in CMOS sensors also support the automatic target detection for identifying the spot. The spot represents the emitted laser beam of the IATS and is used to identify the direction to the target point [115].

The measurement system Modular Digital Imaging Total Station (MoDiTa) developed at the i3mainz of the Mainz University of Applied Sciences expands an IATS modularly by means of a digital industrial camera. Applications for non-contact optical Structural Health Monitoring (SHM) of technical building structures are to be evaluated with it [99]. Images from the external ocular camera are used by means of template matching in conjunction with cross correlation to observe discrete points on the object of interest [36]. Another example of using an external camera is DAEDALUS of ETH Zurich [14, 67]. Image processing algorithms such as multi-ellipse matching are used to detect targets [58]. Another design offers the advantage of a fixed camera with an instrument such as commercial IATS. The fixed camera provides constant calibration parameters as opposed to the modular version which requires calibration after reconfiguration. An early prototype is mentioned in [71] and the prototype series IATS2 from the manufacturer Leica Geosystems is mentioned in [66, 72]. In [77], algorithms such as blob analysis, edge detection and feature matching are cited as particularly successful for object recognition. [89] uses images from the telescope camera and target centroid detection to align the IATS and extract the target pattern using line features.

Artificial intelligence (AI) methods are increasingly being used in a wide variety of areas and are no longer limited to the software industry. The machine learns to solve a precisely defined task through extensive training. The intelligence is based on the recognition of patterns and regularities within the data. In the field of image recognition, Convolutional Neural Networks (CNNs) have proven to be an effective tool. The great success of CNNs is due, on the one hand, to the large number of training images available, which are processed under high computational power to determine the relevant features for the task. On the other hand, CNNs allow an unlimited depth of their layered architecture, which makes it possible to learn even very complex features and defines the term Deep Learning (DL). One Machine Learning (ML) task in this context is classification. Based on matching features, the classification is made into a certain class. A distinction can be made between binary, multi-class, multi-label and imbalanced classifications. Object detection is the recognition of instances of semantic objects of a certain class. Bounding boxes are formed around the classified object. Two tasks are solved. Within the image the class is recognised and localised [116, 117].

We describe two different use cases below. The first is to minimise manual intervention by the operator by classifying images with different backgrounds. The second application is to use the cameras that have already been integrated in total stations to support or extend already proven methods for the detection of targets and to make additional information about their properties available to the user. We focus on two topics where we expect DL to provide complementary, meaningful benefits to the traditional methods currently being used successfully. We describe the two approaches with self-generated datasets and evaluate them for their practical application.

3.C.2 Classification

The following classification application refers to the Modular Digital Imaging Total Station (MoDiTa) measurement system developed at i3mainz. Here we extend a total station with an industrial camera [24, 69]. The controlling is carried out via the self-developed control software and enables both the camera and the total station to be addressed exclusively via this. The system calibration can be carried out semi-automatically directly on site. For the calibration, the crosshair for each project is determined once in the image. This is done using the traditional image processing approach that identify the longest lines in the image. The crosshair is determined with subpixel accuracy using edge detection and compensation adjustment. In order for the detection to be successful, a uniform background is required. This should not be too dark, too bright nor irregular so that the six longest lines can be determined. Interfering lines and edges must be avoided; otherwise, false detection will occur. So far, this selection has been done manually. The user has to decide for himself whether the image sufficiently represents the crosshairs. The aim is to fully automate this process. The selection of a suitable image from the current stream of images should be as automatic as possible, thus eliminating the need for a manual step. After initial detection, crosshair tracking is already highly reliable during an on-site project, which is why we do not consider ML for ongoing crosshair tracking. The complete approach ensures that all mechanical camera movements are accurately corrected computationally. This represents a binary problem. The method we use is supervised learning. The requirements for this are annotated image data, known as training data. These provide the learning goal for the machine. In the form of pairs, the learning goal is

defined on the basis of examples of input and corresponding correct output. During training, the model learns to predict the correct output for new inputs [117]. Binary image classification is a comparatively simple task where the machine only has to learn whether the searched object is represented. We based the development of the image classifier on the DL library of MVTec HALCON. The software was made available free of charge as part of the "MVTec on Campus" programme (www.mvtec.com). The pre-trained model provided is trained on internal, industry-related images. Re-training thus requires a significantly smaller amount of training datasets. The pre-trained network contains several hidden layers. This makes it suitable for more complex classification tasks, but it is more computationally and memory intensive. This leads to small batch sizes. The exact structure of the network is not described [118].

3.C.2.1 Data Resources and Training

For the acquisition of the necessary dataset, we used the MoDiTa measuring system and its control software. We recorded the images with different industrial cameras. We took care to capture images that were as ideal as possible for the images to be classified as detectable. In these images, the complete crosshair is shown with an even and sufficiently bright background. For the dataset that is not to be used for crosshair detection, we used synthetic data and original image captures (*Figure C.2-1 a-c*). This dataset contains images that show the geodetic crosshairs with overlaps, overexposures, inhomogeneous backgrounds, additional lines (which can lead to false detections) or with images that are too dark for the classical digital image processing algorithm to achieve a successful result. The classification of the images into the two classes was carried out manually and checked with the support of the control software. Based on previous research [119], the original dataset is expanded. Our first results showed that the model has learnt to decide whether the image is suitable for detection based on the background. It does not check whether a crosshair is visible. Thus, a uniform light grey image is classified as suitable, even if there is no crosshair in it. To take this into account, we captured additional images that have extreme overexposure or underexposure. The images have a uniform background but represent the crosshairs inadequately. The developed software cannot detect the crosshairs in these images using the traditional image processing programs. The aim of this extension is not to make the classification exclusively dependent on a uniform background, but to specify the concrete detectability of the crosshairs in the training data in a better way, i.e. much more clearly distinguished than before. To account for the unbalanced dataset, we determined class weights according to the number of images. To synthetically enlarge the image dataset and avoid divergence effects such as overfitting, we enlarged the dataset by 50 % using data augmentation [120]. We mirrored the images around the two axes and performed rotations up to 50 gon. The dataset was split 70 : 15 : 15 (training : validation : test). While the validation and test data are used, among other things, for performance evaluation, the largest dataset, the training images, are used to optimise the CNN. During training each image is abstracted through a series of filters as well as activation functions, summarised as weights, and down sampling strategies. The resulting feature maps assign importance and thus distinctiveness to different regions or objects in the image in the context of the global query. In supervised learning, the prediction of the CNN is compared to the ground-truth labels. Due to the given classification, the CNN can optimise the weights independently, which is why we also speak of end-to-end trained networks. The adjustment of the weighting is carried out

automatically by a backpropagation of error and represents a gradient descent procedure that is specially optimised for multi-layer networks.

To control the learning process, there are so-called hyperparameters in ML. These differ from other parameters that are determined during training. Depending on the specific use case and the basic network topology, there are different hyperparameters. They influence the speed and quality of the learning process. The goal of hyperparameter tuning is to generate the best possible model based on the training images. Two important hyperparameters for tuning the filter weights are the learning rate and the momentum. The learning rate describes the influence of the gradient on the adjustment of the weights. Too low a learning rate results in an unnecessary number of iterations, while too high a learning rate results in divergence. In our case, we start with a higher learning rate and reduce it during training [121]. Momentum describes the influence of previous adjustments. Instead of using only the gradient of the current step to determine the search, the gradients of the past steps are also used to determine the direction. This dampens the oscillation [118]. Since we have a small batch size, six in this case, we determine the momentum with a high value.

We visualise the progress of the training by means of a time-dependent loss function. The function gives the mean error over a batch of image data. As long as the loss function steadily decreases and converges to the mean error from the validation data, the CNN learns generalised object representations.

3.C.2.2 Evaluation

To be able to evaluate the model, the model with the highest F1-Score is examined. The F1-Score represents a metric for assessing the quality of the model on the dataset. For this we need precision and recall (also called sensitivity) [122]. The F1-Score represents the harmonic mean between these two values. A value of 1 corresponds to a perfect model. The recall indicates which proportion of correct positive results were correctly identified (*Equation (3.43)*). Precision is the ratio of all correctly predicted positive instances to all positively predicted instances. It indicates the proportion of positive identifications that were actually correct (*Equation (3.42)*) [39, 122].

$$precision = \frac{TP}{TP + FP} \quad (3.42)$$

$$recall = \frac{TP}{TP + FN} \quad (3.43)$$

$$F1 - Score = 2 * \frac{precision*recall}{precision+recall} = 0.974 \quad (3.44)$$

True Positive (TP) means correctly identified, False Positive (FP) means incorrectly identified and False Negative (FN) means incorrectly rejected.

The high value shows that the amount of incorrectly classified cases is very low. By expanding the dataset with additional overexposed and underexposed images, the F1-Score marginally increased from 0.967 to 0.974.

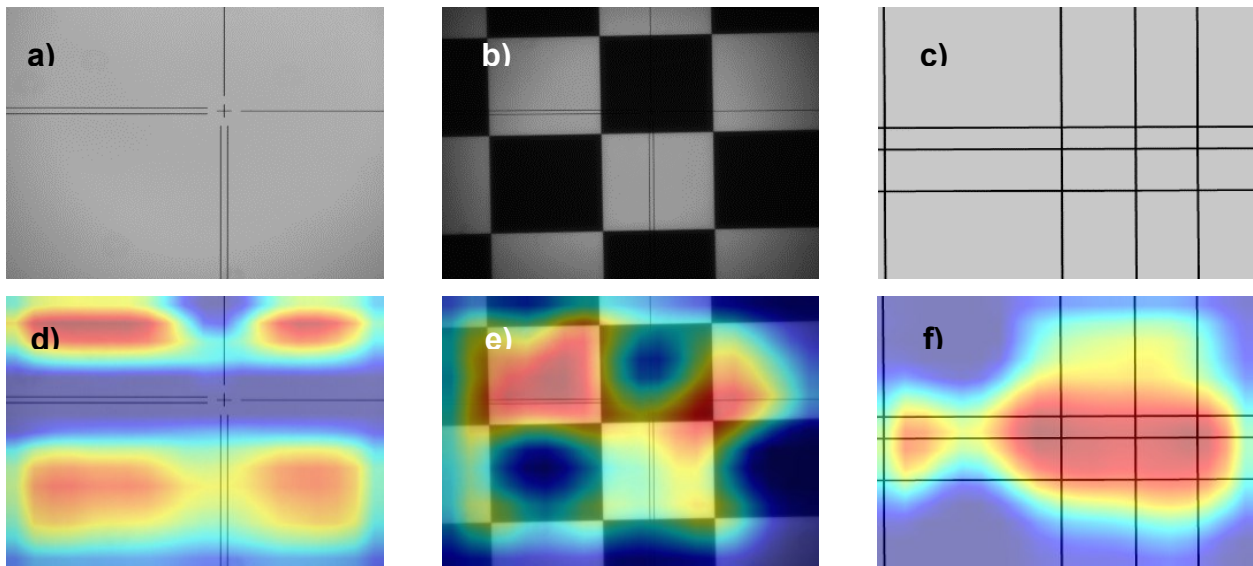



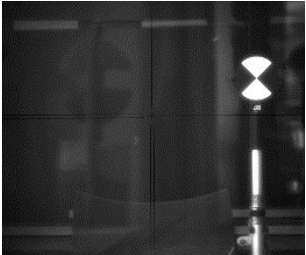


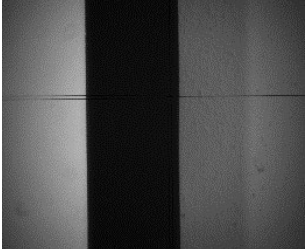
Figure C.2-1 Training images and heatmaps. **Top:** Images from the training dataset. a) shows an ideal image for the class of detectable images. An original take without modification. b) shows an original image, but with an inconsistent background, which makes detection difficult or even impossible. c) A completely synthetically generated image without geodetic crosshair but with too many lines. **Bottom:** Heatmaps. d) classified as suitable for detection with a confidence value of 0.97. e) and f) are classified as unsuitable for detection with a confidence value of 1.00.

We use heatmaps to evaluate whether the classified object features correspond to what is being searched for. These highlight in colour the relevant areas for classification (see *Figure C.2-1 d-f*). This does not correspond to segmentation but can be used for evaluation. The calculation is carried out using the “Gradient-weighted Class Activation Mapping” (Grad-CAM) method [123]. *Figure C.2-1 d-f* show the calculated heatmaps. The sections highlighted in colour show the relevant areas for classification for the CNN. In all three examples, the classification is performed correctly. In example *d*) the areas in the background are highlighted in particular. Examples *e*) and *f*) show that the interfering lines or the irregular and partly too dark background are decisive for judging these images as unsuitable for the detection of the crosshairs by means of traditional image processing algorithms.

For further evaluation, we examined random samples. We compare the images classified by CNNs with conventional image processing methods. For this, we use the control software for MoDiTa. This software does not classify the images, but it directly triggers the recognition of the longest lines in the image in order to detect the crosshair using digital image processing algorithms. If this is not successful, an error message is shown to the user. *Table C.2-1* shows a comparison of images that are suitable for conventional detection and those that are not. Overall, the CNN trained with extended dataset has a higher matching rate with traditional image processing. With this CNN, better predictions can be made as to whether the captured image can be used to detect the crosshairs. *Table C.2-1* shows results from inference. In inference, the CNN evaluates previously unknown images [124]. *Table C.2-1* image b) is classified as unsuitable by both CNNs,

although a detection can be performed successfully. Here it can be seen that the combination of an unstable background with low exposure continues to be challenging for the trained CNNs. Whereas images a) and c) represent an improvement in classification given the original CNN. The bright and even backgrounds alone are no longer sufficient to trigger the crosshair detection and are consequently classified as unsuitable.

Table C.2-1 Comparison of classification by ML and traditional image processing. Shown are them results of the inference with the calculated confidences of the trained CNN and the result of the triggered detection using the control software of MoDiTa. We highlighted in colour the conformity of the results with traditional image processing.

Image	Trained with original dataset	Trained with extended dataset	Traditional image processing
	☑ = suitable ☒ = unsuitable		
a) 	☑ = 0.23 ☒ = 0.77	☑ = 0.02 ☒ = 0.98	☒
b) 	☑ = 0.00 ☒ = 1.00	☑ = 0.12 ☒ = 0.88	☑
c) 	☑ = 0.91 ☒ = 0.09	☑ = 0.05 ☒ = 0.95	☒
d) 	☑ = 0.89 ☒ = 0.11	☑ = 0.68 ☒ = 0.32	☑
e) 	☑ = 0.76 ☒ = 0.24	☑ = 0.07 ☒ = 0.93	☒

3.C.3 Object Detection

The use of IATS is common practice. However, the use of the built-in camera or cameras is often limited. In many cases, the telescope and overview cameras are only used for documentation or to support control via touchpad. Applications of photogrammetric methods for automatic target acquisition of non-retroreflective objects by a total station, e.g. using SURF (speeded-up robust feature) or SIFT (scale invariant feature transform), are described in [125]. In the geodetic context, varieties of different targets with different properties are in use, depending on the instrument and application used. These differ, among other properties, by a different offset, which ranges from several millimetres to centimetres. It is often necessary for the user to select or locate them manually. Likewise, in order to be able to use applications such as automatic targeting or electro-optical distance measurement (EDM), a target with retroreflective properties must be aimed sufficiently accurately. Automatic target detection requires a clearly visible spot, i.e. the energy must be set correctly. However, this depends on the distance, which is often unknown. With EDM, the field of view is smaller, which means that the prism must be aimed sufficiently accurately beforehand. This in turn requires automatic targeting [84, 115]. In both cases, the overview or telescope camera that have been installed could be used for support.

The aim is to distinguish between different targets and to roughly determine the position of the object being searched for. It is not a matter of locating the centre exactly. For this investigation we have defined four different classes of geodetic targets. The images are intended to serve as an additional source of information and to support methods that have already been implemented. Object detection is about finding the object you are looking for and assigning it to a class. If an object is found, bounding boxes are formed and its localisation is compared with the design of the ground truth. During training, the network learns to adapt these so that they match the search results as well as possible.

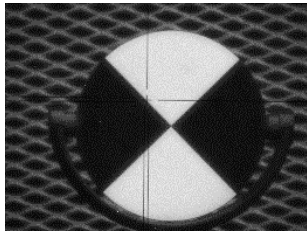
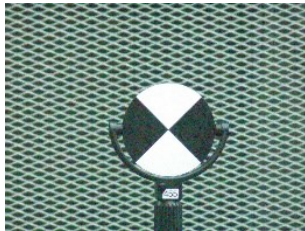

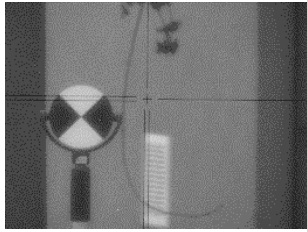


3.C.3.1 Data Resources and Training

To collect the necessary image data, we used the MoDiTa measuring system and the internal cameras of a Leica Nova TS60 [3]. The overview camera has a field of view of 19.4° and the telescope camera a field of view of 1.5° . The industrial camera used changed and so did the field of view. The image data was captured simultaneously from all three cameras. We recorded a black and white target, a 360° prism, a circular prism and a metal sphere with different backgrounds and distances from the instrument. The distances were between 8 m and 55 m. *Table C.3-1* shows original images recorded by the different cameras at the same time. It illustrates the different fields of view and their restrictions. We limited the distance to 55 m, because beyond this distance the targets are only represented by about 20×20 pixel in the overview camera. Due to the change in dimensions, not enough information about the targets remains in the image. The aim is to use images from the different cameras with equal balance in the input data. We expanded the dataset again by means of augmentation. Especially deeper architectures benefit from data augmentation [126]. Flipping, histogram equalisation, blurring, sharpening, morphological transformation and cut out/random erasing were used [127, 128]. In total we generated about 1000 images per class. The division (training : validation : test) is again in the ratio 70 : 15 : 15. Since the supervised learning method is used, notations are

necessary. The annotations were created semi-automatically. Since the images are captured using MoDiTa and the developed software, it was possible to determine the searched target with pixel accuracy by means of template matching during the capture. By means of cross-correlation, the searched pattern is detected in the images and the coordinate information is passed on for further use.

Before training, some parameters are set. For example, based on the input data, we determine that only one target is presented per image and therefore only one determination is required. The minimum confidence level is set at 0.7. Later in the evaluation respectively inference, we reduce this value for the assessment to 0.3. The initial learning rate is set higher and reduced in succeeding epochs. The loss function is adjusted using this gradient weighting. In this case, the batch size is four images, which are processed simultaneously. Since the batch size is small, the momentum is set high.

Table C.3-1 Examples of original images from different cameras and distances of a black and white target. The crosshairs are clearly visible in the images from the industrial camera.

Distance	Industrial camera	Telescope camera	Overview camera
~13 m			
~40 m			

3.C.3.2 Evaluation

The results of the evaluation show a precision (*Equation (3.42)*) of 98.3 % true positives over all classes. Half of the false positive results are due to incorrect or insufficiently accurate localisation. The recall (*Equation (3.43)*) value is calculated as 91.7 % for all classes. Here the 360° degree survey prism shows the worst results of the individual targets.

A common method for evaluating object detection is the mean Average Precision (mAP) [39, 81]. This averages the Intersection over Union (IoU) over all classes as well as threshold values. It is therefore a measure of how well the instances were found and classified. The IoU is a measure of the accuracy of the localisation. For the proposed box, the ratio of the intersection and the union with the ground truth is formed. The IoU is the ratio of the two areas intersection and union [39, 81]. *Table C.3-2* shows the class AP and mAP for the four different targets.

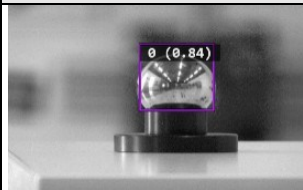
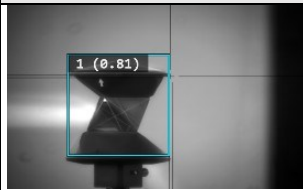
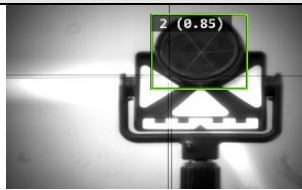

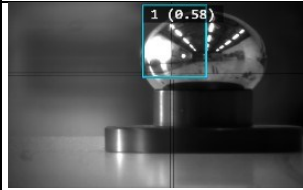
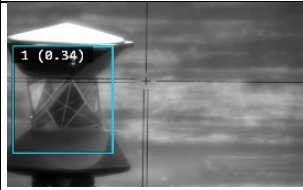
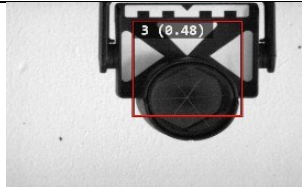

The exact location of the targets is not relevant to the question. Nevertheless, the results, with a mAP for all classes of 82 % (mean IoU), show the successful detection of the different classes.

Table C.3-2 Evaluation results: class and mean Average Precision for the test dataset.

Class (Target)	AP for IoU70	AP for IoU80	AP for IoU90	mean IoU AP
Metal solid sphere	0.98	0.96	0.91	0.93
360 degree survey prism	0.79	0.77	0.65	0.72
Circular prism	0.93	0.88	0.69	0.83
Black & white target	0.95	0.80	0.59	0.81
Mean class AP	0.91	0.85	0.71	0.82

In the inference, the network is given a dataset that is completely unknown so far. In order to be able to evaluate the object detection more comprehensively, we have set the value for the minimum confidence to 0.3 in this phase. Exemplary results are shown below in *Table C.3-3*. The images above illustrate a successful example of each target detection. It should be noted that the representation of the searched objects in the images is ideal, so that high confidence values are achieved. In the second row, both incorrect classifications and results with a small confidence are represented. The bounding boxes determined by the CNN are shown with the respective confidence. It is clear that the bounding boxes do not exactly enclose the targets but are sometimes determined to be too small or too large. Since in this case this is not important for the question, these deviations are acceptable. With the lower threshold value, the metal sphere shows the highest number of false classifications. Since the sphere has a reflective surface, the representation of the sphere in the image changes depending on the environment. Nevertheless, it reaches the highest value with the mAP. However, since the sphere is also very small, the detection is limited to a much smaller distance. It has been shown that in the case of images, especially in the nearer range, the net delivers less favourable results. Often the target is not detected, particularly if it is only partially visible due to the detail.

Table C.3-3 Examples from inference. The enclosing rectangles are the bounding boxes determined by the CNN. The different colours represent the different classes. cs stands for confidence score.

Metal solid sphere	360 degree survey prism	Circular prism with a removable target plate	Black and white target
			
Correctly classified cs = 0.84	Correctly classified cs = 0.81	Correctly classified cs = 0.85	Correctly classified cs = 0.71
			
Incorrectly classified as 360 degree prism cs = 0.58	Correctly classified but with a low cs = 0.34	Incorrectly classified as a black and white target, cs = 0.48	Correctly classified but with a low cs = 0.45

3.C.4 Conclusion

The application for classifying the image scene for the subsequent crosshair detection shows promising approaches. Looking at the results in detail, it becomes apparent that the ML approach shows its advantage in the case of multiple and distinct disturbances of the background. The previous approach using traditional image analysis always calculates the longest lines even if they do not belong to the crosshair. This can lead to an erroneous determination and is only noticeable when checked by the user. The CNN successfully classified these images as unsuitable. By expanding the dataset to include images with uniform backgrounds but insufficient crosshair representation, the classification could be made more robust. Images that were previously still classified as suitable are successfully sorted out in the second trained net. However, deficits are still evident in images with irregular dark backgrounds. Overall, the decision-making process is supported significantly and the user can be relieved of the responsibility for this work step.

An integration of the network into the existing software package is unproblematic due to the implemented image processing software from MVTec. The package would increase significantly in necessary storage space but would remain reasonable (~200 MB). The resulting time loss for the classification before the detection by means of the already implemented traditional image processing is 200-500 ms per image (2-6 MB, 2456 x 2054 x 8).

The application for object detection of different geodetic targets shows good results. The images are correctly classified with high precision and recall, and the objects are correctly located. Support for a successful application in the measurement process is therefore given. The image-based approach can also be used to identify non-retroreflective targets. Retroreflective targets such as prisms can be differentiated to a certain level and the offset can be applied automatically. A critical factor here is the distance of the target or the sufficient representativeness of the searched object in the image. Since the object detection delivers less accurate results for close-up images and when the target is partially cut off or obscured, one approach would be to expand the dataset specifically for this purpose. It has been shown that the overview camera, due to its specific characteristics, is much more limited in its distance than the telescope camera. From a distance of over 55 m, the images are not usable for training purposes. The dataset could, however, be expanded with images from the other cameras. Investigations into the distance up to which successful object detection can be carried out using these cameras and machine learning are still pending. The size of the target used is also crucial. As the aim is not to detect the target as precisely as possible in the image but only to find it sufficiently precisely to trigger further steps in the measurement process, the largest possible mAP is not necessary. Here, further investigations could be focused on more precise localisation. The determined bounding boxes could be used as an approximation to search for features in a smaller area using traditional image analyses. In this study, we have limited the possible targets per image to one. This does not correspond to every real-life situation, so there are still more configurations that are possible. Likewise, the minimum confidence value was set at 0.7. This value could be set lower. A generally valid value would first have to be proven in practice.

4. Outlook

In recent years, many manufacturers have been focusing on the development of total stations towards multi-sensor systems. The combination of sensors can be technically implemented in hardware, allowing the respective strengths of the systems to be combined. 3D laser scanning, GNSS or digital imaging technologies are often already standardised functionalities in high-end products. Thus, the benefits of the integrated cameras are further expanded and no longer refer only to a supporting visual aid. The display of survey, planning and sketch data on the instrument display is expected and positively accepted by the user of premium instruments. The on-board software offers apps for operating the measurement tasks and is designed with user-friendliness and simplicity of use in mind. Functionalities such as aligning the instrument via touchpad show the progress of the development towards interactive tactile instrument control, as such a functionality is already well known to all due to smartphones. The user no longer has to look through the long-range optics; this is done by the built-in camera sensors. If this information is mirrored onto a tablet, for example, the instrument can be controlled wirelessly at a distance. The direction and distance are recorded by simply triggering the measurement with a tap on the touchpad. The precise determination of the position is carried out by the automatic target detection. Developments such as the Leica CS35 [129] or the Trimble TSC7 [130] controller already demonstrate the successful use of wireless control of robotic total stations. It would be preferable to select the built-in camera sensor and transmission technology in such a way that it delivers the image information synchronously at full resolution together with the determination of the position (direction and distance). For this purpose, the complete photogrammetric calibration of the installed optics would be useful in order to be able to use the image information for precise image processing. In this way, a single engineer, without having to stand directly at the instrument, could precisely capture the survey scene both discretely and area-wise. With a fixed camera, the calibration itself can be performed at the factory and the parameters automatically adjusted. With a modular implementation, calibration is carried out in-situ by means of simultaneous calibration. Compared to other sensor systems, calibration is easy to implement in the measurement process and the software.

On the software side, greater automation is possible for image analysis. Due to the increasing growth in computing power, more and more methods can be implemented directly on site and not only in the post-process. More complex algorithms and methods can be implemented, for example, edge detection, template matching, feature matching, or ML approaches such as classification and object detection. Signalling can thus possibly be dispensed with and fields of application can also be extended to areas with a high risk of trespassing. The reliability of the possible algorithms is still dependent on the environmental conditions such as lighting and atmospheric influences. The achievable accuracy also depends on the object size, the image resolution and the quality of the optical focus [72]. Which of these methods and algorithms will prove successful in practice is still the subject of applied research. A major advantage of image analysis is its independence from signalling. As shown in the example of the MoDiTa measuring system, distinctive points are sufficient to be able to track them in the image. The use of AI will continue to develop in this area and find its place in practice. Narrowly defined tasks, as described in the example in *Section 3.C.2*, can be carried out without the need for interaction by the engineer. For example, the approach in *Section 3.C.3* for recognising

common survey targets demonstrates how it can also assist in finding non-reflective targets. The use of an overview camera to provide a large-scale view of the measurement area is a suitable option here. Due to the larger FoV, a larger area can be covered simultaneously and examined for possible targets. The larger FoV offers the advantage of being able to detect and track several targets at the same time. However, it does not have the magnification of the telescope and thus has a lower resolution, which affects the accuracy. It is conceivable here that the AI approach finds the targets it is searching for by means of object detection. The predicted bounding boxes are used to roughly determine the angular position of the targets. Subsequently, the motorised alignment of the IATS to the detected targets is carried out. Traditional image processing algorithms or already common automated target recognition (for prisms) then determine the exact centre point. A schematic representation is shown in *Figure 4-1*. This eliminates the need for the observer to manually target. Inaccuracies caused by the subjective perception of the observer are eliminated.

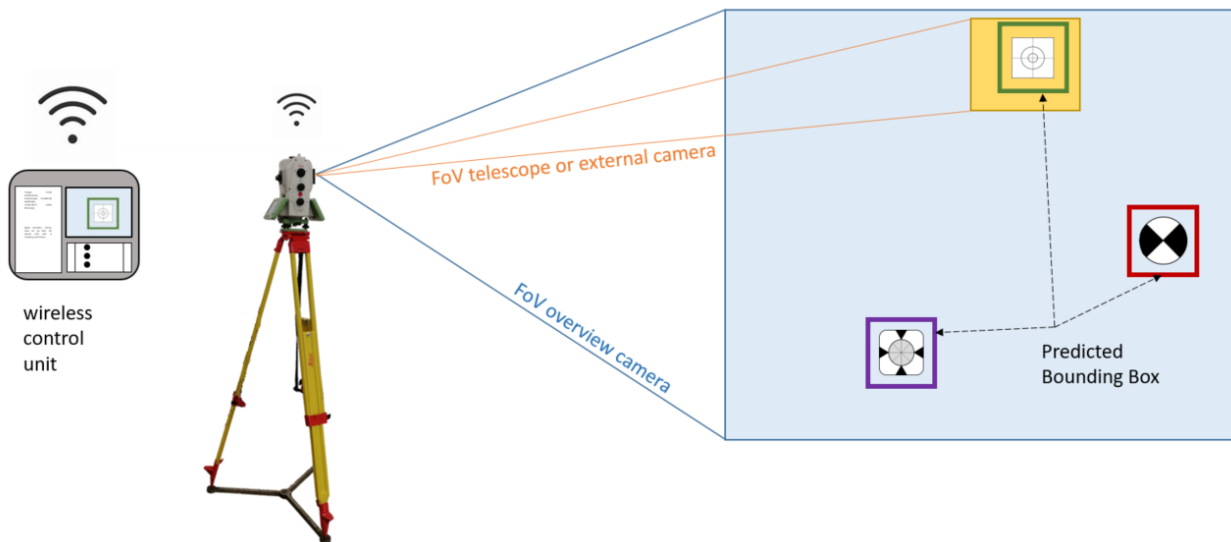


Figure 4-1 Schematic representation of target recognition. The overview camera finds the targets using AI and object detection. The IATS is then automatically aligned and the centre point is precisely determined using automated target recognition (for prisms) or traditional image processing algorithms. The control software additionally displays information about the target.

The Trimble SX12 [64] already omits an eyepiece and thus also the technical effort to integrate it. This has the advantage that complex beam splitter elements can be avoided and therefore more light can be guided to the camera chip [65]. Mobile autonomous robots equipped with measurement sensors achieve the ability to independently recognise and differentiate target marks in the environment. Additional information could be made available to the user. Depending on the target marks, the distance measurement can be independently switched to reflectorless or prism mode. Additional constants could be added or omitted independently. With corresponding visible numbering of the targets, the additional use of optical character recognition (OCR) would be conceivable. With known 3D coordinates, autonomous position determination of the robot carrier is possible with low-cost adhesive markers.

A more professional control of the camera parameters in fixed systems is inevitable. Camera-side settings such as frame rate, frame duration, subsampling, binning, area of

interest, line scan mode, temporal synchronisation and triggering of the image capture are only some of the necessary settings and requirements that should be available in order to be able to adapt to the application and environment in practice.

Whether global shutter or rolling shutter will become established in practice will depend strongly on the applications. Especially when it comes to capturing moving objects or when the camera moves during the exposure, motion blur can occur with long exposure times. Global shutter has the advantage that there is no time difference within the image, but the speed of readout is slower. The rolling shutter method results in a time delay between the first line and all subsequent lines and can add up on large sensors. The use of various filtering techniques prior to image analysis is also purposeful. For example, the use of smoothing filters such as mean value filters, Gaussian filters or median filters to counteract image noise would be conceivable. Noise occurs during the recording of the image or the transmission of the image. The most important source of noise is mentioned by [131] with:

- environmental conditions;
- insufficient light conditions and temperature of the sensor;
- interference during transmission;
- dust particles.

Noise removal algorithms reduce or remove the visible noise. Since the use of IATS is intended for use in very different locations, the ambient and lighting conditions cannot be determined in advance. In this way, a degradation of the image in case of insufficient exposure can be counteracted, adapted to the situation. In this area of image processing, the development towards the use of AI to automatically recognise which approach to denoising is optimal is also evident (cf. [132]) and will continue to be part of research in the future.

The modular principle, as used in MoDiTa, shows its advantages through the possibility of adapting the sensor technology to the application. For example, special cameras such as hyperspectral cameras can be relatively easily attached to an existing total station and calibrated using the control software. The barrier to the combined use of tacheometry and photogrammetry is alleviated because total stations already in the portfolio can be expanded. It would be advantageous if the power supply and data transmission of the camera could be installed within the housing of the instrument. With the current setup, the cables run past the outside by means of a mounting bracket. This can possibly lead to a tensile load on the telescope and therefore to faulty data acquisition. Since the camera is attached to the outside of the telescope without any interference with the instrument, no heat generation affecting other sensors has been detected so far. This would have to be taken into account and investigated more closely in the case of internal cabling. The weight of the camera itself must also be considered. The counterweight is currently attached to the object lens by means of a clamping screw. The motors could be strengthened and the axes could be clamped during direction reading, distance measurement and image acquisition to make the counterweight unnecessary. At this stage, it is not possible to perform image capture and distance measurement simultaneously at high capture rates, as distance measurement requires significantly more time. As an example, the Trimble S9/S9 HP with a measurement time for the

distance in Direct Reflex Plus mode (without a prism) standard with 0.2 - 1 Hz and in tracking mode with 2.5 Hz is specified [63]. At the same time, total stations with a scanner function also show that a distance measurement with a recording rate of up to 30 kHz can be achieved [26, 64].

Overall, a combination of scan data and high-resolution images using an IATS is a way to develop new evaluation methods. This combines the RGB+D data (D stands for distance determined from the laser scanner data). Laser scans are used for distance changes in the line of sight, and image data is used to determine displacements in two axes perpendicular to the camera's line of sight [6]. Classical applications such as stationing can be performed with one instrument. In this case, too, a complete photogrammetric calibration of the camera would be highly desirable. [93] mentions here the use of HDR images (High Dynamic Range) as a helpful functionality to counteract inhomogeneous lighting. As described earlier in this chapter, extended access to the camera's controls is appreciated.

If we look at IATS in relation to Building Information Modelling (BIM) and construction monitoring, the potential of using IATS also becomes obvious here. The repetitive comparison of the actual state to the intended state is increasingly becoming the focus of surveying work. Since stake-out tasks are standard tasks on site and are carried out using IATS, it makes sense to use this instrument for progress monitoring as well. IATS represent an intersection of the sensor technology used here. Previous data is often represented in the form of graphs and charts. Thus, there is no real picture of the object to be monitored, but only the provision of the individual information. Visual monitoring is often a key factor in structural monitoring. The captured 3D data are directly georeferenced and could therefore be integrated into a BIM without media breaks. Monitoring during the use phase of the structure can also be taken over by IATS, even if only in part. As already shown, it is also possible to record high-frequency movements of a structure or parts of a structure. Simple passive targets, thanks to image analysis, can replace accelerometers and their power supply and data processing units. For the implementation of stationary IATS for autonomous permanent monitoring, a permanent power supply and stable communication infrastructure for remote access would be a requirement. This could positively influence the inspection and maintenance period for a walk-through by specialised professionals. For the analysis of the delivered images, the use of ML would be applicable for the inspection of the object. Superficial changes in the structure such as spalling or cracks in the concrete would be automatically recorded by the AI and made available geo-referenced via BIM as an information management system [133, 134]. The connected FEM model could be expanded or updated using the prepared information from the SHM data, and thus again allow conclusions to be drawn about the condition of the structure. At the same time, AIs also show their potential in the computer-aided assessment and interpretation of results. This involves the recognition of changes in previous regularities or laws [135]. A BIM enhanced in this way thus also becomes an early warning system. In fact, the use of this measurement system proves to be a practical tool over the entire life cycle of a building.

Improvements in both hardware and software are required and desired for the simplified and automated application of IATS. The potential for innovation exists both in the installed sensor technology, the transmission and supply technology as well as in the research of application-related evaluation and analysis strategies. AI in particular will continue to

establish itself in the field of image-based surveying. The algorithms and methods shown in this work hopefully represent a contribution to future developments.

List of Figures

Figure A.2-1	Chart for identifying relevant papers.....	19
Figure A.4-1	Observed point and measuring results.....	31
Figure B.2-1	MoDiTa.....	36
Figure B.4-1	Geodetic crosshair with pixel coordinate system.....	40
Figure B.4-2	provides a simplified overview of the single steps.....	41
Figure B.4-3	Detection of the outer crosshair.....	42
Figure B.4-4	Rough detection of the inner crosshair.....	42
Figure B.4-5	Precise detection of the inner crosshair.....	43
Figure B.4-6	Point detection of the outer crosshair.....	44
Figure B.4-7	Principle of adjustment (simplified example).....	44
Figure B.4-8	Crosshair matching.....	46
Figure B.5-1	Exemplary view of two IATS (MoDiTa) on an elongated structure with different forced distances.....	47
Figure B.5-2	Results of the self-calibration.....	50
Figure B.5-3	Plot of estimated parameters and a-posteriori/empirical standard deviation.....	51
Figure B.5-4	Deformation measurement on a steel bridge.....	53
Figure C.2-1	Training images and heatmaps.....	59
Figure 4-1	Schematic representation of target recognition.....	67

List of Tables

Table 2.2-1	Different designs of IATS according to the position of the image sensor	6
Table A.1-2	Structure of confusion matrix.....	13
Table A.3-1	Main specifications and set-up of IATS prototypes.....	23
Table A.3-2	Comparison of the results for correlation based and shaped based matching.....	27
Table B.5-1	Results of the adjustment.....	48
Table C.2-1	Comparison of classification by ML and traditional image processing...	60
Table C.3-1	Examples of original images from different cameras and distances of a black and white target.....	62
Table C.3-2	Evaluation results: class and mean Average Precision for the test dataset.	63
Table C.3-3	Examples from inference.....	64

Abbreviations

AI	Artificial Intelligence
ANN	Artificial Neural Network
AP	Average Precision
ATR	Automatic Target Recognition
AVT	Allied Vision Technologies
BIM	Building Information Modeling
BRIEF	Binary Robust Independent Elementary Features
CCD	Charge-Coupled Device
CET	Central European Time
CMOS	Complementary Metal-Oxide-Semiconductor
CNN	Convolutional Neural Network
DL	Deep Learning
EDM	Electro-optical Distance Measurement
FEM	Finite Element Method
FFT	Fast Fourier Transformation
FoV	Field of View
GNSS	Global Navigation Satellite System
HDR	High Dynamic Range
IASTS	Image Assisted Scanning Total Station
IATS	Image Assisted Total Station
IoU	Intersection over Union
JPEG	Joint Photographic Experts Group
KBS	Knowledge-Based Systems
LoG	Laplacian of Gaussian
LSM	Least Squares Matching

mAP	mean Average Precision
ML	Machine Learning
MoDiTa	Modular Digital Imaging Total Station
NCC	Normalised Correlation Coefficient
NIR	Near Infrared
OCR	Optical Character Recognition
OTR	Optical Target Recognition
RGB	Red, Green, Blue
ROI	Region of Interest
SHM	Structural Health Monitoring
SIFT	Scale Invariant Feature Transform
SURF	Speeded Up Robust Feature
TLS	Terrestrial Laser Scanner
UAV	Unmanned Aerial Vehicle
VGA	Video Graphics Array

References

1. Goodfellow, I.; Bengio, Y.; Courville, A. *Deep Learning*; The MIT Press: Cambridge, Massachusetts, 2016, ISBN 978-0262035613.
2. Leica Geosystems AG. *Leica Nova TM60: Datasheet*; Heerbrugg, Schweiz, 2020.
3. Leica Geosystems AG. *Leica Nova TS60: Datasheet*; Heerbrugg, Schweiz., 2020. Available online: <https://leica-geosystems.com/en-gb/products/total-stations/robotic-total-stations/leica-nova-ts60> (accessed on 4 March 2021).
4. Topcon Corporation. *GTL-1200 Datasheet*, 2022.
5. Peroš, J.; Paar, R.; Divić, V. Application of Fused Laser Scans and Image Data—RGB+D for Displacement Monitoring. *Conferences on Engineering Surveying*; Springer, Cham: Germany, 2021; pp 157–168.
6. Peroš, J.; Paar, R.; Divić, V.; Kovačić, B. Fusion of Laser Scans and Image Data—RGB+D for Structural Health Monitoring of Engineering Structures. *Applied Sciences* **2022**, *12*, doi:10.3390/app122211763.
7. Trimble Inc. *SX10 Scanning Total Station: Datasheet*. Available online: <https://geospatial.trimble.com/products-and-solutions/trimble-sx10> (accessed on 12 June 2021).
8. Rosten, E.; Drummond, T. Machine Learning for High-Speed Corner Detection. In *Computer Vision - ECCV 2006*; Leonardis, A., Bischof, H., Pinz, A., Eds.; Springer Berlin Heidelberg: Berlin, Heidelberg, 2006; pp 430–443, ISBN 978-3-540-33833-8.
9. Calonder, M.; Lepetit, V.; Strecha, C.; Fua, P. BRIEF: Binary Robust Independent Elementary Features. In *Computer Vision - ECCV 2010*, Heraklion, Crete, Greece, 5. - 11. September; Daniilidis, K., Maragos, P., Paragios, N., Eds.; Springer Berlin Heidelberg: Berlin, Heidelberg, 2010; pp 778–792, ISBN 978-3-642-15561-1.
10. Rublee, E.; Rabaud, V.; Konolige, K.; Bradski, G. ORB: An efficient alternative to SIFT or SURF. In *2011 International Conference on Computer Vision*, Barcelona, Spain, 6. - 13. November, 2011; pp 2564–2571.
11. Luhmann, T.; Robson, S.; Kyle, S.; Boehm, J. *Close-Range Photogrammetry and 3D Imaging*; de Gruyter, 2020.
12. Schirmer, W. Universaltheodolit und CCD-Kamera - ein unpersönliches Meßsystem für astronomisch-geodätische Beobachtungen. PhD thesis; Technische Universität München, München, Germany, 1994.
13. Paar, R.; Roić, M.; Marendić, A.; Miletić, S. Technological Development and Application of Photo and Video Theodolites. *Applied Sciences* **2021**, *11*, doi:10.3390/app11093893.
14. Bürki, B.; Guillaume, S.; Sorber, P.; Oesch, H.-P. DAEDALUS: A versatile usable digital clip-on measuring system for Total Stations. In *2010 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, Zurich, Switzerland, 15. - 17. September, 2010; pp 1–10.
15. Hauth, S.; Schlüter, M. Digitalkameratachymeter – Einsatzmöglichkeiten für modulare und integrierte Systeme. In *30. Wissenschaftlich-Technische Jahrestagung der*

- Deutschen Gesellschaft für Photogrammetrie und Fernerkundung. Dreiländertagung OVG, DGPF und SGPF, Wien, Austria, 01. - 03. July; Kohlhofer, G., Franzen, M., Eds.; Deutsche Gesellschaft für Photogrammetrie, Fernerkundung: Wien, 2010; pp 477–484.*
16. Wasmeier, P. Videotachymetrie – Sensorfusion mit Potenzial. *avn - Allgemeine Vermessungsnachrichten*; VDE VERLAG GmbH, Berlin - Offenbach, 2009; pp 261–267.
 17. Leica Geosystems AG. *Leica Nova MS50: Datasheet*. Available online: https://w3.leica-geosystems.com/downloads123/zz/tps/nova_ms50/brochures-datasheet/Leica_Nova_MS50_DAT_en.pdf (accessed on 17 June 2021).
 18. Hecht, E. *Optik*, 3rd ed.; Oldenburger Wissenschaftsverlag: München, Germany, 2001, ISBN 3-486-24917-7.
 19. Ehrhart, M.; Lienhart, W. Development and evaluation of a long range image-based monitoring system for civil engineering structures. *Structural Health Monitoring and Inspection of Advanced Materials, Aerospace, and Civil Infrastructure 2015*; Proceedings of SPIE Vol. 9437 (SPIE, Bellingham, WA, 2015), 2015; 94370K, ISBN 9781628415407.
 20. Paar, R.; Marendić, A.; Wagner, A.; Wiedemann, W. Using IATS and digital levelling staffs for the determination of dynamic displacements and natural oscillation frequencies of civil engineering structures. In *7th International Conference on Engineering Surveying. INGENEO 2017*, Lisbon, Portugal, 17. - 18. October, 2017.
 21. Reiterer, A.; Huber, N.B.; Bauer, A. Image-based detection and matching of homologue points using feature-vectors–functionality and evaluation in a deformation measurement system. *The International Archives of Photogrammetry* **2010**, *38*, 510–515.
 22. Ehrhart, M.; Lienhart, W. Monitoring of Civil Engineering Structures using a State-of-the-art Image Assisted Total Station. *Journal of Applied Geodesy* **2015**, *9*, 174–182, doi:10.1515/jag-2015-0005.
 23. Lienhart, W.; Ehrhart, M. Statische und dynamische Überwachung von Infrastrukturbauten mit kommerziellen Videotachymetern. In *20. Internationale Geodätische Woche Obergurgl 2019*, Obergurgl, Austria, 10. - 16. February; Hanke, K., Weinold, T., Eds.; Wichmann, VDE Verlag GmbH, 2019; pp 145–156.
 24. Schlüter, M.; Hauth, S.; Heß, H. Selbstkalibrierung motorisierter Digitalkameratheodolite für technische Präzisionsmessungen. In *zfv – Zeitschrift für Geodäsie, Geoinformation und Landmanagement*; DVW e. V. – Gesellschaft für Geodäsie, Geoinformation und Landmanagement, Ed.; Wißner-Verlag: Augsburg, 2009; pp 22–28.
 25. Mugnai, F.; Caporossi, P.; Mazzanti, P. Exploiting Image Assisted Total Station in Digital Image Correlation (DIC) displacement measurements: insights from laboratory experiments. *European Journal of Remote Sensing*, 2022; pp 115–128.
 26. Leica Geosystems AG. *Leica Nova MS60: Datasheet*; Heerbrugg, Switzerland, 2020. Available online: https://leica-geosystems.com/-/media/files/leicageosystems/products/datasheets/leica_nova_ms60_ds.ashx?la=en

- gb&hash=DC24A3605EE8B0DED66F30240A8B63DC (accessed on 16 August 2021).
27. NIKON Corporation. *D3200 Kompakthandbuch*, 2012. Available online: <https://downloadcenter.nikonimglib.com/de/products/20/D3200.html> (accessed on 16 March 2023).
 28. Leprince, S.; Ayoub, F.; Klinger, Y.; Avouac, J.-P. Co-Registration of Optically Sensed Images and Correlation (COSI-Corr): an operational methodology for ground deformation measurements. In *2007 IEEE International Geoscience and Remote Sensing Symposium*, Barcelona, Spain, 23. - 28. July; IEEE, Ed., 2007; pp 1943–1946, ISBN 978-1-4244-1211-2.
 29. Ayoub, F.; Leprince, S.; Avouac, J.-P. Co-registration and correlation of aerial photographs for ground deformation measurements. *Isprs Journal of Photogrammetry and Remote Sensing* **2009**, *64*, 551–560, doi:10.1016/j.isprsjprs.2009.03.005.
 30. Becker, D.; Klonowski, J. Object Recognition of a GCP Design in UAS Imagery Using Deep Learning and Image Processing—Proof of Concept Study. *Drones* **2023**, *7*, doi:10.3390/drones7020094.
 31. Janßen, J.; Medic, T.; Kuhlmann, H.; Holst, C. Decreasing the Uncertainty of the Target Center Estimation at Terrestrial Laser Scanning by Choosing the Best Algorithm and by Improving the Target Design. *Remote Sensing* **2019**, *11*, doi:10.3390/rs11070845.
 32. Daramouskas, I.; Meimetis, D.; Patrinooulou, N.; Lappas, V.; Kostopoulos, V.; Kapoulas, V. Camera-Based Local and Global Target Detection, Tracking, and Localization Techniques for UAVs. *Machines* **2023**, *11*, doi:10.3390/machines11020315.
 33. Jung, H.-K.; Choi, G.-S. Improved YOLOv5: Efficient Object Detection Using Drone Images under Various Conditions. *Applied Sciences* **2022**, *12*, doi:10.3390/app12147255.
 34. MVTec Software GmbH. Available online: www.mvtec.com (accessed on 6 July 2023).
 35. Steger, C.; Ulrich, M.; Wiedemann, C. *Machine Vision Algorithms and Applications: Second Completely Revised and Enlarged Edition*; Wiley - VCH, 2007.
 36. Zschiesche, K.; Fitzke, M.; Schlüter, M. Self-Calibration and Crosshair Tracking with Modular Digital Imaging Total Station. *PFG – Journal of Photogrammetry, Remote Sensing and Geoinformation Science* **2022**, doi:10.1007/s41064-022-00220-0.
 37. Bethmann, F.; Luhmann, T. Least-squares Matching with Advanced Geometric Transformation Models. *Photogrammetrie - Fernerkundung - Geoinformation* **2011**, *2011*, 57–69, doi:10.1127/1432-8364/2011/0073.
 38. MVTec Software GmbH. *Solution Guide I*; München, Germany.
 39. Padilla, R.; Netto, S.L.; Da Silva, E.A.B. A Survey on Performance Metrics for Object-Detection Algorithms. In *2020 International Conference on Systems, Signals and*

- Image Processing (IWSSIP)*, Niterói, Rio de Janeiro, Brazil, 1 – 3 July; IEEE, 2020; pp 237–242, ISBN 978-1-7281-7540-9.
40. Sokolova, M.; Japkowicz, N.; Szpakowicz, S. Beyond Accuracy, F-Score and ROC: A Family of Discriminant Measures for Performance Evaluation. In *AI 2006: Advances in Artificial Intelligence*, Hobart, Australia, 4. - 8. December; Sattar, A., Kang, B., Eds.; Springer Berlin Heidelberg: Berlin, Heidelberg, 2006; pp 1015–1021, ISBN 978-3-540-49788-2.
 41. Novaković, J.D.; Veljovic, A.; Ilić, S.S.; Papic, Ž.M.; Tomović, M. Evaluation of Classification Models in Machine Learning. In *Theory and Applications of Mathematics & Computer Science*; Aurel Vlaicu University of Arad Editing House, Ed., 2017; pp 39–46.
 42. MVTec Software GmbH. *Reference Manual HALCON/HDevelop: 20.05.0.3*, 2021.
 43. Wenzel, H.; Pichler, Dieter. *Ambient Vibration Monitoring*; John Wiley & Sons Ltd: West Sussex, England, 2005, ISBN ISBN-13 978-0-470-02430-0 (HB).
 44. Wenzel, H. *Health monitoring of bridges*; Wiley: Chichester U.K., 2009, ISBN 978-0-470-03173-5.
 45. Federal Ministry of Transport and Digital Infrastructure. *Bericht „Stand der Ertüchtigung von Straßenbrücken der Bundesfernstraßen“: Vorlage an den Ausschuss für Verkehr und digitale Infrastruktur des Deutschen Bundestages*, 2016.
 46. Morgese, M.; Ansari, F.; Domaneschi, M.; Cimellaro, G.P. Post-collapse analysis of Morandi's Polcevera viaduct in Genoa Italy. *Journal of Civil Structural Health Monitoring*, 2020; pp 69–85.
 47. National Transportation Safety Board. *Pedestrian Bridge Collapse Over SW 8th Street: Accident Report*. Highway Accident Report NTSB/HAR-19/02; Washington, DC., 2019.
 48. Sawo, F.; Kempkens, E. Model-based approaches for sensor data monitoring for smart bridges. In *2016 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI)*, Baden-Baden, Germany, 19. - 21. September; IEEE, Ed., 2016; pp 347–352.
 49. Sabato, A.; Niezrecki, C.; Fortino, G. Wireless MEMS-Based Accelerometer Sensor Boards for Structural Vibration Monitoring: A Review. *IEEE Sensors Journal* **2017**, *17*, 226–235, doi:10.1109/JSEN.2016.2630008.
 50. Gentile, C.; Ruccolo, A.; Canali, F. Continuous monitoring of the Milan Cathedral: dynamic characteristics and vibration-based SHM. *Journal of Civil Structural Health Monitoring*, 2019; pp 671–688.
 51. Gnewuch, H.; Smeu, E.; Jackson, D.A.; Podoleanu, A. Long range extensometer for civil structure monitoring using fibre Bragg gratings. *Measurement Science and Technology* **2005**, *16*, 2005–2010, doi:10.1088/0957-0233/16/10/016.
 52. Minehane, M.J.; O'Donovan, R.; Ruane, K.D.; O'Keeffe, B. The use of 3D laser scanning technology for bridge inspection and assessment. *Structural Health Monitoring (SHM)* **2014**, *13*, 14.

53. Schill, F. Überwachung von Tragwerken mit Profilsclannern. PhD thesis; Technische Universität Darmstadt, Darmstadt, Germany, 2018.
54. Marendić, A.; Paar, R.; Duvnjak, I.; Buterin, A. Determination of dynamic displacements of the roof of sports hall Arena Zagreb. In *INGEO 2014: Proceedings of the 6th International Conference on Engineering Surveying*. INGEO 2014 – 6th International Conference on Engineering Surveying, Prague, Czech Republic, 3. - 4. April; Czech Technical University, Faculty of Civil Engineering, 2014, 2014, ISBN 8001054691, 9788001054697.
55. Stempfhuber, W. Verification of the Trimble universal total station (uts) performance for kinematic applications. In *Optical 3-D measurement techniques*. Conference on Optical 3-D Measurement Techniques IX, Vienna, Austria, 1. - 3. July; Grün/Kahmen, Ed., 2009; 211 - 221.
56. Biliszczyk, J.; Hawryszków, P.; Teichgraber, M. SHM System and a FEM Model-Based Force Analysis Assessment in Stay Cables. *Sensors (Basel)* **2021**, *21*, doi:10.3390/s21061927.
57. Wagner, A.; Wasmeier, P.; Reith, C.; Wunderlich, T. Bridge Monitoring by Means of Video-Tacheometer – A Case Study. *avn - Allgemeine Vermessungs-Nachrichten* **2013**, 283–292.
58. Guillaume, S.; Clerc, J.; Leyder, C.; Ray, J.; Kistler, M. Contribution of the Image-Assisted Theodolite System QDaedalus to Geodetic Static and Dynamic Deformation Monitoring. In *Conference and Seminar Proceedings: 3rd Joint International Symposium on Deformation Monitoring (JISDM)*. 3rd Joint International Symposium on Deformation Monitoring (JISDM), Wien, Austria, 30. March - 1. April; International Federation of Surveyors, FIG: Copenhagen, 2016; pp 66–75.
59. Page, M.J.; McKenzie, J.E.; Bossuyt, P.M.; Boutron, I.; Hoffmann, T.C.; Mulrow, C.D.; Shamseer, L.; Tetzlaff, J.M.; Akl, E.A.; Brennan, S.E.; et al. *The PRISMA 2020 statement: an updated guideline for reporting systematic reviews*, 2021. *BMJ*, 372.
60. Page, M.J.; Moher, D.; Bossuyt, P.M.; Boutron, I.; Hoffmann, T.C.; Mulrow, C.D.; Shamseer, L.; Tetzlaff, J.M.; Akl, E.A.; Brennan, S.E.; et al. PRISMA 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews. *BMJ* **2021**, 372, doi:10.1136/bmj.n160.
61. Förstner, W.; Wrobel, B.P. *Photogrammetric Computer Vision: Statistics, Geometry, Orientation and Reconstruction*; Springer International Publishing: Cham, 2016.
62. Topcon Corporation. *GTL-1000 English Brochure: Datasheet*, 2020. Available online: https://www.topcon.co.jp/en/positioning/products/pdf/GTL-1000_E.pdf (accessed on 8 June 2021).
63. Trimble Inc. *Trimble S9/S9HP Total Station: Datasheet*. Available online: <https://geospatial.trimble.com/products-and-solutions/trimble-s9-hp> (accessed on 16 August 2021).
64. Trimble Inc. *Trimble SX12: Datasheet*. Available online: <https://geospatial.trimble.com/SX12> (accessed on 26 August 2021).
65. Wagner, A. New Geodetic Monitoring Approaches using Image Assisted Total Stations. PhD thesis; Technische Universität München, München, Germany, 2017.

66. Wagner, A.; Wiedemann, W.; Wasmeier, P.; Wunderlich, T. Monitoring Concepts using Image Assisted Total Stations. In *SIG 2016*. SIG 2016 – International Symposium on Engineering Geodesy, Varaždin, Croatia, 20. - 22. May; Paar, R., Marendić, A., Zrinjski, M., Eds.; Croatian Geodetic Society, 2016, 2016, ISBN 953590180X, 9789535901808.
67. Guillaume, S.; Bürki, B.; Griffet, S.; Mainaud Durand, H. QDaedalus : Augmentation of Total Stations by CCD Sensor for Automated Contactless High-Precision Metrology. In *FIG Working Week 2012*, Rome, Italy, 6. - 10. May, 2012.
68. Charalampous, E.; Psimoulis, P.; Guillaume, S.; Spiridonakos, M.; Klis, R.; Bürki, B.; Rothacher, M.; Chatzi, E.; Luchsinger, R.; Feltrin, G. Measuring sub-mm structural displacements using QDaedalus: a digital clip-on measuring system developed for total stations. *Applied Geomatics* **2014**, 7, doi:10.1007/s12518-014-0150-z.
69. Hauth, S.; Schlüter, M.; Thiery, F. Schneller und ausdauernder als das menschliche Auge: Modulare Okularkameras am Motortachymeter. *avn - Allgemeine Vermessungsnachrichten*; VDE VERLAG GmbH, Berlin - Offenbach, 2013; 210 - 216.
70. Atorf, P.; Heidelberg, A.; Schlüter, M.; Zschiesche, K. Berührungslose Positionsbestimmung von spiegelnden Kugeln mit Methoden des maschinellen Sehens. *zfv – Zeitschrift für Geodäsie, Geoinformation und Landmanagement* **2019**, 144., 317–322, doi:10.12902/zfv-0268-2019.
71. Walser, B.H. Development and calibration of an image assisted total station. PhD thesis; ETH Zurich, Zurich, Switzerland, 2004.
72. Reiterer, A.; Wagner, A. System Considerations of an Image Assisted Total Station – Evaluation and Assessment. *avn - Allgemeine Vermessungs-Nachrichten* **2012**, 83–94.
73. Wagner, A.; Huber, B.; Wiedemann, W.; Paar, G. Long-Range Geo-Monitoring Using Image Assisted Total Stations. *Journal of Applied Geodesy* **2014**, 8, doi:10.1515/jag-2014-0014.
74. Wasmeier, P. Grundlagen der Deformationsbestimmung mit Messdaten bildgebender Tachymeter. PhD thesis; Technische Universität München, München, Germany, 2009.
75. Heppe, M. Erweiterung der Multistation Leica MS50 zum Modularen Digitalkameratachymeter. Bachelorthesis; University of Applied Sciences Mainz, Mainz, Germany, 2015 (not published).
76. Rottensteiner, F. Semi-automatic extraction of buildings based on hybrid adjustment using 3D surface models and management of building data in a TIS. Zugl.: Wien, Techn. Univ., PhD thesis, Wien, Austria, 2001.
77. Wagner, A.; Wasmeier, P. Flächen- und Feature-basiertes Monitoring mit Videotachymetern. In *Multi-Sensor-Systeme – Bewegte Zukunftsfelder*. 138. DVW-Seminar, Hamburg, Germany, 18. - 19. September; DVW – Gesellschaft für Geodäsie, Geoinformation und Landmanagement e.V., Ed.; Wißner-Verlag, 2014; pp 75–88, ISBN 3896399950.
78. Weinhuber, A.; Wiedemann, W.; Wasmeier, P.; Wunderlich, T. No second Chance– Monitoring destructive Bridge Load Tests. In *Contributions to International*

- Conferences on Engineering Surveying; INGEO&SIG 2020*. 8th INGEO International Conference on Engineering Surveying and 4th SIG Symposium on Engineering Geodesy, Online Event, 22. - 23. October; Kopáček, A., Kyrinovič, P., Erdélyi, J., Paar, R., Marendić, A., Eds., 2020; pp 29–38.
79. Reiterer, A.; Wasmeier, P. Anwendungsfelder bildgebender Tachymetrie – Herausforderungen und Perspektiven. In *Terrestrisches Laserscanning – TLS 2011 mit TLS-Challenge*, Beiträge zum 106. DVW-Seminar, Fulda, Germany, 1. - 2. December; Wunderlich, T., Ohlmann-Lauber, J., Eds.; Wißner-Verlag, 2011; pp 189–202, ISBN 978-3-89639-838-3.
 80. Ehrhart, M.; Lienhart, W. Image-based dynamic deformation monitoring of civil engineering structures from long ranges. In *Image Processing: Machine Vision Applications VIII*; Lam, E.Y., Niel, K.S., Eds.; SPIE, 2015; pp 139–152.
 81. Ulrich, M.; Follmann, P.; Neudeck, J.-H. A comparison of shape-based matching with deep-learning-based object detection. *tm - Technisches Messen* **2019**, *86*, 685–698, doi:10.1515/teme-2019-0076.
 82. Ma, J.; Jiang, X.; Fan, A.; Jiang, J.; Yan, J. Image Matching from Handcrafted to Deep Features: A Survey. *Int J Comput Vis* **2021**, *129*, 23–79, doi:10.1007/s11263-020-01359-2.
 83. Reiterer, A. A Knowledge-Based Decision System for an On-Line Video-Theodolite-Based Multisensor System. PhD thesis; Technische Universität Wien, Wien, 2004.
 84. Grimm, D.; Kleemaier, G.; Zogg, H.-M. *ATRplus White paper*, 2015. Available online: https://leica-geosystems.com/-/media/files/leicageosystems/products/white-papers/atrplus_wp.ashx?la=en&hash=91E09F6A1CF79DE10D9B742391F7B35F (accessed on 10 October 2021).
 85. Topcon Corporation. *X-pointing Technology White Paper*; Livermore, California, USA, 2013.
 86. Wagner, A. A new approach for geo-monitoring using modern total stations and RGB + D images. *Measurement* **2016**, *82*, 64–74, doi:10.1016/j.measurement.2015.12.025.
 87. Marendić, A.; Paar, R.; Damjanovic, D. Measurement of bridge dynamic displacements and natural frequencies by RTS. *JCE* **2017**, *69*, 281–294, doi:10.14256/JCE.1804.2016.
 88. Neitzel, F.; Schwarz, W. Schwingungsuntersuchungen. In *Ingenieurgeodäsie: Handbuch der Geodäsie, herausgegeben von Willi Freeden und Reiner Rummel*; Schwarz, W., Ed.; Springer-Verlag GmbH Deutschland, 2017; pp 463–506, ISBN 978-3-662-47188-3.
 89. Omidalizarandi, M.; Kargoll, B.; Paffenholz, J.-A.; Neumann, I. Accurate vision-based displacement and vibration analysis of bridge structures by means of an image-assisted total station. *Advances in Mechanical Engineering*, 2018; 168781401878005.
 90. Ehrhart, M.; Kalenjuc, S.; Lienhart, W. Monitoring of bridge vibrations with image-assisted total stations. In *SMAR 2017 - Fourth Conference on Smart Monitoring*,

Assessment and Rehabilitation of Civil Structures, Zürich, Switzerland, 13. - 15-September, 2017, ISBN 978-3-905594-66-9.

91. Lienhart, W.; Ehrhart, M.; Grick, M. High frequent total station measurements for the monitoring of bridge vibrations. *Journal of Applied Geodesy* **2017**, *11*, 1–8, doi:10.1515/jag-2016-0028.
92. Omidalizarandi, M.; Neumann, I.; Kemkes, E.; Kargoll, B.; Diener, D.; Ruffer, J.; Paffenholz, J.-A. Mems based bridge monitoring supported by image-assisted total station. In *Remote Sensing and Spatial Information Sciences. GeoSpatial Conference 2019 – Joint Conferences of SMPR and GI Research*, Karaj, Iran, 12. - 14. October; The International Archives of the Photogrammetry, Ed., 2019; pp 833–842.
93. Wiedemann, W.; Wagner, A.; Wasmeier, P.; Wunderlich, T. Monitoring mit scannenden bildgebenden Tachymetern. In *Terrestrisches Laserscanning 2017(TLS 2017)*; DVW – Gesellschaft für Geodäsie, Geoinformation und Landmanagement e. V., Ed.; Wißner-Verlag: Augsburg, 2017; pp 31–44, ISBN 978-3-95786-145-0.
94. Reiterer, A.; Lehmann, M. An Image-Based Measurement System for Quality Control in Civil Engineering. In *5th International Symposium on Image and Signal Processing and Analysis*. ISPA 2007, Istanbul, Turkey, 27. - 29. September; IEEE, Ed., 2007; pp 113–118.
95. Reiterer, A. Object monitoring/reconstruction with automated image-based measurement systems. In *3rd IAG Symposium on Geodesy for Geotechnical and Structural Engineering and 12th FIG Symposium on Deformation Measurements*, Baden, Austria, 22. - 24. May, 2006.
96. Reiterer, A.; Lehmann, M.; Miljanovic, M.; Ali, H.; Paar, G.; Egly, U.; Eiter, T.; Kahmen, H. A 3D optical deformation measurement system supported by knowledge-based and learning techniques. *Journal of Applied Geodesy* **2009**, *3*, 1–13, doi:10.1515/JAG.2009.001.
97. Zschesche, K.; Rau, L.; Schlüter, M. Optische Schwingungsmessungen: Status, Integration, Pros und Contras. In *GeoMonitoring 2020*. GeoMonitoring 2020, Braunschweig, Germany, 12 - 13 March; Institutionelles Repositorium der Leibniz Universität Hannover, Ed., 2020; pp 49–63.
98. Staiger, L.; Werner, Y.; Ziegler, S. Structural Health Monitoring mit Modularem Digitalkameratachymeter. Compulsory master project work; University of Applied Sciences Mainz, Mainz, Germany, 2021 (not published).
99. Zschesche, K. Image Assisted Total Stations for Structural Health Monitoring—A Review. *Geomatics* **2022**, *2*, 1–16, doi:10.3390/geomatics2010001.
100. Grimm, D.; Zogg, H.-M. *Leica Nova MS50 White paper*; Leica Geosystems AG: Heerbrugg, Switzerland, 2013.
101. Huang, Y.D.; Harley I. Calibration of Close-Range Photogrammetric Stations Using a Free Network Bundle Adjustment. In *Proceedings of first Conference on Optical 3-D Measurement Techniques*; Gruen and Kahmen, Ed.; Wichmann Verlag: Vienna, Austria, 1989; pp 49–56.

102. International Organization for Standardization. *Optics and Optical Instruments - Field Procedures for Testing Geodetic and Surveying Instruments - Part 3: Theodolites*, 2001, 17.180.30 (ISO 17123-3).
103. Steger, C. Extraction of curved lines from images. In *Proceedings of 13th International Conference on Pattern Recognition*, Vienna, Austria, 25. - 29. August; IEEE, Ed., 1996; 251-255 vol.2, ISBN 0-8186-7282-X.
104. Steger, C. An unbiased detector of curvilinear structures. *IEEE Trans. Pattern Anal. Machine Intell.* **1998**, 20, 113–125, doi:10.1109/34.659930.
105. Steger, C. Extracting curvilinear structures: A differential geometric approach. In *Computer Vision – ECCV '96*, Cambridge, UK, 14. - 18. April; Buxton, B., Cipolla, R., Eds.; Springer Berlin Heidelberg: Berlin, Heidelberg, 1996; pp 630–641, ISBN 978-3-540-49949-7.
106. Canny, J.F. Finding Edges and Lines in Images. Master's thesis; Massachusetts Institute of Technology, 1983.
107. Haralick, R.M.; Shapiro, L.G. *Computer and Robot Vision: Vol. 2*. Chapter 14.9; Addison-Wesley Longman Publishing Co., Inc: USA, 1992, ISBN 0201569434.
108. Suesse, H.; Voss, K. Adaptive Ausgleichsrechnung und Ausreißerproblematik für die digitale Bildverarbeitung. *Mustererkennung*, 1993; pp 600–607.
109. Kampmann, G.; Renner, B. Vergleich verschiedener Methoden zur Bestimmung ausgleichender Ebenen und Geraden. *avn - Allgemeine Vermessungsnachrichten*; VDE VERLAG GmbH, Berlin - Offenbach: Offenbach, 2004; pp 56–67.
110. Leica Geosystems AG. *TS30 Technical Data*; Switzerland, 2009. Available online: https://leica-geosystems.com/sftp/files/archived-files/TS30_Technical_Data_en.pdf (accessed on 4 February 2022).
111. IDS Imaging Development Systems GmbH. *UI-3250ML-M-GL: Datasheet*, 2015.
112. Lachinger, S.; Vorwagner, A.; Reiterer, M.; Fink, J.; Bruschetini-Ambro, S.-Z. Entwicklung eines neuen Regelwerkes für dynamische Messungen von Eisenbahnbrücken der ÖBB. *Baudynamik*, 2022; pp 53–66.
113. Bruschetini-Ambro, S.-Z.; Fink, J.; Lachinger, S.; Reiterer, M. Ermittlung der dynamischen Kennwerte von Eisenbahnbrücken unter Anwendung von unterschiedlichen Schwingungsanregungsmethoden. *Bauingenieur*, 2017; 2 -13.
114. IDS Imaging Development Systems GmbH. *UI-3080CP-M-GL: Datasheet*, 2016.
115. Grimm, D.; Hornung, U. Leica ATRplus – Leistungssteigerung der automatischen Messung und Verfolgung von Prismen. *avn*; VDE VERLAG GmbH, Berlin - Offenbach, 2015; pp 269–276.
116. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, 521, 436–444, doi:10.1038/nature14539.
117. Papp, S.; Weidinger, W.; Munro, K.; Ortner, B.; Cadonna, A.; Langs, G.; Licandro, R.; Meir-Huber, M.; Nikolić, D.; Toth, Z.; et al. *Handbuch Data Science und KI: Mit Machine Learning und Datenanalyse Wert aus Daten generieren*, 2.th ed.; Carl Hanser Verlag München, 2022, ISBN 978-3-446-46947-1.

118. MVTec Software GmbH. *HALCON Operator-Referenz 22.05.0.0*, 2022.
119. Zschesche, K.; Schlüter, M. Machine Learning for Classification and Detection in Image Assisted Total Station Applications. In *Ingenieurvermessung 23*. 20. Internationalen Ingenieurvermessungskurs, Zurich, Switzerland, 11. - 14. April; Andreas Wieser, Ed.; Wichmann, 2023; pp 249–261.
120. LeCun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-based learning applied to document recognition. *Proceedings of the IEEE* **1998**, *86*, 2278–2324, doi:10.1109/5.726791.
121. Wilson, D.R.; Martinez, T.R. The need for small learning rates on large problems. In *Proceedings of the IEEE International Joint Conference on Neural Networks IJCNN'1*, Washington, DC, USA, 15. - 19. July; IEEE, 2001; pp 115–119, ISBN 0780370449.
122. Xu, H.; Su, X.; Wang, Y.; Cai, H.; Cui, K.; Chen, X. Automatic Bridge Crack Detection Using a Convolutional Neural Network. *Applied Sciences* **2019**, *9*, doi:10.3390/app9142867.
123. Selvaraju, R.R.; Cogswell, M.; Das, A.; Vedantam, R.; Parikh, D.; Batra, D. Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. In *2017 IEEE International Conference on Computer Vision (ICCV)*, Venice, Italy, 22. - 29. October; IEEE, Ed., 2017; pp 618–626.
124. Prince, S.J. *Computer Vision: Models, Learning and Inference*; Cambridge University Press, 2012.
125. Schwieger, V.; Kerekes, G.; Lerke, O. Image-based target detection and tracking using image-assisted robotic total stations. In *Machine vision and navigation*; Sergiyenko O., Flores-Fuentes W., Mercorelli P., Ed.; Springer, Cham, 2020; pp 133–169, ISBN 978-3-030-22587-2.
126. Bengio, Y.; Bastien, F.; Bergeron, A.; Boulanger-Lewandowski, N.; Breuel, T.M.; Chherawala, Y.; Cissé, M.; Côté, M.; Erhan, D.; Eustache, J.; et al. Deep Learners Benefit More from Out-of-Distribution Examples. In *AISTATS*, 2011; pp 164–172.
127. DeVries, T.; Taylor, G.W. Improved Regularization of Convolutional Neural Networks with Cutout. *arXiv preprint arXiv:1708.04552* **2017**.
128. Zhong, Z.; Zheng, L.; Kang, G.; Li, S.; Yang, Y. Random Erasing Data Augmentation. *Proceedings of the AAAI Conference on Artificial Intelligence* **2020**, *34*, 13001–13008, doi:10.1609/aaai.v34i07.7000.
129. Leica Geosystems AG. *Leica CS35: Datasheet*; Heerbrugg, Schweiz, 2020. Available online: <https://leica-geosystems.com/de-de/products/total-stations/controllers/leica-cs35> (accessed on 16 May 2023).
130. Trimble Inc. *Trimble TSC7 Controller: Datasheet*. Available online: <https://de.geospatial.trimble.com/products-and-solutions/trimble-tsc7> (accessed on 16 May 2023).
131. Verma, R.; Ali, J. A Comparative Study of Various Types of Image Noise and Efficient Noise Removal Techniques. *International Journal of Advanced Research in Computer Science and Software Engineering* **2013**, 617–622.

132. Liu, B.; Liu, J. Overview of Image Denoising Based on Deep Learning. *Journal of Physics: Conference Series* **2019**, *1176*, 22010, doi:10.1088/1742-6596/1176/2/022010.
133. Musella, C.; Serra, M.; Menna, C.; Asprone, D. BIM & AI: advanced technologies for the digitalisation of seismic damages in masonry buildings. In *International fib Symposium on Conceptual Design Of Structures*. International fib Symposium on Conceptual Design Of Structures, Madrid, Spain, 26. - 28. September, 2019; pp 1–8.
134. Jang, K.; Kim, J.-W.; Ju, K.-B.; An, Y.-K. Infrastructure BIM Platform for Lifecycle Management. *Applied Sciences* **2021**, *11*, 10310, doi:10.3390/app112110310.
135. Resnik, B. Anwendung von künstlichen neuronalen Netzen bei der Auswertung von Deformationsmessungen. *avn - Allgemeine Vermessungs-Nachrichten*, *2021*, 94–103.

List of Publications

Articles in peer-reviewed journals:

- I. Philipp Atorf, Arno Heidelberg, Martin Schlüter, Kira Zschiesche (2019) **Berührungslose Positionsbestimmung von spiegelnden Kugeln mit Methoden des maschinellen Sehens**, zfv - Zeitschrift für Geodäsie, Geoinformation und Landmanagement. Bd. 155. H. 5. Wißner-Verlag 2019 pp. 317 - 322
- II. Bastian Plaß, Kira Zschiesche, Tamer Altinbas, Daniel Karla, Linda Rau, Martin Schlüter (2020) **Künstliche Intelligenz als Strategie in der Ingenieurgeodäsie – erste Schritte im Bahnumfeld**, zfv - Zeitschrift für Geodäsie, Geoinformation und Landmanagement. Bd. 145. H. 4/2020. 2020 pp. 236 - 240
- III. Kira Zschiesche, Martin Schlüter, Linda Rau (2020) **BIM in der Praxis – Ansätze zur Integration von Structural Health Monitoring in ein Bestands-BIM**, DVW e. V.; Runder Tisch GIS e. V. (Hrsg). Leitfaden Geodäsie und BIM. Version 2.1. Augsburg: Wißner-Verlag 2020 pp. 166 - 167 (DVW-Merkblatt; 11 (2020))
- IV. Kira Zschiesche (2021) **Image Assisted Total Stations for Structural Health Monitoring—A Review**, Geomatics. Bd. 2. H. 1. MDPI AG 2021 pp. 1 - 16. DOI: 10.3390/geomatics2010001
- V. Kira Zschiesche, Marvin Fitzke, Martin Schlüter (2022) **Self-Calibration and Crosshair Tracking with Modular Digital Imaging Total Station**, Journal of Photogrammetry, Remote Sensing and Geoinformation Science (PFG). Bd. 90. H. 5. Springer-Verlag GmbH 2022 pp. 1 - 15. DOI: 10.1007/s41064-022-00220-0
- VI. Kira Zschiesche, Martin Schlüter (2023) **Classification and object detection with image assisted total station and machine learning**, Journal of Applied Geodesy. pp. 1 - 9. DOI: 10.1515/jag-2023-0011.

Conference papers/posters:

- VII. Kira Zschiesche, Martin Schlüter, Linda Rau (2020) **Optische Schwingungsmessungen: Status, Integration, Pros und Contras**, Tagungsband Geomonitoring 2020. Hannover. 2020 pp. 49 - 63

- VIII. Kira Zschiesche, Denise Becker (2021) **Application of deep learning for image-based pattern recognition in surveying engineering**, Posterbeitrag, AGIT Symposium & Expo 7. - 8. July 2021. Salzburg, Austria.
- IX. Zschiesche, Kira; Schlüter, Martin (2023): **Machine Learning for Classification and Detection in Image Assisted Total Station Applications**. In: Andreas Wieser (Hg.): Ingenieurvermessung '23. 20. Internationalen Ingenieurvermessungskurs. Ingenieurvermessung '23. Zurich, Swiss, 11. - 14. April 2023: Wichmann, pp. 249 - 261.

Supervised thesis/ project reports:

- Christian Veith (2019) **Photogrammetrische Kalibrierung koaxialer Tachymeterkameras – vergleichende Untersuchungen an der Leica Multistation TS60**, bachelor thesis, Hochschule Mainz (University of Applied Sciences), Supervising Prof. Dr.-Ing. Martin Schlüter
- Linda Rau (2019) **Integration der periodischen Zustandsüberwachung von Bauwerken (SHM) in die Bauwerksdatenmodellierung (BIM)**, master thesis, Hochschule Mainz (University of Applied Sciences), Supervising Prof. Dr.-Ing. Martin Schlüter
- Rebecca Maria Schmitt (2020) **Standardisierung der Feldprüfung von Mobile Mapping Systemen (MMS)**, bachelor thesis, Hochschule Mainz (University of Applied Sciences), Supervising Prof. Dr.-Ing. Martin Schlüter
- Victoria Glanc (2020) **Geometrische Überwachung von Brückenbauwerken mit 3D-Laserscanning**, master thesis, Hochschule Mainz (University of Applied Sciences), Supervising Prof. Dr.-Ing. Martin Schlüter
- Marvin Fitzke (2021) **Kontinuierliches Strichkreuztracking für Modulare Digitaltachymeter**, bachelor thesis, Hochschule Mainz (University of Applied Sciences), Supervising Prof. Dr.-Ing. Martin Schlüter
- Linda Staiger, Yannik Werner, Simon Ziegler (2021) **Structural Health Monitoring mit Modulen Digitalkameratachymeter (MoDiTa)**, project work, processing period: 01.09.2020 - 25.02.2021, Hochschule Mainz (University of Applied Sciences), Supervising Prof. Dr.-Ing. Martin Schlüter

- Simon Ziegler (2021) **Berührungslose Zustandserfassung von Brückenbauwerken auf der Basis von Referenzfahrten**, master thesis, Hochschule Mainz (University of Applied Sciences), Supervising Prof. Dr.-Ing. Martin Schlüter
- Michael Biele, Lukas Haas, Cedric Jager, Alexander Milch (2022) **Structural Health Monitoring (SHM) für Bestandsobjekte**, project work, processing period: 01.09.2021 - 28.02.2022, Hochschule Mainz (University of Applied Sciences), Supervising Prof. Dr.-Ing. Martin Schlüter
- Alexander Milch (2022) **Objektmonitoring mit dem Lasertracker**, master thesis, Hochschule Mainz (University of Applied Sciences), Supervising Prof. Dr.-Ing. Martin Schlüter
- Eric Bach, Marvin Fitzke, Darius Minor (2023) **SHM - Praxistauglichkeit von Messsystemen an Großbauwerken**, project work, processing period: 01.09.2022 - 28.02.2023, Hochschule Mainz (University of Applied Sciences), Supervising Prof. Dr.-Ing. Martin Schlüter