Graz University of Technology

DI (FH) Niko Benjamin Huber

Design and implementation of a rock slope monitoring system based on “Image Assisted Total Stations“

DOCTORAL THESIS

For obtaining the academic degree of
Doktor der technischen Wissenschaften

Doctoral Programme of Technical Sciences
Doctoral School Geosciences, Class of Geodesy

Graz University of Technology

Institute of Remote Sensing and Photogrammetry

Head: Univ.-Prof. Dr.rer.nat. Dipl.-Forstwirt Mathias Schardt

Supervisor: Ao.Univ.-Prof. Dipl.-Ing. Dr.techn. Viktor Kaufmann

Graz, October 2014

Statutory Declaration

I declare that I have authored this thesis independently, that I have not used other than the declared sources/resources, and that I have explicitly marked all material which has been quoted either literally or by content from the used sources.

Graz, ____________________________ ____________________________

Date Signature

Eidesstattliche Erklärung

Ich erkläre an Eides statt, dass ich die vorliegende Arbeit selbstständig verfasst, andere als die angegebenen Quellen/Hilfsmittel nicht benutzt, und die den benutzten Quellen wörtlich und inhaltlich entnommenen Stellen als solche kenntlich gemacht habe.

Graz, ____________________________ ____________________________

Date Signature

Abstract

The objective of this thesis is the development of a photogrammetric processing chain for oriented image bundles, acquired with a stereo set-up of "Image Assisted Total Stations" (IATS) to continuously measure rock slope deformation. The system supplies 3D deformation vectors which can be analyzed by experts or processed in geodetic deformation analysis software reducing risks to infrastructure and even live posed by rock fall events.

The developed processing chain was designed to automatically detect corresponding points in the respective stereo-images with sub-pixel accuracy, track them in consecutive measurement epochs and calculate their 3D position using a forward intersection. The used IATS devices need about five seconds to capture one stereopair of images with a ground sampling distance (GSD) of around 1 mm at a distance of 100 m covering a field of view of about 1.56 gon (2.45 m) in horizontal and 1.17 gon (1.84 m) in vertical direction. This constrains the minimum time frame available for data processing and evaluation for a system which permanently acquires data.

The proposed solution consists of a fully configurable modular system containing optimized processing steps to increase the matching accuracy, the number of detected points and to reduce the amount of mismatches. Test results of the individual processing chain components were evaluated using synthetic ground truth data. The whole system was assessed using data-sets from field experiments and compared to high-resolution terrestrial laser scans, and conventional theodolite measurements of signalized points. The accuracies lie in a similar range as manually executed theodolite based point measurements on signalized targets. However, these results can be retrieved automatically based on image texture and the detection of dozens or even hundreds of deformation vectors is only a matter of seconds. Deformation experiments using artificial and natural structures were carried to demonstrate the temporal tracking of deformations on individual points leading to 3D deformation vectors.

The IATS system combines the strength of a theodolite based measurement system in terms of highest accuracy with the high point density and automated processing of a terrestrial laser scanner. Furthermore, a statistical evaluation of each measured point is possible. Due to the fact that errors in an IATS system, set up with a shorter baseline to increase matching accuracy, mainly occur perpendicular to the base, the system also shows a high complementarity with TLS or polar distance measurements.

Acknowledgements

I would like to thank my supervisor Viktor Kaufmann for all his help and the detailed review of this thesis. His thorough revisions lead to significant improvements.

Special thanks go to Alexander Reiterer for giving me the opportunity to work in this interesting field and his continuous support until the end of this thesis.

Further, I want to thank Mathias Schardt as my prior supervisor and boss for his support and Ruth Hödl for the equally important great administrative help.

Great thanks also belong to my colleagues at the JOANNEUM RESEARCH and the Technical University of Munich, especially to Andreas Wagner for the great support and cooperation.

None of this would have been possible without the support from my family and friends. Especially Josi was of the utmost importance to my mental health with his charming stories about how working on a PhD can also look like. Thank you deeply for all the laughs my friend.

Last but not least I want to thank Birgit for sharing my life exactly during these many years of work on this thesis. We had a great time, I don’t regret anything.

I want to dedicate this thesis to my late grandma Reinlinde Huber, who kept asking me when I was going to be finished. I am sorry I didn’t make it in time for you to see Oma.

Contents

1 Introduction

Geo-hazards such as rock falls and landslides endanger, injure and kill a large number of people all around the world. Research conducted at the International Landslide Centre at Durham University, UK indicates that from 2004 to 2010, 2620 non-earthquake triggered landslide events killed a total of 32 322 people. However, this figure may still be too low as countries such as North Korea or Ethiopia with multiple fatal events each year only report a fraction of the total incidents. Events are likely to cause fatalities where the following factors come together: heavy rainfall, steep slopes and a dense population (Petley, 2012).

In addition to the cost of human lives a huge economic impact has to be considered. In the USA alone average annual damages of one to two billion USD occur (U.S. Geological Survey, 2013). Similar or higher figures are to be expected in all densely populated areas all around the world. Furthermore, ancillary economic costs caused by infrastructure failure and downtime also have to be taken into account.

Norway, to name a European example, registered more than 31 500 sites endangered by landslides or rock falls (Devoli, et al., 2011). Due to monitoring costs and inaccessible terrain only five of these sites are currently continuously monitored. One of these sites is the Åknes landslide illustrated in Figure 1. High-resolution long range remote monitoring could help to strongly decrease costs and risks to surveying personnel.

As rock falls cannot be avoided the mitigation of accompanying risks to human lives and infrastructure is most important. Knowledge about dynamics of unstable regions may in many cases allow the prediction of hazardous events (Poisel & Preh, 2004) early enough for life saving precautionary measures to be taken. This calls for monitoring methods capable of detecting deformations in high accuracy to gain an insight in the spatial distribution and expected dynamics of the event. These insights derived from monitoring data are crucial for a reliable risk assessment and consequent reactions such as focused, more expensive on-site monitoring, the application of protection installations, or in last consequence the evacuation of endangered population. In post disaster event monitoring, remote surveying may be the only possibility to assess the amount of mass movement, occurred damages or further hazards while at the same time reducing the risk for emergency staff and field personnel.

In this thesis the design and implementation of a high accuracy image-based remote monitoring system capable of dealing with the above mentioned issues is described and evaluated. The system is based on data acquisition from a stereo set-up of "Image Assisted Total Stations" (IATS) and subsequent image processing steps in order to derive 3D positions of un-signalized points on the object and to track them over multiple measurement epochs creating deformation vectors.

An IATS is a conventional tacheometer extended with an imaging sensor within the optical path of its telescope. Using the built-in robotic capabilities of the base instrument it is possible to capture a specific region of interest (ROI) through camera rotation. The data is represented as a bundle of images forming a panoramic mosaic. Using an exact calibration of the tacheometer and its imaging sensor the measured horizontal and vertical angles of rotation can be transferred into a coordinate system centered at the camera's center of projection. This allows the calculation of specific angle values for every position on the image. Matched homologous points in stereo-images can then be used to perform a spatial forward intersection to calculate their 3D position.

Figure 1: Continuously multisensor monitored rockslide, Åknes Norway (Kristensen, et al., 2010)

The author developed a modular photogrammetric processing chain for the processing of IATS image data from multiple measurement epochs calculating the above mentioned high-resolution deformation vectors.

The IATS devices used in this work are prototypes from Leica (Wasmeier, 2009) based on a modified total station TCRA1201 with an ocular camera with a resolution of 2560x1920 pixels and a field of view of about 1.56 gon in horizontal and 1.17 gon in vertical direction (for a focus position at ∞).

The work presented in this thesis originates from the FWF funded research project "i-MeaS – An Intelligent Image-Based Measurement System for Geo-Hazard Monitoring" (project number "Translational Research L514") lead by the Institute of Geodesy and Geophysics of Vienna University of Technology in which the author participated as project assistant over the course of two years and was further supported by the European Community's Seventh Framework Programme (FP7/2007-2013) research project DE-MONTES ("Deformation Monitoring by High-resolution Terrestrial Long Range Sensing" under grant agreement 285839) coordinated by the JOANNEUM RESEARCH, Graz in which the author participated as a key researcher.

1.1 Problem statement

The measurement system including its photogrammetric processing chain developed in this work is targeted at the task of continuous rock slope deformation measurement. These measurements are taken in consecutive scanning intervals; so called measurement epochs, to monitor rock walls which may be at risk of undergoing deformation. Knowledge about the kinematics[1] of unstable surfaces in an accuracy range of 2 - 20 mm of deformation per day (Figure 2) may allow a subsequent prediction of hazardous events (Kristensen, et al., 2010). Figure 2 illustrates a schematic diagram of the movements at the Norwegian Åknes rockslide derived from measurement data of previous events. Continuous monitoring and data analysis is crucial to identify the movement behaviour and velocity trends.

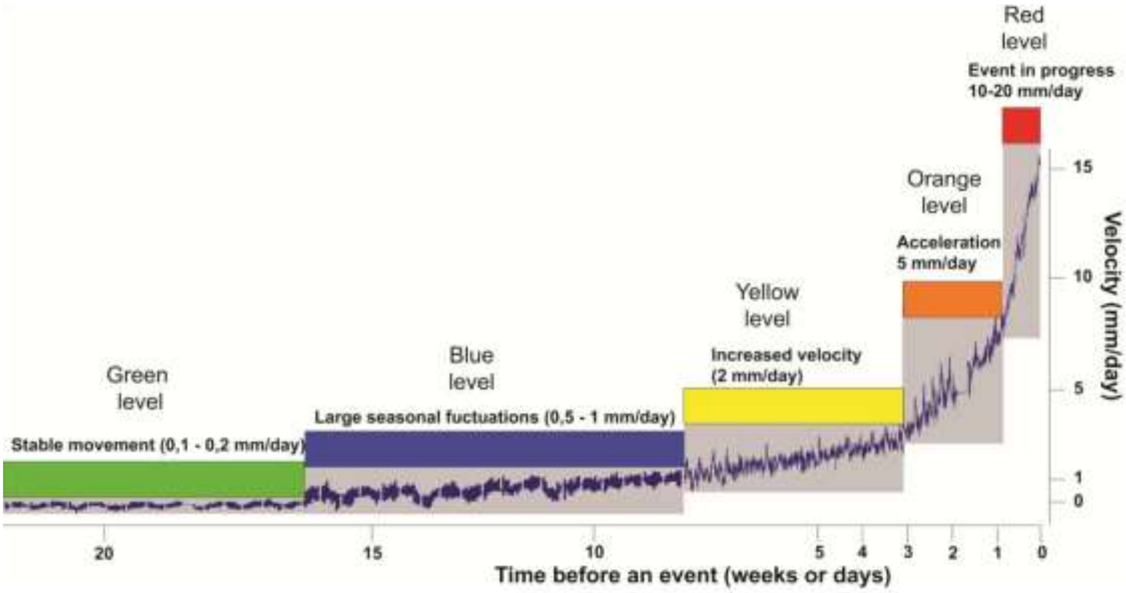

Figure 2: Velocity and acceleration as criteria for different alarm levels (Kristensen, et al., 2010)

Different extensive long term multisensory monitoring campaigns and evaluations of instable sites are documented in the literature. An exceptionally

---

[1] Description of motions without explanation of their source

well documented example is the analysis of the kinematics of the Gradenbach landslide based on photogrammetric measurements, GPS and seismic monitoring data by Brückl, et al. (2006). In Brückl, et al. (2013) the evaluations were extended by extensometer measurements, hydrostatic water levels in boreholes and a strain monitoring using fiber optic sensors to analyze underlying dynamic processes.

If a deformation of on object occurs, not only the geometrical changes are of interest but also the dynamics of the underlying process. This means that "geodetic deformation analysis" more and more means "geodetic analysis of dynamic processes" (Welsch & Heunecke, 2001). The classical geodetic deformation analysis as referred to in this thesis however, excludes the modelling of dynamic processes. This merely descriptive approach only evaluates the temporal development of the process (Welsch, et al., 2000).

In a classical geodetic deformation analysis the calculated 3D object point coordinates, reference point coordinates and raw angle measurements are evaluated and analyzed in a common network adjustment. Connected measurements from different temporal epochs, as provided by the analysis of oriented image measurements, may lead to the identification of unstable reference points, a set of points that have moved significantly and the according covariance information. This evaluation can be carried out in geodetic deformation analysis software such as GOCA developed at the University of Applied Sciences Karlsruhe (Kälber, et al., 1999). The interpretation of the results is carried out either manually by experts or semi-automated relying on knowledge based geo-risk assessment (Vicovac, et al., 2010) or fuzzy systems (Chmelina, et al., 2006). The retrieved information can aid in the reduction of risks to infrastructure and even live posed by rock fall events.

The measurement of rock slopes is traditionally done using monitoring techniques with a low data density such as TPS, DGPS or extensometer measurements leading to a point distance of about 50-100 meters (Abellán, et al., 2014). Another approach is a photogrammetric analysis offering far higher point density at the cost of lower accuracy. An alternative approach for rock slope monitoring is the application of laser scanning. To be more precise ground based terrestrial laser scanning (TLS) is used due to its higher point density and accuracy compared to the inadequate viewing angle from the air. The point distance on the object can be in cm range, the distance component can reach millimeter to centimeter accuracy. However, the system suffers from poor repeatability of single measurements and generally low temporal resolution. The lack of repeatable single measurements is made up for by the fact that whole scans in the form of TLS-DEMs can be overlaid and subtracted to detect 1D differences (Bauer, et al., 2005; Abellán, et al., 2010; Figure 3). Aryal, et al. (2012) developed a 2D comparison based on a sliding window correlation approach. Other techniques focus on the 3D alignment of different point clouds or sub-regions of those point clouds over multiple temporal epochs. This can be done by a manual or semi-automatic tracking of targets or common features or automatically for example using the ICP algorithm (Abellán, et al., 2014).

The evaluation of measured and illustrated patterns of displacements can help to increase the understanding of the failure mechanisms. Furthermore, small scale rockfall (Figure 3) is rarely detected by current techniques due to the lack of point density. A prediction of possible rock slope failures has to be based on indicative signs of precursory activity such as precursory rockfalls and small pre-failure deformation which may lie in the same order of magnitude as instrument errors and are therefore hard to detect using TLS (Abellán, et al., 2014).

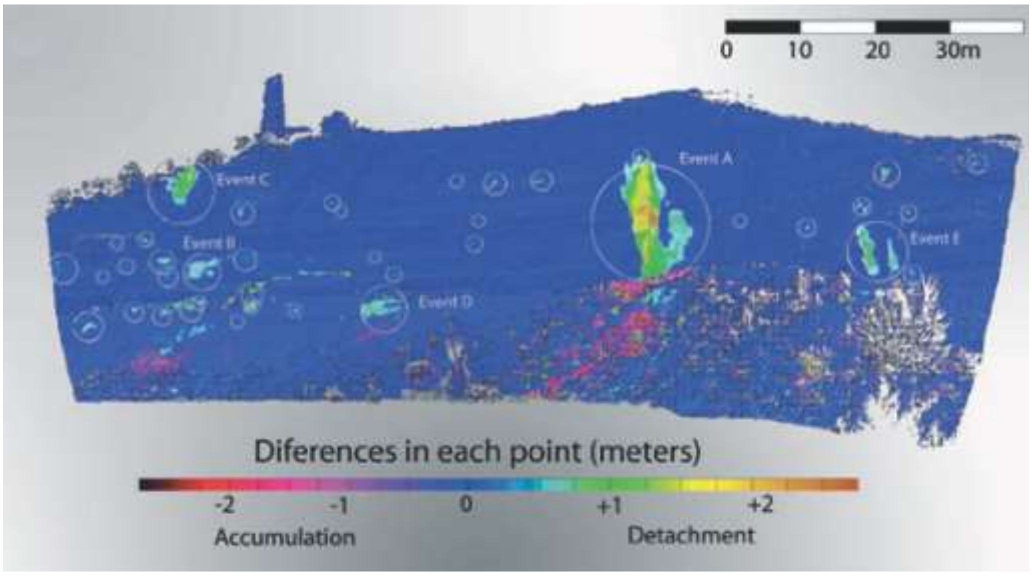

Figure 3: TLS point cloud comparison showing differences. Main rockfall events manually marked (Abellán, et al., 2010)

The proposed work targets the application of continuous rock slope monitoring in a high temporal cadence using a stereo set-up of IATS. Optimally data evaluation using the developed photogrammetric processing chain can be carried out in parallel at the same speed as the data acquisition. This approach tries to retain the high accuracy values that can be reached with theodolite measurements of signalized points while automatically measuring non-signalized points in high density using image based processing, and tracking them over multiple measurement epochs. An effort is undertaken to tackle the shortcomings of conventional TPS measurements such as the low point density and low level of possible automation as well as the shortcomings of TLS such as lower accuracy compared to TPS measurements, lower temporal resolution and poorly repeatable single point measurements. In addition, existing deformation analysis methods and software developed for the analysis of TPS measurements should be directly usable for a further evaluation of the produced results.

Due to the work's focus on high temporal resolution and continuous monitoring, the production of high-resolution dense point clouds is not part of this work although possible using the acquired data (Section 8.4.2.2). Compared to the already more than 10-15 years of history of TLS applied to this area of applications and the huge amount of research work and instrument expertise that has been accumulated, the proposed work does not aim at a direct competition with the established concepts but rather demonstrates new concepts and shows achievable accuracies and other parameters of the new system. The recent appearance of commercially available IATS makes the presented ideas even more relevant.

1.2 Objectives

The main objective of this thesis and therefore the main driver of the conducted research was the development of a photogrammetric processing chain for oriented image bundles acquired with an IATS stereo set-up.

The processing chain has to be able to produce homologous points in a single temporal epoch, temporally connect them to adjacent measurement epochs and calculate 3D deformation vectors.

In order to fulfill the objective of supplying temporally connected measurements multiple technical problems have to be solved with the highest possible degree of accuracy and the lowest possible computational runtime. The author proposes the following process.

- Detection of non-signalized points in the images
- Mutual matching of these points to detect correspondences in the highest possible accuracy considering computational constraints
- Tracking matched points over multiple measurement epochs
- Using the calibration parameters and the tracked stereo-correspondences to calculate 3D deformation vectors

One main side requirement to the developed system was to find a solution how to optimally balance the two competing requirements of accuracy and limited runtime. The latter is aggravated by the fact that the used processing chain should be able to run on standard field capable hardware.

A further objective is the determination of an optimal system stereo set-up. This again comes to a trade-off between the point matching requirement of having similar viewpoints for better matching results and the requirement of having an optimal base-to-distance ratio for the spatial forward intersection.

Also it had to be evaluated how the used approach performs in comparison to other state of the art remote sensing solutions like TPS or TLS measurements in terms of accuracy and runtime.

1.3 Structure

This thesis describes a modular photogrammetric processing chain using data from a prototype measurement system based on two stereo IATS. The work's structure shall allow a consecutive reading wherein each chapter is logically based on the information given in the sections before. However, it is also possible to read individual chapters which have their own introductions and conclusions making them self-sufficient parts of work.

Chapter 1 contains an introduction, the specification of objectives for this thesis and an outline of its structure.

Chapter 2 gives an overview of the relevant measurement devices. This includes an account of the background of IATS, a report on the hardware of the used Leica IATS prototypes, a description of the system calibration and a short overview of the used device control software.

Chapter 3 presents the system set-up and design. This section contains theoretical considerations of achievable accuracies as well as a system design review describing the data acquisition as well as the subsequent photogrammetric evaluation. Example application scenarios are presented from which key parameters regarding accuracy and system runtime are derived and linked to the successively described processing chain.

Chapter 4 shortly outlines image pre-processing methods that can be applied in order to increase the efficacy of further processing.

Chapter 5 describes applied image analysis methods for the detection and tracking of homologous points. As the photogrammetric analysis applied is based on the detection of non-signalized points in the images, the most suitable detector had to be found based on a review of state of the art methods. The same procedure was applied to the consecutive steps of feature description and matching.

Chapter 6 discusses the importance of the detection of mismatches in a set of correspondences. In case of sufficient available runtime the accuracies of the matches can be refined via a Least Squares Matching which is described in this chapter. Furthermore, strategies for temporal tracking of detected homologous points and the resulting calculation of deformation vectors are presented.

Chapter 7 contains a description of the implementation of the before described processing steps as well as evaluations of the algorithms' accuracy and runtime based on synthetic rendered ground truth data. To improve the general outline and readability of this thesis only a sub-set of exemplary evaluation charts are illustrated within this chapter. The remaining charts were placed in the Appendix Sections B to E.

Chapter 8 evaluates the developed measurement system using data acquired during field experiments in comparison with theodolite measurements on signalized targets.

Chapter 9 states the key findings and main conclusions and gives future prospects for the use of the developed system.

Chapter 10 lists the references; Chapter 11 contains a list of abbreviations appearing in the thesis and states their meaning.

The work is concluded with the Appendix containing tables from the evaluation of feature detection methods in Section A, figures for the visual evaluation of detected feature distribution in Section B, feature detector performance evaluation charts in Section C, feature descriptor performance evaluation charts in Section D and least squares refinement performance evaluation charts in Section E which are linked to the evaluations given in Chapter 7 as mentioned before. Appendix Section F and G comprise the network adjustment reports of field experiments whereas Section H contains the command line parameters of the feature detection, description and matching tool developed by the author.

2 The IATS measurement system

2.1 Introduction

The terms used to address total stations with differing levels of imaging support and capabilities range from "Imaging Total Station" (Leica Geosystems AG, 2010) or "Multistation" (Leica Geosystems AG, 2013a) to "Imaging Station" (Topcon, 2008). Based on Walser (2005) the term "Image Assisted Total Station" (IATS) will be used consistently throughout this thesis.

The terms measurement accuracy and precision are used in this thesis as defined in the International vocabulary of metrology (BIPM, 2008). Measurement precision is the agreement of independent measurements under stable and repeatable conditions. A high precision describes a low variation of measurement results. The deviation of the mean of a large set of measurements from a given reference value describes the trueness or accuracy of the mean. This value does not give information about the spread of the data. The measurement accuracy describes the agreement of a single measurement and a given reference value. In this context a measurement error is the measurement value minus a reference quantity. There are two components of a measurement error, the systematic error which remains constant or varies predictably in replicable measurements and the random error varying in an unpredictable manner.

The above mentioned IATS are a new kind of tacheometers offering in addition to 3D point measurements the acquisition of oriented images of high angular resolution showing the scene visible in the system's telescope. Mosaics of adjacent images can be captured with camera rotation executed by computer controlled motors moving the tacheometer's axes (Reiterer, et al., 2010).

Using an exact calibration of the tacheometer and its imaging sensor the measured horizontal and vertical angles of rotation can be transferred into a coordinate system centered at the camera's center of projection. This allows the calculation of specific angle values for every position on the image. This means that matched homologous points in stereo-images can be used to perform a spatial forward intersection to calculate the 3D position of this point.

One of the most interesting aspects of this fusion of photogrammetry and angular measurements using a total station is the possibility to measure non-signalized points in the images with higher accuracy than possible for a manual operator. This can be achieved through sub-pixel point detection and matching. The instrument used in this work the IATS2 based on a Leica TCRA1201+ total station (Section 2.2) has an angle measurement accuracy of 0.3 mgon (Leica Geosystems AG, 2007a), the built-in camera has a resolution of around 0.6 mgon per pixel.

This technology poses multiple new possibilities like the detection of non-signalized points in areas of rich texture within the captured images in a high density in seconds. The positions of certain geometrical patterns such as drill holes can be automatically detected or refined from approximate values. Signalized points in the form of specific target patterns or prisms can also be found and precisely localized. Furthermore, a direct image-based point to point correspondence in a stereo set-up or over measurement epochs can be established automatically.

In all of the above stated cases the positioning accuracy is improved as points are defined by distinctive areas and rounded or linear features which can be used to calculate the center position as opposed to sighting the point directly with the center of the telescope (Wasmeier, 2009, p. 9).

2.1.1 Background and on-going research

The idea of using a combination of cameras and theodolites dates back to Albrecht Meydenbauer to the middle of the 19th century. He was the first to use the term Photogrammetry in 1867, in the same year he developed a combined camera and measurement system (Figure 4a), a first predecessor of devices that would later be called Phototheodolites (Albertz, 2001). Phototheodolites consisted of measuring cameras on a tripod which could be leveled and turned around vertical and horizontal axes with the possibility to read out angle values of the taken images. These could further be used for photogrammetric evaluations. Among the first manufacturers was the company Koppe which produced the phototheodolite shown in Figure 4b (Engler, et al., 1897). Over the years these devices became one of the most important instruments of terrestrial photogrammetry.

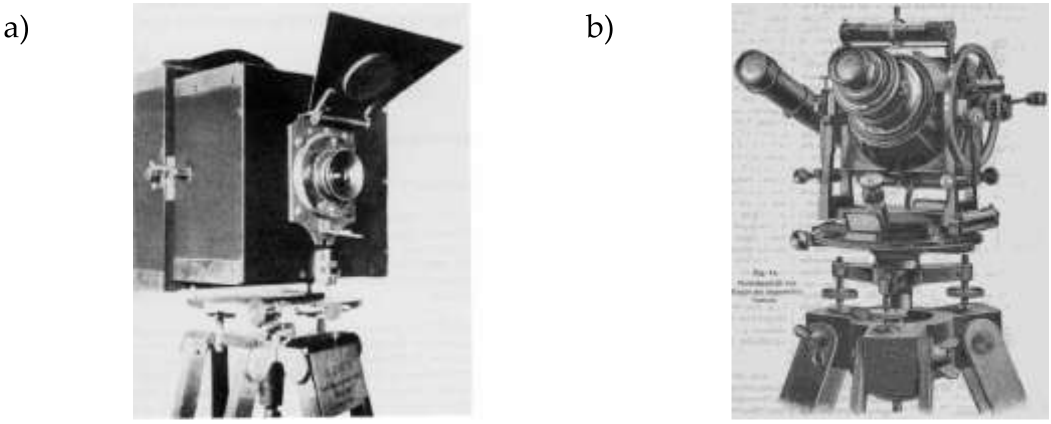

Figure 4: a) First photogrammetric camera built by Meydenbauer 1867 (Albertz, 2001), b) Phototheotolite by Koppe 1897 (Engler, et al., 1897)

With the emergence of electronic theodolites for industrial surveying purposes in the early 1980s the instruments started making use of computers to automatically calculate spatial forward intersections from stereo measured points. So called electronic triangulation systems were developed based on the principle of computational spatial forward intersection of angles measured from two or more instruments using automated aiming mechanisms.

Among the first of these systems were the Keuffel & Esser AIMS (Vyner & Hanold, 1982), the Zeiss IMS and the Brunson Electronic Triangulation System (BETS). The company Kern developed the industrial measurement Software ECDS which together with the image analysis software SPACE could be used to calculate forward intersections of angles measured with their instrument Kern E2-SE (Gottwald, 1987).

In order to increase automation even further autonomous motors where used to aim at targets (Kahmen & Steudel, 1988) and an automated recognition of targets with imaging sensors was developed (Haag, et al., 1997).

The main field of application for these systems were industrial high precision measurements. Due to the introduction of Laser-trackers by Leica in 1991 the imaging based systems were completely replaced within only a couple of years. Only after the invention and broad availability of small scale, reliable industrial digital cameras manufacturers of surveying instruments have again taken up the interest in image-assisted instruments in the beginning of the first decade of the 21st century. Most of these new systems used wide angle imaging components to create an overview and documentation of a surveyed area. A review of the different available image-based and image-assisted instruments on the market will be given in the next chapter.

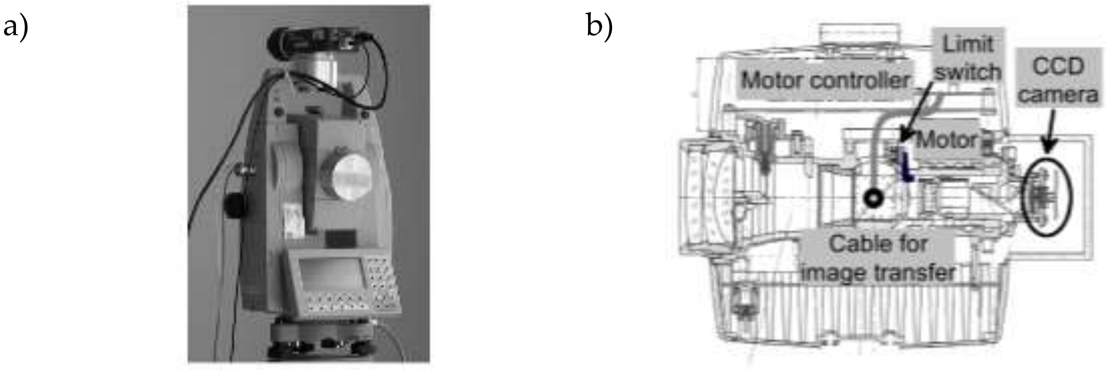

Figure 5: a) Modified TCRA 1101 total station with ocular and overview camera, b) Vertical cross-section of instrument (Walser, 2005)

Important for the current academic research are a prototype IATS system based on a Leica TCRA1101 total station placing a camera in the instruments ocular (Figure 5) and an enhanced small batch series of IATS2 prototypes based on a Leica TCRA1201 (Figure 6). This thesis is based on data captured with the latter of the two devices.

The TCRA1101 based prototype was developed by Bernd Walser in close cooperation with Leica Geosystems (Walser, 2005). Walser used an ocular camera with a resolution of 488 x 572 pixel and a field of view of around 1.3 x 1 gon. An overview camera with a field of view of 20 gon served as pointing helper. Walser developed an extensive procedure for system calibration and already performed some accuracy evaluations with template matching of signalized targets (Walser, 2005). This system was evaluated by Vicovac, 2008 and used in first tests of deformation measurements (Reiterer, et al., 2009). Based on the experience gained in the development of this first IATS, Leica Geosystem developed a small batch series of IATS2 based on their 2007 released TCRA1201+ R1000 total station. The new IATS provided a more stable image acquisition and a greatly enhanced image resolution of 2560 x 1920 pixel of the ocular camera, however, due to cost reasons the overview camera is missing in this device.

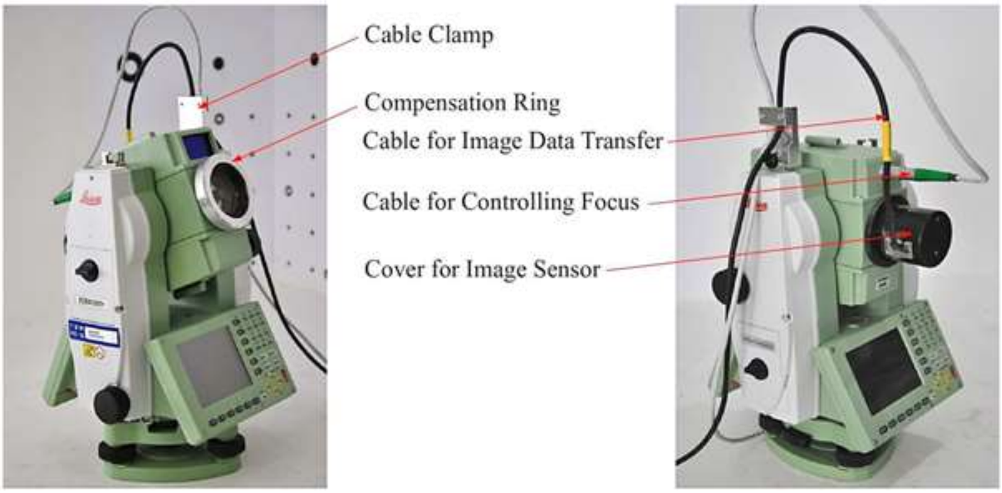

Figure 6: Leica IATS2 prototype based on Total Station TCRA1201+ (Wagner, et al., 2013b)

Research activities in the past years were dealing with a number of IATS related issues starting from the important system calibration (Walser, 2005; Vogel, 2006; Wasmeier, 2009) over to image pre-processing and first solutions for the detection of non-signalized points (Roic, 1996). Further work was dealing with manual point detection (Scherer, 2004) and reachable accuracies (Mischke & Kahmen, 1997; Wasmeier, 2009). Automated point detection was evaluated as well by Mischke & Kahmen (1997) and Reiterer (2004).

Juretzko (2005) developed a concept of a fully integrated system based on three cameras with different focal lengths, a wide angle camera for a proper documentation and overview images, a long-focus-lens-camera to identify details and an ocular camera for the actual measurements.

At the institute of Geodesy and Geophysics of the Vienna University of Technology multiple research projects were dedicated towards developing a stereo IATS based system for the purpose of geo- and deformation monitoring. Last of these was the project “i-MeaS – An Intelligent Image-Based Measurement System for Rock Fall Monitoring” in which the author worked as project assistant from the end of 2008 to the beginning of 2011, developing the foundations for this PhD thesis. The used measurement system was based on a stereo set-up of two modified Leica TCRA 1201 tacheometers (Reiterer, et al., 2010).

The above mentioned IATS2 system from Leica was also an integral part of the European Community's Seventh Framework Programme (FP7/2007-2013) research project DE-MONTES (“Deformation Monitoring by High-resolution Terrestrial Long Range Sensing” under grant agreement 285839) coordinated by the Joanneum Research in which the author participated as a key researcher.

This project aimed to combine the IATS high accuracy long range measurement capabilities with larger scale digital photogrammetry reconstructions.

Other academic developments try to develop their own IATS systems as add-on modules for standard total stations. One of these solutions is developed at the Geodesy and Geo-dynamics Lab (GGL) of the ETH Zurich. The system is called “Daedalus” and consists of multiple components which can be clipped onto regular total stations without any further changes to the measurement hardware (Figure 7).

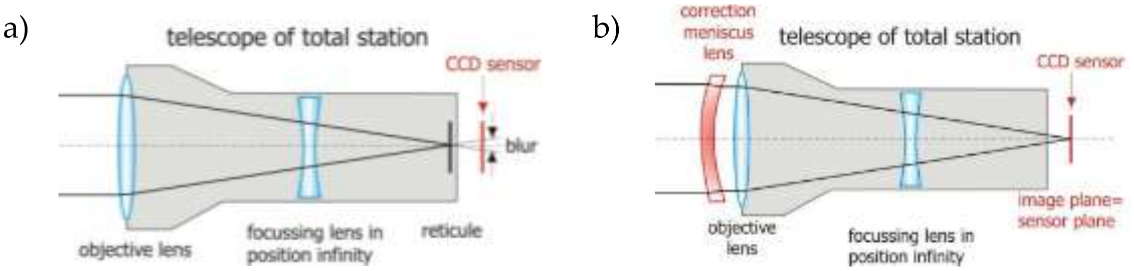

Figure 7: a) Original telescope b) Adjusted point of focus towards the appended CCD sensor plate behind the telescope using a correction lens (Guillaume, et al., 2012)

All adaptions to the hardware are made in a non-destructive way allowing a further use of the theodolite for standard surveying. The system works with a range of Leica total stations such as the models TCA 1800, TCA 2003 or TDA 5005. The camera including a cable based interface and a focus motor is clipped onto the viewing end of the telescope whereas a correction lens to shift the focal plane about four millimeters back onto the clipped-on projection plane of the sensor is attached to the front end of the telescope (Figure 8).

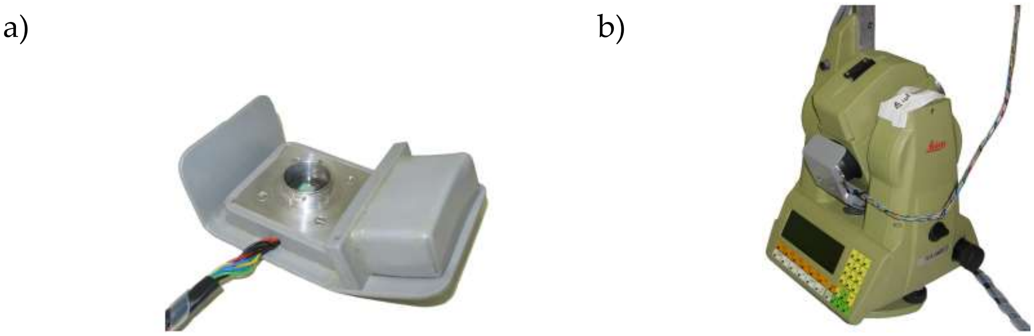

Figure 8: a) Clip-on industrial camera and correction lens b) Application on Leica TCA 1800, zenith measurements are still possible (Guillaume, et al., 2012)

The used industrial camera has a resolution of 1024 x 768 pixels and can capture and transfer 30 full frames per second. The angular resolution per pixel is around 1.2 mgon.

The system was developed mainly for the application of astro-geodesy but was further improved to be also applied in automated terrestrial and engineering surveying as well as deformation, vibration, and frequency analysis (Bürki, et al., 2010; Guillaume, et al., 2012).

Another solution was developed by the "i3mainz - Institute for Spatial Information and Surveying Technology" which has been working on combinations of industrial digital cameras and polar measurement systems since 2006 resulting in multiple versions of digital camera-theodolites (Schlüter, et al., 2009).

Figure 9: i3Mainz Digitalkameratachymeter mounting on a Leica TCRM 1103 (Hauth & Schlüter, 2010)

The newest system consists of an industrial digital camera with a resolution of 2560 x 1920 pixels which is mounted onto the ocular of the total station (Leica TCRM1103) with an adapter (Figure 9). The focus of the camera is fixed onto the crosshair plane of the telescope; therefore no further interference with the optical path of the sensor is necessary. A fixed mounting plate is screwed to the back of the eye-piece to allow a fast assembly of the system while not interfering with normal mode of operation of the instrument (Hauth & Schlüter, 2010).

2.1.2 Current commercial systems

In the beginning of the 21st century the availability of a wide range of compact industrial digital camera hardware on the one side and growing computational power, data transfer and storage capacities on the other side strongly increased the interest in the topic of image-assisted technologies.

Manufacturers aimed at developing functionalities for interactive, semi-automatic surveying modes to support the measurement process and have developed total stations with varying degrees of imaging capabilities. In the following the currently available models including their features are described.

Figure 10: a) Topcon Imaging Station, b) Trimble VX Spatial Station, c) Pentax R-400VDN Visio, d) Leica Nova MS50

In the year 2004 Topcon released the system GPT-7000i with an ocular camera of 640 x 480 pixel within the optical path of the telescope utilizing its magnification factor of 30. In addition a wide angle overview camera was available (Topcon, 2004) in order to allow aiming of the instrument by clicking on the displayed image. Countering the previous model's lack of automated focus and remotely controllable actuators the Topcon Imaging Station (IS) was released in 2008. The resolution of both cameras was increased to 1280 x 1024 pixel with a maximum frame rate of 10 Hz (Topcon, 2008). The instrument was equipped with radio antennas and could be controlled remotely. In 2012 an update with similar specifications and an added long range Wi-Fi control was released under the name IS-3 (Topcon, 2012).

The Trimble VX Spatial Station released in 2007 has an eccentric overview camera with a field of view of 18.3 x 13.7 gon with a focus range from 3 meters to infinity. The camera resolution is 2048 x 1536 pixel and a frame rate of 5 Hz can be reached. Through camera calibration the image pixels can be related to the ocular measurements. This can be used to remotely control the instrument and initiate measurements using a virtual target on the captured images or to document and overlay executed measurements with image texture (Trimble, 2007).

In 2009 the company Pentax released the R-400VDN Visio series which follows a similar concept. The instruments are equipped with an overview RGB camera above the telescope which can capture and document all executed measurements. The maximum camera resolution is 2048 x 1536 pixel with a maximum frame rate of 10 Hz. Images are captured in a field of view of around 9.8 x 7.3 gon at a fixed focus with a range from 20 m upwards. Images are stored on an SD card (Pentax, 2009).

Using the experience gained from their first IATS prototype version from 2003 (Walser, 2005) and the small batch series of IATS2 produced in 2008 (as described in Section 2.2) Leica Geosystems included two 2560 x 1920 pixel imaging sensors in their new total station MS50 (Figure 10d) which was released in 2013. The maximum frame rate that can be reached is 20 Hz. The wide angle overview camera has a field of view of around 17.2 x 13 gon whereas the camera coaxial to the optical path of the telescope has a field of view of around 1.4 x 1.1 gon and a magnification factor of 30. The focus of the overview camera is kept fix and has a range of 2 m and further while the telescope-camera is able to variably focus from 1.7 m to infinity. The angular resolution in pixel is 0.0005 gon and 0.0067 gon for the telescope and the overview camera respectively (Leica Geosystems AG, 2013a; 2013b).

2.2 Prototype IATS2 – Leica TCRA1201+ hardware components

Until the introduction of the Leica MS50 in late 2013 the apparent restrictions in terms of interfaces and accessibility the commercial systems could not be integrated into a self-developed photogrammetric processing chain necessary to independently evaluate system performance and exploit the full possible potential of the technology. Academic researchers therefore focused on creating their own creative and open solutions as described in Section 2.1.1. These adaptions to existing instruments are however accompanied by unwanted side effects such as additional inaccuracies introduced by custom made lenses or mount-on camera equipment. As the main know-how of how to build highly accurate measurement systems avoiding the mentioned deficiencies resides at the instrument manufacturers the IATS2 built directly by Leica Geosystems represents an attractive solution to the academic community. The interfaces to the instrument and the camera are documented and accessible, however, the device is still not fully integrated and market ready as some issues with telescope balance caused by the additional weight of the camera and power supply and data transfer cables which are not run out through the vertical axis at the base of the unit (Figure 6), remain.

Similarly to the IATS2 the Leica MS50 can also be remotely controlled over defined interfaces. It is even possible to apply custom written hardware drivers to communicate with the instrument. An adaption of the abstract hierarchical high and low level driver model developed within DE-MONTES for TLS and IATS2 control (Wagner, et al., 2013a) was developed to also control the MS50 (Briechle, 2014) making the instrument fully compatible with the developed photogrammetric processing chain. The full integration of the imaging hardware and connector cabling in the housing of the MS50 however, leads to a significant reduction in data transfer rates to the controlling computer. This effect is caused by the fact that the direct camera data connection cable was replaced by slip ring contacts into the interior of the total station to facilitate measurements. Test measurements have shown that the data transfer rate from the device was reduced by a factor of four (Wagner, et al., 2014). In this work only the IATS2 instrument described in this section was used.

The following sections give a description of the individual components of the Leica IATS2 instrument upon which the image analysis conducted in this work is based. Information about the base system the Leica TCRA1201+ R1000 will be given, followed by the adaptations necessary to include an ocular camera together with a focusing mechanism and the appropriate interfacing to control the camera as well as the instrument.

2.2.1 Angle measurements / Base instrument

The most important component of an IATS system is the total station used as base instrument. In the following the instrument specifications of the IATS2 in terms of measurement accuracy will be given.

The base instrument of the IATS2 is the Leica Total Station TCRA1201+ R1000 which was introduced in 2007 and is in term based on the previous TPS1200 class of instruments (Leica Geosystems AG, 2007a). The angle measurement accuracy of the device is 0.3 mgon. Distance measurements of reflectors can be executed at an accuracy of 1 mm ±1.5 ppm up to a distance of 1800 meters at strong sunlight or haze, 3000 meters at light haze or moderate sunlight and up to 3500 meters with overcast skies and no haze. Reflector-less distance measurements up to 500 meters reach an accuracy of 2 mm ±2 ppm and 4 mm ±2 ppm above that distance in favorable atmospheric conditions (object in the shade or overcast sky). Distances of up to and above 1000 meters can be measured in the reflector-less mode at night time or twilight and go down to a maximum of 800 meters at favorable daytime conditions and 600 meters in direct sunlight assuming the use of a Kodak Gray card with 90% reflectance. For a Kodak Gray card with 18% reflectance only half of the respective distances can be reached (Leica Geosystems AG, 2007a) whereas in practice natural targets regularly lead to even lower values. When using reflector-less measurements, beam divergence as a function of distance also has to be taken into account. The size of the laser beam’s footprint on the surface is depending on this divergence. For this instrument it is 12x40 mm at a distance of 100 meters but already 60x200 mm at a distance of 500 meters (Leica Geosystems

AG, 2007b) limiting the usability for precise point measurements. The figures given by the manufacturer are 1σ standard deviations (Leica Geosystems AG, 2007a).

2.2.2 Instrument adaptations / Ocular camera

The by far most important adaption of the base instrument in order to transform it into an IATS was the integration of a digital camera in the optical path of the telescope of the total station. The camera is mounted so that the imaging sensor lies in the focal point of the telescope replacing the crosshairs. A defined physical reticle is not necessary anymore as a virtual one can be drawn in the resulting images at any time (Wasmeier, 2009, p. 18).

Leica used an industrial camera from the German manufacturer imaging Development Systems (IDS), namely the model UI-148xLE-C from the uEye series. The camera has a CMOS color chip with a resolution of 2560 x 1920 pixels and a sensitive area of 5.63 x 4.22 mm. The used color filter array is implemented in the form of a Bayer pattern (Bayer, 1976) with a 25% blue and red and 50% green coverage. The physical side length of each pixel is therefore 2.2 μm. Due to the inherent properties of the used CMOS sensor of lower power consumption and higher readout rates compared to a CCD sensor the maximal frame rate is 6.3 full frames per second. With reduced resolution or the readout of only certain regions of the full image higher frame-rates are possible. Outcrops with the HD resolution of 1920 x 1080 Pixels can be acquired in a frequency of 13 frames per second, XGA (1024 x 768) and VGA (640 x 480) images at a frequency of 27 and 52 frames per second respectively (IDS GmbH, 2009). For certain applications the measurement frequency can be more important than the resolution and can be exploited for high frequency measurements of construction oscillations as done by Bürki, et al. (2010) with their own IATS system (Section 2.1.1) or to model atmospheric refraction influence by optical turbulences (Reiterer, 2012) with the here described Leica IATS2.

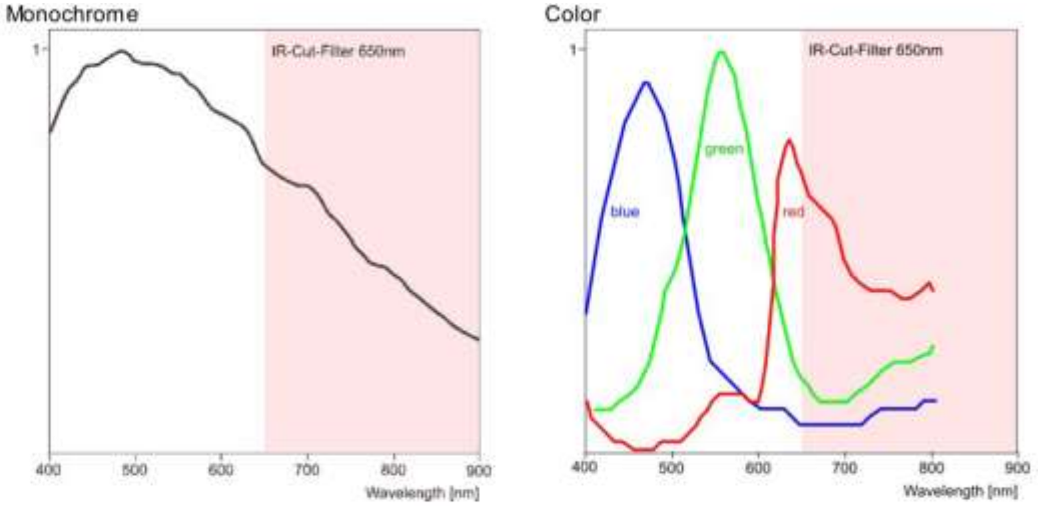

Figure 11: Relative Sensor Sensitivity of UI-148x (IDS GmbH, 2009, p. 134)

Figure 11 shows the relative sensor sensitivity of the used camera. The optionally available IR-Cut-Filter is not used within the IATS2 system therefore signals in the near infrared range from 650 to 900 nm are also present in the images. The brightness information captured by the Bayer pattern is transferred to the controlling PC where it is converted to a color image and an additional gray channel representation with a function of the uEye API (IDS GmbH, 2009, p. 93). The evaluations conducted in this work as well as measurements on acquired IATS data generally rely on the monochromatic image representations. However, for the documentation of measured points or remote aiming in live stream of images the color information is especially helpful. Furthermore, color can be used for a rough segmentation of a measured region of interest, e.g. based on vegetation.

The point of focus of the ocular camera can be changed moving a lens by manually turning a focusing ring. Through the use of a connected servo motor the focusing can additionally be automated or executed remotely. The servomotor 3564B from the company Faulhaber (Faulhaber GmbH & Co.KG., 2007) was applied for this purpose as it offers a serial interface to a controlling computer. A limit switch at the focus position of infinity serves as a reference position from which the motor can reproducibly actuate every specified focus position using a count of its encoder steps. Consequently focus positions are simply specified as certain encoder step values (Wasmeier, 2009, p. 20). According to Wasmeier (2009) the focusing lens can be kept at a stable position for most applications of geo-monitoring as the measuring distances generally surpass the 75 to 80 meters after which focusing steps only change in an insignificant range. To compensate for the additional weight of the camera a compensation ring was mounted on the front face of the telescope (Figure 5).

2.3 System calibration

An IATS system provides images in addition to precise angular measurements. In combination with the long focal length and the ability to easily change vertical and horizontal viewing angles this is the main difference and advantage compared to a conventional photogrammetric system. In contrast to a normal theodolite multiple targets can be selected within an image to retrieve their spatial vector. These targets are no longer at the center of the telescope but at arbitrary positions in the field of view. This results in an angular offset from the measured point to the theodolite reference axis intersection point.

To make use of acquired images for measuring purposes it is essential to find a calibration procedure that allows retrieving a mathematical function relating every image position to an angle value with respect to the original theodolite readings. While for fixed-focus cameras one set of calibration parameters is sufficient the calibration has to be repeated for different focus positions as these change the optical path of the telescope (Wasmeier & Reith, 2013).

In addition to this external orientation of the camera in relation to the theodolite, for an optimal result the intrinsic camera parameters have to be calibrated as well. Due to the long focus and the used industrial camera the distortion values are negligible around the center of the images but have to increasingly be taken into account the closer to the border of an image an angle reading is required.

Therefore the integrated calibration of the whole IATS device consists of three individual calibration steps.

1) The base instrument calibration of the theodolite,
2) the calibration of image pixels in respect to theodolite readings and
3) the intrinsic camera calibration.

2.3.1 Instrument axis calibration

In order to execute precise angular measurements with a theodolite the main condition that has to be fulfilled is that all axes of the instrument are perpendicular to each other and intersect in one point. However, in practice small deviations will always remain due to inaccuracies in the production process as well as caused by small internal movements in the device while in productive use. The following errors occur and have to calibrated and re-calibrated in accordance with recommendations from the manufacturer: line-of sight error, tilting axis error, vertical axis error, vertical index error (Leica Geosystems AG, 2003).

2.3.2 Camera calibration

Based on the pinhole camera model describing the relationship of 3D coordinates projected onto a 2D image plane through a single point aperture, the center of projection, a photogrammetric camera system is defined with the parameters of intrinsic and extrinsic orientation.

The intrinsic parameters to be specified are (Hartley & Zisserman, 2000, p. 153ff) the principle point (x_0, y_0), the focal length (c) and lens distortion parameters (Hartley & Zisserman, 2000, p. 189). These parameters are generally modeled in a calibration matrix **K** as illustrated in (Hartley & Zisserman, 2000, p. 157) which is part of a 3x4 general projection matrix **P** mapping from image to world coordinates. It is formed by a multiplication of **K**, the 3x3 rotation matrix **R** and the 3x4 matrix $[I| - C]$ combining the 3x3 identity matrix **I** and the vector $\boldsymbol{C}$ describing the position of the camera's projection center $\boldsymbol{X_0} = (X_0, Y_0, Z_0)^T$ in world coordinates.

Distortion parameters are not included in this form and are modelled using the spatially dependent corrections $\Delta x, \Delta y$. These stem from a combination of symmetric lens distortion, tangential lens distortion, out-of-plane distortion and in-plane-distortion with a predominant influence from symmetric distortion (Hartley & Zisserman, 2000, p. 191).

The intrinsic orientation can be calculated from bundle adjustment of test-field images. In the case of the calibration procedure developed by Wasmeier (2009) these parameters will be estimated together with exterior orientation based on the photogrammetric collinearity equations as the distortion of the IATS camera is rather small due to the long focal length and the high quality of the lens.

In contrast to the projection matrix **P** used in computer vision the photogrammetric approach of camera calibration is based on the collinearity equations which are formulating the basic principles of photogrammetry and are used to relate 3D object space coordinates $\boldsymbol{X} = (X, Y, Z)^T$ to their respective 2D image coordinates (x, y) from known camera calibration parameters, combining the above mentioned intrinsic calibration with extrinsic camera parameters. For further information and the according equations refer to Kraus, 2007, p. 436f.

Based on these equations a combined determination of intrinsic and extrinsic orientation is possible given a test-field with a sufficient amount of spatially well distributed known targets or additional geometric constraints (Wasmeier, 2009, p. 35).

2.3.3 Combined calibration approach

While the intrinsic parameters of an IATS are independent from the theodolite the extrinsic parameters are tightly connected to the instrument readings as these form an essential component of the extrinsic orientation. This requires a combined calibration of theodolite readings and extrinsic camera parameters.

Vogel (2006) illustrated the different components of a theodolite with a photogrammetric device including the parameters to be calibrated. He extended the collinearity equations (Kraus, 2007, p. 436f) with the parameters describing the position of the camera in relation to the theodolite. According to him the IATS parameters to be calibrated are the principle point, the first term of radial distortion, focal length, theodolite line of sight error and tilting axis error, photogrammetric line of sight error and vertical index error (horizontal and vertical deviation of camera center from theodolite center) and the distance between camera center and theodolite center. In addition the camera sensor tilting angles causing deviations from perpendicularity with theodolite line of sight also have to be determined.

In case of the Leica IATS2 not all of these parameters can be calibrated. In contrast to the IATS used by Vogel (2006) the projection center of the camera and the theodolite is in the range of <1 mm to <10 mm depending on the focus position. These small values of lead to ill-conditioned normal equations, therefore the horizontal and vertical deviation of the center of projection from the theodolite center cannot be calibrated (Wasmeier, 2009, p. 39).

Another problem is the detection of the tilting of the camera sensor which consists of a generally small rotation around the x or y axis of the sensor (ω, φ). A combined extrinsic parameter estimation of camera and instrument will show these errors as small constant components of the global rotation with a strong correlation to the position of the principle point. In order to still calibrate these components this calls for calibration measurements with known reference angle readings for all targets. In contrast a rotation of the sensor around its z axis (κ) is easily detectable as the base instrument is not capable of such a movement (Wasmeier, 2009, p. 31f).

Another approach to the instrument calibration was developed by Walser (2005). It describes the calibration in a more image driven approach. The relation between theodolite and image coordinates is modelled as an affine transformation. The parameters to be calibrated in Walser's approach are the vertical index error, line of sight error, tilting axis error, column, line scale factors, a skew factor and the rotation of the sensor around its Z axis.

This approach is not strictly a photogrammetric calibration as the intrinsic orientation is not calibrated. The focal length and the radial distortion have to be known. The principle point can be chosen at an arbitrary reference pixel and is an input parameter for the procedure that calculates correction parameters for theodolite readings in relation to the reference principle point. As the principle point is variable with changing focus positions so are these correction parameters (Wasmeier, 2009, p. 43).

Wasmeier (2009) combined these two approaches using the first radial distortion term and the focal length calculated with Vogel's method and the affine correction parameters from Walser using the reference principle point calculated with Vogel's method. This allowed Wasmeier a comparison and mutual verification of the two sets of calibration parameters which were found to coincide well (Wasmeier, 2009, p. 43ff). In further calibrations the principle point position at the geometrical centre of the camera sensor was chosen.

The actual calibration process was carried out on a virtual calibration field created with the use of a collimator. Using the motorization of the total station it is possible to create an arbitrarily dense target field by capturing the collimator with the IATS with the virtual target appearing at a different position in each image. Each observed target yields two sub-pixel coordinates and two angular readings from the theodolite. The number of unknowns to adjust is the amount of calibration parameters plus two angles of the target. These stay the same in each image as the target is stationary while only the camera is moved. The observed target field leads to an over-determined system which can be linearized and evaluated (Wasmeier & Reith, 2013).

As the calibration has to be done in a controlled environment a collimator with a movable tube is used. The light of an illuminated target in the collimator is aligned parallel so that a virtual target at infinity is created. Lenses in the moveable tube bring the light to diverge making the target appear at a specific reference distance. The target of the collimator consisting of a crosshair and multiple circles can be detected and matched accurately using special image analysis techniques. The sub-pixel accuracy of these methods lies in the range of $\frac{1}{20}$ pixel or better (Wasmeier & Reith, 2013). The calibration process has to be executed for each focus position to be calibrated leading to a stable set of parameters. Ten sampling points of focus position and object distance were found to be enough to determine the recession curve for the full focussing range of the instrument (Wasmeier, 2009, p. 34).

2.4 Control software

One main target of research work in connection with new hardware is the possibility to access the full range of functionality of the device and to configure and evaluate all instrument parameters. Due to proprietary closed software environments aimed at binding a customer closer to the respective manufacturer this is generally not fully possible with surveying instruments on the market. As a result of its prototype status this is not the case for the used Leica IATS2, however, this brought up the necessity to create custom control software for the device. In parallel to this work data acquisition software was developed by Andreas Wagner at the Technical University of Munich which is described in detail in Wagner & Reith (2013).

The software allows using the instruments either in an interactive mode wherein parameters can be set and feedback is shown in a graphical user interface (GUI) or with command line calls in order to allow a connection to on overlying software architecture to automatically execute measurements in certain time intervals for mid or long term automated monitoring tasks.

Data acquisition takes around 5 - 5.5 seconds per image on each device including automated pointing of the instrument and transfer of the uncompressed image to data storage. The field of view of each camera at focus position ∞ is around 1.56 x 1.17 gon. The measurement time for a certain region of interest is defined by the field of view and the set overlap of images.

The software can either be used to control one single device or a stereo setup of two instruments. This depends on the used stereo baseline. Smaller baselines can be measured using one controlling PC (usually a field capable laptop) with cable connections to both devices whereas wider baselines generally require two instances of the control software on computers each directly connected to one instrument via cable. To temporally synchronize the measurements the system clocks of the controlling computers have to be synchronized as well. The controlling computers can be connected using a Wi-Fi bridge (Wagner, et al., 2013b) and operated remotely.

The command line batch processing mode can be accessed by external software managing automated continuous measurements and data processing. Data acquisition was tested with the "DE-MONTES Data Acquisition Module" (Furuya, et al., 2013) and the "DE-MONTES Monitoring Overlay" (Nauschnegg, et al., 2013). The latter allows a specification of continuous monitoring tasks such as data acquisition, processing and evaluation dependent on the successful evaluation of the previous step.

3 System set-up and design

The following chapter gives an insight into recommended system set-up configurations including theoretical accuracy considerations for the used stereo set-up. Furthermore, the data processing scheme will be introduced. It consists of a modular photogrammetric processing chain representing the core component of this work. Due to its modular design it can be adapted towards the needs of different workflow requirements allowing to adjust the trade-off between reduced processing time and higher accuracy results.

The typical set-up consists of stereo configurations of two instruments. Measurements from more than two instruments would allow redundant evaluations of processing results but the increased effort of system set-up and data acquisition minimize the potential benefits. Therefore only stereo measurements are covered in this section.

3.1 Theoretical accuracy considerations

An accuracy evaluation of the spatial forward intersection of different stereo set-ups depends on the geometric configuration. The positional errors to be expected can be calculated based on the base-to-distance ratio, errors of the station coordinates, instrument errors and matching accuracy. Matching errors are hereby harder to evaluate than geometrical constraints as they are not a stable factor but strongly depend on the content of the evaluated images. In general automated matching algorithms depend on well-defined image texture and show an increase in reliability and accuracy with decreasing variations between analyzed images. Hence, larger viewpoint changes stemming from wider baselines and varying illumination conditions from different measurement epochs may decrease the accuracy of matching results.

Digital cameras capture data in a fixed raster containing intensity value information. That way the Leica IATS2 produces 2560 x 1920 individual measurements in the field of view per captured single channel image. The algorithms used and evaluated in this work (Chapter 4) can reach sub-pixel accuracy by fitting mathematical functions to (sections of) the image content. Calculated values such as the continuous function's point of inflection may now lie in between the discrete raster of pixels. Different algorithms can reach different levels of accuracy ranging from less than 0.1 pixels to around one pixel. In this first theoretical accuracy estimation a fixed value of 0.3 pixel was chosen as a compromise between the different mentioned algorithms. To further reduce complexity possible matching accuracy degradation with increasing viewpoint changes was also not integrated in the following estimations.

The distance to the measured object in relation to the length of the baseline between two instruments is the most important factor in the measurement set-up. The measured object should be located close to the perpendicular bisector of the base line. Errors in perpendicular direction to the baseline increase with a decreased baseline-to-distance ratio (e.g. from 1:3 to 1:4). Errors parallel to the baseline however do not suffer from this effect. This leads to the conclusion that the direction of expected deformations has to be taken into account when planning system set-up and measurement configuration.

Another important factor restricting the possibilities of the system set-up is the automated point measurement as the matching quality decreases with larger viewpoint changes between images. A base-to-distance ratio of 2:1 would theoretically lead to evenly distributed errors in the direction of and perpendicular to the base line. This results in a change of view of 90 degrees and makes point matching of natural targets impossible and therefore prohibits the use of such a configuration. From the perspective of image processing a maximum base-to-distance ratio of 1:2 or better 1:3 is required, leading to view-point changes ~28 and ~19 degrees respectively.

The accuracy that can be achieved with a specific configuration can be calculated based on the accuracy of the station coordinates, the instrument's angle measurement accuracy as given by the manufacturer, and the accuracy of the matching process. Using the values from the Pellheim field experiment (Section 8.2) an accuracy of σ_x = 2 mm is assumed for the coordinates of the stations. The measurement accuracy of the used IATS is given as σ_t = 0.3 mgon. The matching accuracy value of 0.3 pixel with a resolution of ~0.6 mgon per pixel translates to an angle of σ_a= 0.18 mgon. The horizontal accuracy $\sigma_{t'}$ is based on the two direction measurement of the orientation and the direction to the target and can be calculated with Gaussian error propagation (Equation (3.1)). Due to the levelling of the instrument only a single measurement is used for the vertical direction.

$$\sigma_{t'} = \sqrt{2} \cdot \sqrt{\sigma_t^2 + \sigma_a^2} \qquad \sigma_{v'} = \sqrt{\sigma_t^2 + \sigma_a^2} \tag{3.1}$$

This leads to values of ~0.35 mgon for vertical $\sigma_{v'}$ and ~0.49 mgon for horizontal accuracy $\sigma_{t'}$. The forward intersection can be expressed as a constrained adjustment of the minimum distance of rays each going through the camera center of one of the IATS using the Gauß-Helmert model (Niemeier, 2008, p.

176; Equation (6.8)). The eigenvalues of the resulting covariance matrix $\mathbf{Q}_{xx}$ express the square of the semi-major and semi-minor axes of the error ellipse. The direction of the error ellipse is expressed by the eigenvectors. Using a principle component transformation $\mathbf{Q}_{xx}$ can be rotated into the direction of the baseline to calculate transversal and longitudinal errors (Wagner, et al., 2014). The diagonal of the cofactor matrix of the observations $\boldsymbol{Q}_{ll}$ (Equation (6.9); Section 6.4) can be inserted with the squared values of the above calculated figures to derive the error ellipse. Simulated Helmert point errors (Niemeier, 2002, p. 260) for a baseline of 100 meters using the example values introduced above are illustrated in Figure 12.

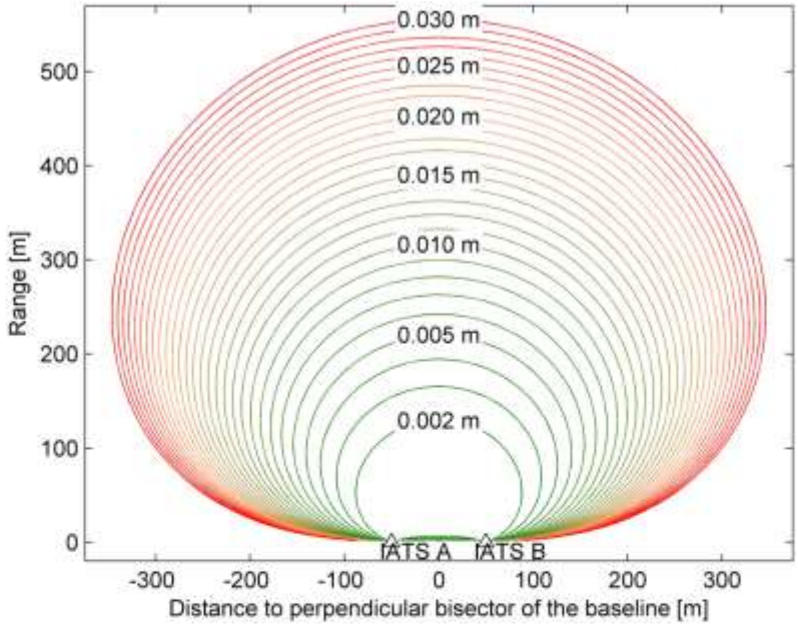

Figure 12: Helmert point error for a baseline of 100 m, angle measurement accuracy 0.49 mgon, station accuracy 2 mm (Wagner, et al., 2014)

The above shown evaluations allow a suitable estimation of the influence of different measurement configurations on the achievable accuracies.

3.2 Data acquisition and processing scheme

The IATS based measurement system described in this work can be used with varying data acquisition strategies targeted at different use cases. The key parameters of each use case are the required accuracy, time constraints and the size of the measured target region. Unfortunately these are conflicting objectives which require a well-balanced trade-off as higher accuracy and larger target regions lead to longer processing times.

3.2.1 Continuous monitoring

The first main use case is continuous monitoring of an endangered site requiring a fixed installation and continuous data acquisition and processing. The time constraints for processing are defined by the measurement frequency. Full processing of each epoch has to be finished by the time the data of the next epoch is captured. The accuracy of the produced results will typically be limited by these constraints. The measurement process for this use case is executed using the automated data acquisition module as mentioned in Section 2.4. Data acquisition takes around 5-6 seconds per image, processing however requires access to both stereo-partners of an image pair. This means that the data transfer between the two instruments over a customized long distance Wi-Fi link also has to be taken into account. According to tests this adds another ~1.2 to 2 seconds for stereo distances up to 600 meters and ~5 to 80 seconds for distances around 1.5 km (Wagner & Reith, 2013). Depending on connection speeds a fully continuous monitoring would therefore leave around 7-8 seconds for processing. This requires the processing chain to be flexible enough to run a fast coarse matching continuously and re-process parts of the data in higher accuracy at night time or as soon as the illumination conditions prohibit new data acquisition.

3.2.2 Single measurement campaign

The second main use case consists of single measurement campaigns in which one or multiple data-sets are captured. The devices are packed up after each campaign and the data acquisition is done with the interactive mode of the IATS control software. In this case processing is typically done after the measurement process is complete without time constraints. However, hybrid set-up configurations are also possible wherein a measured data-set in a campaign is immediately processed on-site, analyzed and used to plan and define the consecutive measurement epochs. This again sets tough time and hardware constraints for the immediate processing.

3.2.3 Processing chain

The applied processing chain has to be modular and flexible enough to meet the mentioned use case requirements using an appropriate configuration and parameterization. In the case of little available processing time faster processing methods have to be chosen limiting the final accuracy. Up to a certain level the acquisition of more computational power can also be used to increase processing speed, however, costs of powerful high-end hardware may exceed the budget of specific measurement applications. The focus in this work lies on the software side as the development, fine tuning and evaluation of a flexible processing chain consisting of multiple different methods and algorithms forms the core component of this thesis.

The processing chain consists of the following consecutive stages:

- Pre-processing,
- Homologous point detection, and
- Mismatch detection and match refinement

Depending on the available processing time various components can be in- or excluded. The different stages of implemented processing steps are illustrated in Figure 13 - 15. The individual methods and algorithms are described in detail in Chapters 4 and 6. The fastest processing option used in the continuous monitoring case is highlighted in light red in each Figure; all other methods can be used additionally. It has to be noted that the least squares matching based match refinement step is computationally too costly and therefore completely excluded from fast processing.

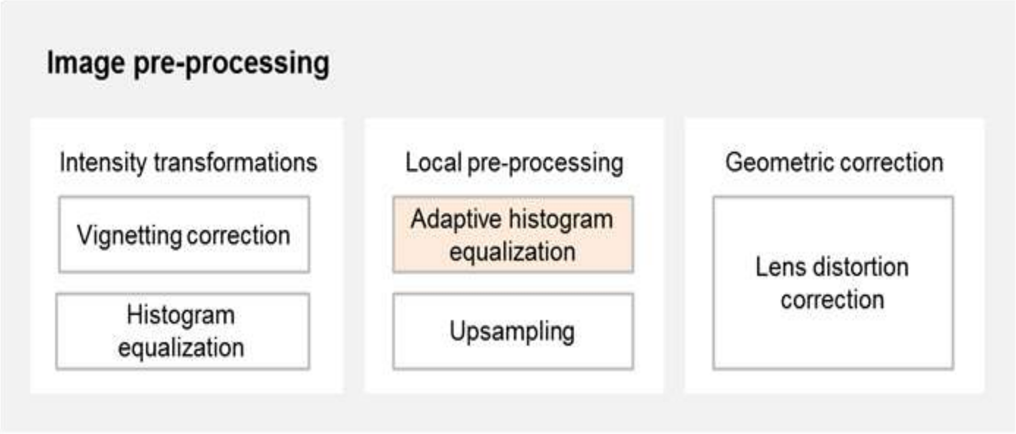

Figure 13: Image pre-processing steps (light red: coarse, fast processing)

Image pre-processing describes operations at the lowest level of abstraction. Input and output images are still raster of intensity values. Pre-processing aims at enhancing certain image features for further processing whereas suppressing unwanted features such as image noise and distortions which might stem from lens deficiencies (Sonka, et al., 2008, p. 113). In the developed processing chain the main pre-processing steps applied aim at evenly distributing intensity levels over the whole image and the correction of lens deficiencies found in the calibration (Figure 13). The illumination falloff towards the periphery of the is strongly limited by the long focal length used. This effect and its compensation are called vignetting correction. More important are the histogram equalization and the adaptive histogram equalization. Both try to equally distribute intensity levels over the full intensity scale. This can be used to increase contrast; especially near histogram maxima (Sonka, et al., 2008, p. 116). Adaptive histogram equalization applies this process not over the whole image but individually on smaller image-tiles to preserve and increase local contrast deviations (Section 4.2). These contrast enhancement methods enhance the quality of the results of the subsequent process of homologous point detection using feature matching. Another simple pre-processing concept that greatly enhances further processing accuracy is up-sampling of the images (Xu, et al., 2012). This describes an artificial increase of image resolution generating new pixel values with interpolation methods. Nonetheless the increase of image size is also reflected in additional further processing time. In the case of global or local intensity transformations the geometry of the image remains unchanged, whereas in the case of up-sampling the principle point and focal length measured in pixels have to be adapted to new image dimensions.

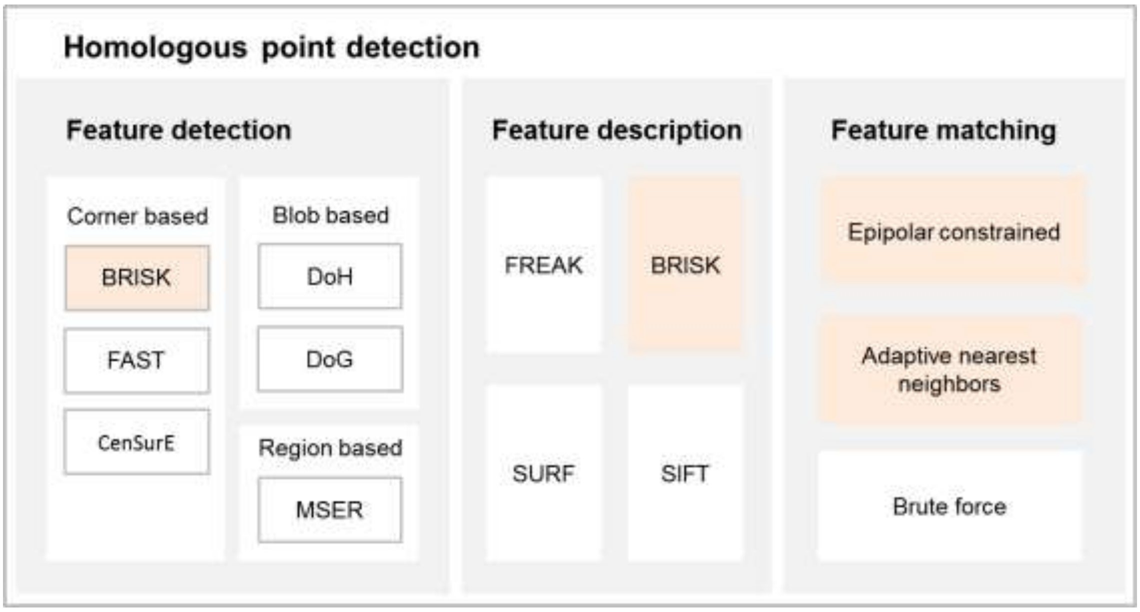

Figure 14: Homologous point detection process consisting of feature detection, description and matching (light red: coarse, fast processing)[2]

The basic principle of the whole measurement system is the detection of homologous non-signalized points consisting of image features in stereo-images as well as over multiple measurement epochs. This allows the calculation and evaluation of deformation vectors in a similar manner as from manual surveying. Dense matching methods are computationally too expensive and time consuming and were therefore excluded as processing option. Homologous point detection can be divided in the sub-processes feature detection, feature description and a feature matching (Figure 14). Various

---

[2] Refer to Chapter 11 List of Abbreviations for the meaning of the mentioned abbreviations

feature detection methods were evaluated and implemented in the processing chain so as to exploit complimentary characteristics of the different methods. Matching of the generated feature description of each point generally requires a mutual comparison of each vector with every other one in the set. This full unrestricted comparison leads to a quadratic relation of the amount of features and the processing time and is referred to as "brute force matching". The amount of points to be compared can be reduced with the known epipolar relation between images. This procedure will further be called epipolar constrained matching. The relation between point set size and computational cost is linear. However, this can only be applied for a stereo set-up and will not match deformations occurring over multiple measurement epochs as they violate the global geometric relation between the images describing only a change in camera pose but not in image content. Another matching possibility is the generation of a tree like structure describing the similarity of feature vectors. Generation of these structures can be done computationally efficient using specialized software libraries; the matching process consists of following the tree content which can also be done in constant time (Muja & Lowe, 2009).

Figure 15: Mismatch detection and match refinement (light red: coarse, fast processing)

The results after the matching process may still contain errors reducing the results of the final measurements. Therefore a third post-processing component is processed consisting of mismatch detection and match refinement (Figure 15). Mismatches can be found as correspondences violating a certain model describing image relation. One of these models is the epipolar relation between images again which can be extended by a local evaluation of the variation of the match position on the epipolar line. That way unlikely large 3D variation in a local neighborhood (spikes) can be detected and excluded. Another local possibility to detect mismatches is the calculation of so called meta-descriptors formed by the spatial relation of groups of features (Liu, et al., 2012). A last step is the refinement of matches using the least squares matching method (Grün, 1985). The implemented process strongly benefits from reliable approximations and well defined points as delivered by the previous processing steps. According to the required accuracy the refinement can be carried out with two free parameters adjusting a shift, four estimating a helmert transformation or six parameters estimating an affine transformation. In addition two radiometric parameters can be adjusted as well consisting of a constant intensity value shift and an intensity value multiplication factor. Least squares matching adjusts and improves the found transformation parameters over multiple iterations until a consistency criterion is met. To improve the reliability of results and to introduce an error estimate the process is applied from search image to reference image and back. The distance from the original starting point to the resulting back-matching position in pixel serves as an accuracy measure. The resulting high matching accuracy comes at the cost of time consuming computation making this method unavailable for fast processing.

4 Image pre-processing

4.1 Introduction

In the case of photogrammetric measurements the influence of illumination plays in important role in how the texture of the measured object appears in the acquired data. This becomes even more of a problem the further apart the different viewpoints are situated in a spatial or temporal context. For later matching purposes it is important to have highest possible similarity which can be reached with suitable image pre-processing. In theory most pre-processing steps could also be integrated into subsequent processing, it makes sense however, to keep the operations separated to benefit from optimized implementations and to avoid unwanted duplication. This leads to the placement of image pre-processing components at the beginning of the processing chain described in Section 3.2.3 (Figure 13).

Image pre-processing is applied at intensity value level. The input as well as the output of the process are images represented as matrices of intensity values. At this stage the images are generally still in the same form as captured by the sensor. The aim of pre-processing is the suppression of unwanted image features or distortions or the enhancement of features that facilitate further processing steps. The elimination of unwanted features can therefore reduce image information while at the same time improving processing results (Sonka, et al., 2008, p. 113).

For the application at hand it is important not to distort the geometry of the images which would introduce measurement errors. Therefore the described methods only apply pixel wise intensity transformations. With the exception of lens distortion correction which corrects the relation of image positions to measurement angles, thereby improving the accuracy of 3D measurements. Image up-sampling does not change the geometry of the image however the intrinsic parameters of principle point and focal length have to be adapted to the new dimensions.

4.2 Intensity transformations

This class of pre-processing methods modifies the intensity value of each pixel. An intensity correction can be applied depending on the pixel position to correct systematic deficiencies. The systematic errors can be represented as a multiplicative coefficient $e(i,j)$ for a corrected image $g(i,j)$ forming the degraded image $f(i,j)$.

$$f(i,j) = e(i,j)\, g(i,j) \qquad\qquad g(i,j) = \frac{f(i,j)}{e(\,i,j)} = \frac{c\, f(i,j)}{f_c(i,j)} \tag{4.1}$$

To obtain $e(i,j)$ a reference image with a known intensity value c can be taken resulting in the degraded image $f_c(i,j)$. The found error coefficients can then be used to correct the original images as long as the transformation is valid. It may be necessary to periodically re-calculate the error coefficients (Sonka, et al., 2008, p. 115). This can also be used to correct vignetting effects describing the illumination falloff from the principle point towards the image borders proportional to cos^4 of the irradiation angle on the sensor (Sonka, et al., 2008, p. 90).

Intensity transformations can also be applied non-position dependent so that $q = T(p)$ describes the transformation from intensities $p\,[p_0, p_k]$ to $q\,[q_0, q_k]$. Intensity transformations are particularly useful to enhance image contrast. The histogram equalization technique can be applied to generate an image with intensity values equally distributed over the full available intensity scale, thus enhancing contrast close to histogram maxima while decreasing contrast near minima (Sonka, et al., 2008, p. 116).

The effect of the application of histogram equalization can be seen in Figure 16. The cumulative probability density shows the equal intensity distribution over the full range of a one channel eight bit image with its 256 intensity values.

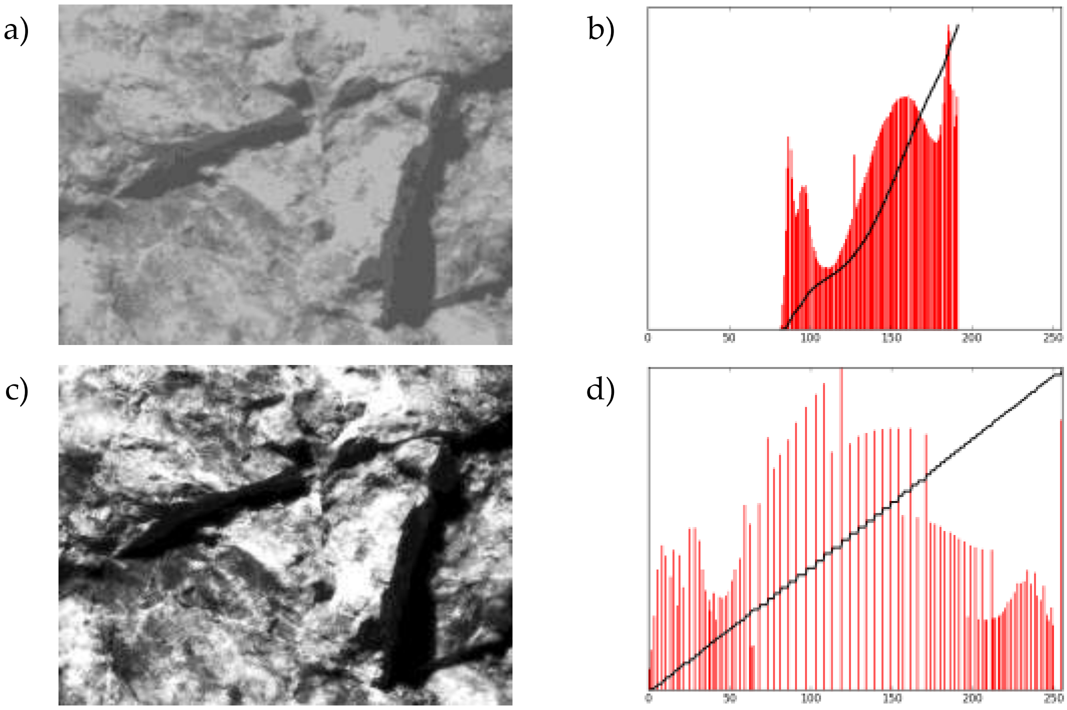

Figure 16: a) low contrast image and c) result of histogram equalization with corresponding histograms (red) and cumulative histograms (black) in b) and d)

Histogram equalization supports a later matching process by adjusting the contrast range in the stereo-images in a similar manner; however, areas deviating from the mean intensity such as shadows are not sufficiently enhanced.

4.3 Local neighborhood based pre-processing

Local neighborhood based pre-processing uses a batch of surrounding intensity values of a pixel to produce new output values.

In the IATS processing chain local neighborhood based pre-processing is used for two purposes. The first is the up-sampling of images to increase the sub-pixel accuracy of results. The second application of local neighborhood based pre-processing is contrast limited adaptive histogram equalization which performs histogram equalization over smaller batches of the image and results in a much more homogeneous contrast distribution than global histogram equalization in cases where the image contains regions that are significantly lighter or darker than the majority of the content.

Image up-sampling is a simple method to improve sub-pixel accuracy by increasing the image size in pixels using interpolation. An extensive analysis and evaluation of the method and its implication on stereomatching was carried out by Xu, et al. (2012). They developed a simple up-sampling framework leading to an accuracy increase of around 20% for dense stereomatching. Both bilinear and bi-cubic interpolation were reported to lead to satisfactory results. This method can also be used in the implemented IATS based processing chain. However, applying this step is only possible for very high accuracy applications without strong time constraints. Although the up-sampling pre-processing step itself is computationally inexpensive the improved quality of results comes at the cost of a quadratic increase in processing time with the magnification factor for almost all further processing steps.

Up-sampling of images can also be used to improve the matching accuracy of algorithms reaching only pixel accuracy to sub-pixel accuracy. This was applied successfully for the normalized cross correlation (NCC) dense matching algorithm by Debella-Gilo & Kääb (2011), Heid & Kääb (2012) and Redpath, et al. (2013) for the calculation of deformation vectors derived from dense matching of remote sensing data illustrating glacier creep. The applied dense matching as well as the large amount of applied up-sampling by a factor of 10 in Redpath, et al. (2013) cause a huge computational effort. This stands in contrast to the high-resolution feature point matching approach developed in this work which neglects dense disparity information in favor of computational efficiency and accurate point matches.

In contrast to the latter application and similar to Xu, et al. (2012) the applied point matching and refinement algorithms described in Section 4.3 and 5.2 already use sub-pixel accuracy refinement methods. The application of image up-sampling as a pre-processing step will therefore increase the algorithms' accuracy in a similar range as for Xu, et al. (2012) which is around 20%. As already mentioned this gain comes at the cost of a greatly increased processing time. The method can therefore only be applied if there are no time restrictions on the processing.

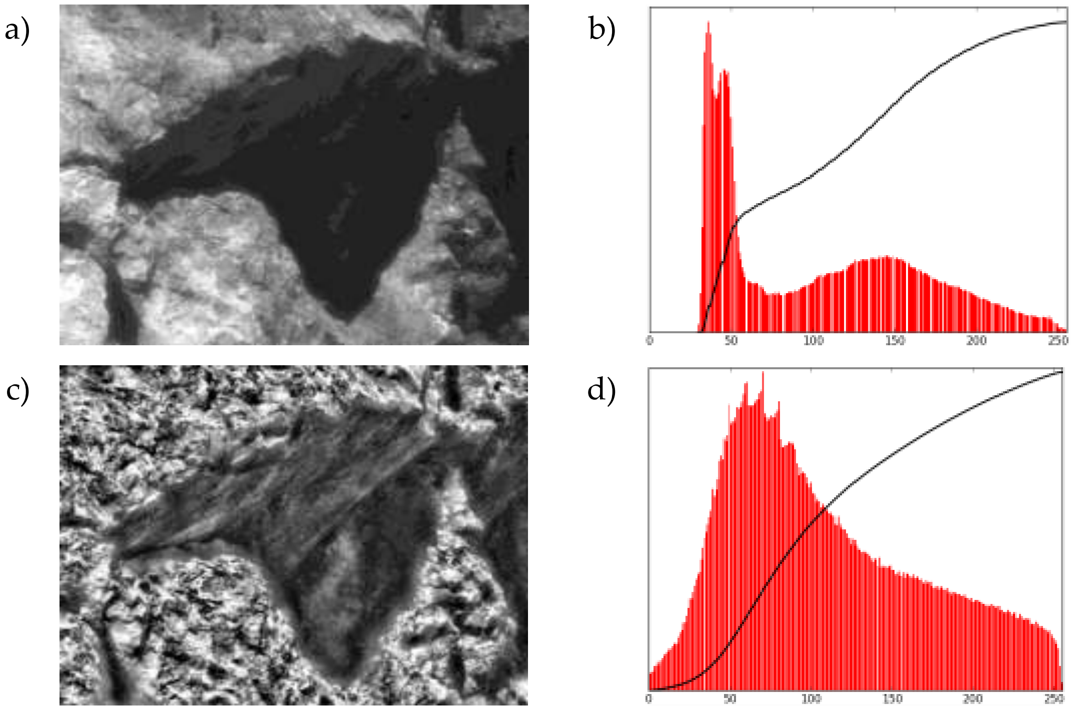

Figure 17: a) image with cast shadow and b) result of adaptive histogram equalization with corresponding histograms (red) and cumulative histograms (black) in b) and d)

The second applied method, adaptive histogram equalization, divides the image in rectangular contextual regions or tiles in which the optimal contrast is enhanced with standard histogram equalization as outlined in Section 4.2. This tiling process would create visible borders at the edges of contextual regions. To avoid this effect only the center pixel of each tile gets the value from the transformation function derived with the local histogram equalization. All other pixels get interpolated values of the result of the transformation functions of their adjacent tiles. This creates a continuous transition between equalized neighboring contextual regions (Zuiderveld, 1994, p. 475).

Adaptive histogram equalization tends to generate or amplify noise in homogeneous areas which are characterized by a peak in the histogram caused by a lot of pixels falling into the same intensity range. Contrast limited adaptive histogram equalization solves this problem by a contrast limiting or maximum contrast clipping factor. This factor is applied to high peaks of equalized histograms on each contextual region by limiting the maximum amount of intensity values falling into a certain local histogram bin. Clipped values are then equally distributed over the whole local histogram. This leads to a minimization of introduced noise (Zuiderveld, 1994, pp. 476-477).

This approach leads to similar results as the Wallis filter (Wallis, 1976) commonly used in photogrammetry. Purpose of the Wallis filter is to adaptively enhance contrast on both ends of the brightness range by forcing image contrast and mean to resemble given values in local image patches. It is a generalization of the local standard deviation based image enhancement which adjusts local differences so that each patch has the same given standard deviation. Wallis (1976) expanded this method by two weighting parameters, the contrast gain constant to increase or decrease the amount of applied enhancement and the brightness forcing constant which defines the weight of the local mean within a tile in contrast to the global mean. To ensure smooth transitions between tiles the results are interpolated from the neighboring block centers as in local histogram equalization.

Rock slope monitoring is strongly influenced by varying illumination conditions causing cast shadows with “false” edges represented as sudden changes in the intensity values of the images which are moving with the changing angles of the sun. Adaptive histogram equalization is an excellent possibility to locally enhance contrast in shadow areas and therefore decrease the effect of cast shadows. This effect is illustrated in Figure 17 showing a section of an image with an overhanging rock-face and the caused shadow. Global histogram equalization might even amplify the shadow borders (Figure 16) whereas the local adaptive variant of the method increases the similarity of shadow and non-shadow regions.

4.4 Geometric correction

Geometric correction or geometry based pre-processing is the rectification of an image following a given model. It can be used to apply a transformation changing image rotation, scale, skewing angle or more complex distortions such as a full perspective transformation or lens distortions.

The application of transformations generally generates new pixel positions in between the original image raster. To generate integer values for the output image raster intensity interpolations of neighboring non-integer values have to be used. However, every interpolation step introduces errors causing a slight degradation of image quality.

In the IATS based processing chain there are two possibilities to correct camera lens deficiencies like radial and tangential distortion. The first is to run all image processing operations on the original images and apply the adjustment parameters found in the calibration at the stage of 3D point reconstruction. The second possibility is to calculate an undistorted image from the found distortion model and then run all other processing methods on the corrected image. This can help to improve the quality of image matching but comes at the cost of interpolation errors in the distortion correction step. As the lens distortion of the IATS is relatively small (maximum ~1.5 pixel at image borders) the first method is generally used and the respective parameters are added to the results of the point detection step.

The second application of geometry based pre-processing in the application at hand is the use of the known mutual orientation of the two IATS stations to calculate epipolar corrected images which only have disparities in row direction. This can also help a consecutive image matching as the search space in the image is restricted to only one dimension. However, the known epipolar relation between images can also be exploited at the point matching or mismatch detection stage (Section 3.2.3) without interpolation errors and additional processing caused by image warping.

5 Feature based methods for the detection of homologous non-signalized points

5.1 Introduction

The key components for the calculation of accurate 3D points using spatial forward intersection is the accurate image orientation acquired from the system calibration as described in Section 2.3 and equally important the detection of homologous points in corresponding images of a stereo set-up with highest possible accuracy. This process is the second part of the developed processing chain described in Section 3.2.3.

Although the images are saved in the form of rasterized intensity values it is possible to detect and match points situated in between this given grid of pixels by interpolating texture from their local neighborhood. This is called sub-pixel accuracy and is one of the main requirements to this task of homologous point detection.

To build a processing chain delivering the best possible results all available technical options were analyzed and evaluated. As mentioned previously there are two conflicting main requirements to be fulfilled which are highest possible accuracy and lowest possible runtime. To find the best solution the process of homologous point detection was split into different stages (Figure 14). These are defined as follows.

- Detection of distinct non-signalized points based on well-defined statistical extrema within the image content
- Description of these points using their local intensity neighborhood forming feature vectors
- Matching of these vectors to recognize homologous points

For each of these steps state of the art image processing algorithms were analyzed and evaluated. The main purpose was to find the methods performing best for the individual tasks keeping in mind the above stated requirements. The resulting system thus has to make use of complementary characteristics of the different hereby evaluated algorithms. The required degree of accuracy or the available runtime have to be configurable leading to different configurations of the processing chain for different tasks.

5.2 Local feature detection

Local features are image patterns forming distinctive extrema in their local neighborhood and the whole image. The following section describes the local feature detection algorithms that were evaluated in the course of this work. It lists general requirements and concepts regarding local features and then gives information of a set of state of the art algorithms. Implementations of the described methods will be evaluated regarding repeatability and sub-pixel localization accuracies under changing environmental conditions as well as runtime requirements. All of this is done in accordance with the processing workflow defined in Section 3.2.3 in order to find the most suitable approach for the different use cases.

5.2.1 Requirements, concepts and terminology

To be able to match data-set over longer time series containing multiple measurement epochs the application at hand requires detected points to have a high invariability towards changed illumination conditions or geometrical transformations of the images like changed viewpoints. This requirement towards high repeatability under varying conditions can generally only be met through the use of image features such as second order derivations which are more complex to compute in comparison to simple intensity value differences. Different classes of feature detectors show varying characteristics which may complement each other. This calls for a careful evaluation of a well-balanced ratio of repeatability and accuracy versus computational effort.

In the literature different terms are used such as interest point, interest region or local feature. Similar to a point defined in geometry an ideal local feature would describe a location without a spatial extent. However, an image consists of a pixel raster defining the smallest unit which always has an implicit spatial extent. The calculation of sub-pixel values can only be done using interpolation from the surrounding neighborhood which again assigns the found feature the spatial extent of this neighborhood. If this extent is ignored in further processing like 3D reconstruction the local feature is generally called interest point, if the spatial context is used to describe the found feature it is typically called region. In accordance with the literature in the following the term local feature will be used describing interest points as well as regions (Tuytelaars & Mikolajczyk, 2008).

According to Tuytelaars & Mikolajczyk, 2008 a suitable local feature should have the following properties.

- *Repeatability:* in multiple images of the same scene under varying viewing conditions a maximum of features should be re-detected.
- *Distinctiveness:* the local neighborhood describing a feature should exhibit high entropy.
- *Locality:* in order to avoid possible occlusions the feature should be local and planar. This is a competing property with distinctiveness as the intensity variation is reduced the more homogenous a feature is.
- *Quantity:* a sufficient number of features should be detectable depending on the application.
- *Accuracy:* localization of the feature in scale space and image location should be as precise as possible.
- *Efficiency:* the feature should be computed inexpensively.
- *Invariance:* expected deformations are to be modeled in the descriptor making it invariant to these transformations. Higher invariance may lead to a decrease in distinctiveness.
- *Robustness:* features should be robust against deformations such as image noise or blur. This may reduce feature distinctiveness as well.

Different feature detection methods are categorized according to the type of features that are found. In the following sections the most important representatives of the corner based and blob based methods are described.

5.2.2 Corner based methods

The definition of corners as detected by corner based feature detection methods are image positions with a high curvature. This can be found at a point of intersection of two edges, close to different dominant edge directions caused by highly textured regions but also at object or shadow boundaries. The found features do not have to correspond to projections of real world 3D corners. However, the important criterion is not the detection of real corners but conformance with the previously defined properties of local features.

Due to the tendency of 2D corners to be detected at occlusion boundaries where the strongest intensity variations appear the features often do not comply with the *Locality* criterion as features are not planar but contain foreground as well as background components. This effect also limits viewpoint invariance as changed viewing angles cause significant variations especially at non-planar regions such as object boundaries. A further problem is the identification of scale as corner positions are rather stable over multiple scale space levels. The

Localization of found features on the other hand is more accurate at edge positions (Tuytelaars & Mikolajczyk, 2008).

5.2.2.1 Harris operator

Harris & Stephens (1988) proposed a corner detector based on a matrix of image gradients, the second moment matrix or structure tensor. This matrix is also called auto-correlation matrix and often used to describe local texture representing gradient distribution around a feature (Equation (5.3)). By integrating a scale parameter the method was expanded to a scale adapted version.

$$I_x(x,y,\sigma_D) = \frac{\partial}{\partial x} g(\sigma_D)\, I(x,y) \qquad I_y(x,y,\sigma_D) = \frac{\partial}{\partial y} g(\sigma_D)\, I(x,y) \tag{5.1}$$

$$g(\sigma) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \tag{5.2}$$

$$M = \sigma_D^2 g(\sigma_I) \begin{bmatrix} I_x^2\,(x,y,\sigma_D) & I_x(x,y,\sigma_D) I_y(x,y,\sigma_D) \\ I_x(x,y,\sigma_D) I_y(x,y,\sigma_D) & I_y^2(x,y,\sigma_D) \end{bmatrix} \tag{5.3}$$

The gradients are computed with kernels of the scale σ_D depicting the window size. If this parameter is chosen too small the result may be affected by noise if it is too large finer corners may be lost. The scale factor σ_I defines a Gaussian smoothing kernel responsible for the size of the corners. Higher smoothing values correspond to coarser scales and larger corners.

Corners are found at positions where both eigenvalues of the matrix are large. Harris & Stephens (1988) defined a corner response measure H as given in equation (5.4) tunable with the parameter κ.

$$H = \lambda_1 \lambda_2 - \kappa\,(\lambda_1 + \lambda_2)^2 = det(M) - \kappa\, trace^2(M) \tag{5.4}$$

Corners can be extracted by a thresholding of the parameter κ. Sub-pixel accuracy is reached in the neighbourhood of a point from a quadratic approximation of maxima in the corner response function H (Tuytelaars & Mikolajczyk, 2008).

The strength of this method is the high localization accuracy whereas the complex processing required depicts its main weakness.

5.2.2.2 *SUSAN – Smallest Uni- value Segment Assimilating Nucleus*

The mentioned Harris operator works well for a wide range of applications but requires a quite complex and time consuming computation. Therefore methods were developed taking a contrary approach to corner detection, which were more focused on simplicity and a resulting faster computation. One of these methods is called SUSAN (Smallest Uni- value Segment Assimilating Nucleus, Smith & Brady, 1997).

A circular mask with a center pixel, described as "nucleus" is evaluated in a way that the intensity of each pixel within the mask is compared to its nucleus. Equation (5.5) shows a simplified comparison of each pixel r with the nucleus r_0 within the mask. The numbers of pixels with a similar intensity as the center are forming an area called the "Uni- value Segment" (USAN). This value ($n(r_0)$ in equation (5.6)) is largest when the nucleus lies in a homogeneously textured area, it is strongly reduced closer to an edge and reaches a minimum when lying on a corner. In the case of a "Smallest USAN" the nucleus differs strongly from the surrounding mask and represents a potential local feature. SUSANs are found by a comparison of n with a threshold of g empirically defined as 0.75 times the maximum possible value of n.

$$c(r, r_0) = \begin{Bmatrix} 1\ if\ |I(r) - I\ (r_0)| \leq t \\ 0\ if\ |I(r) - I\ (r_0)| > t \end{Bmatrix} \tag{5.5}$$

$$n(r_0) = \sum_r c(r, r_0) \qquad\qquad R(r_0) = \begin{Bmatrix} g - n(r_0)\ if\ n(r_0) < g \\ 0\ otherwise \end{Bmatrix} \tag{5.6}$$

The algorithm does not rely on image derivatives and also does not require noise reduction and can be computed inexpensively. However, due to the same properties it is strongly susceptible to changed illumination conditions. The method was also applied for noise reduction and edge detection purposes. Found corners are often located on or close to edges reducing the repeatability of the feature and increasing noise sensitivity (Smith & Brady, 1997).

5.2.2.3 *FAST*

FAST (Features from Accelerated Segment Test) was developed targeted at real time applications and enhanced the methods applied in the SUSAN detector (Smith & Brady, 1997) to further speed up computation. The method was proposed by Rosten & Drummond (2005). Each evaluated pixel is compared to the center of a circular mask with a fixed radius of 16 pixels and classified as one of three states: darker, similar or brighter. Machine learning methods are applied to generate decision trees containing the results of previously tested pixels on sets of training images. These decision trees can be converted into compile-able code of long if, then, else combinations forming the executable feature detector.

This detector is not scale invariant but very fast to compute and offers accurate feature localization. Computation times where found to be 30 to 40 times faster than other methods such as DoG (Section 5.2.3.2) or Harris (Section 5.2.2.1, Rosten & Drummond, 2006). In its original form the detector does not compute feature orientation and scale. These two deficiencies were eliminated in extensions by Leutenegger, et al. (2011) and Rublee, et al. (2011). They added feature orientation calculation and scale invariance by computing the detector on multiple image pyramids with a successive non-maxima suppression. The resulting feature detectors including feature descriptors of the same name are called BRISK (Binary robust invariant scalable keypoints) and ORB.

5.2.3 Blob based methods

In the field of visual perception so called blobs describe cylindrically assembled color sensitive groups of neurons in the visual cortex which were first identified in the year 1979 (Wong-Riley, 1979). Another definition of the term blob is small drop of color. In the field of computer vision it describes areas in an image featuring a unique, homogenous property clearly separating them from their environment. That way they illustrate mathematically defined extrema in the image. Local features that fall into this category generally consist of small areas of undefined shapes with in interpolated maximum defining their exact center. This means that they exhibit complementary properties in comparison to corner based features and are generally more robust to scale and viewpoint changes. On the other hand the localization accuracy of blob like features is generally lower than for corner based methods.

5.2.3.1 Determinant of the Hessian

The second-order partial derivative of a function forming a square matrix is called Hessian. In image processing the intensity function $I(x, y)$ at the position x, y at scale σ_D is used to define the Hessian as follows (Bay, et al., 2008).

$$L_{xx}(x, y, \sigma_D) = \frac{\partial^2}{\partial x^2} g(\sigma_D)\, I(x, y) \quad H(x, y, \sigma_D) = \begin{bmatrix} L_{xx}(x, y, \sigma) & L_{xy}(x, y, \sigma) \\ L_{xy}(x, y, \sigma) & L_{yy}(x, y, \sigma) \end{bmatrix} \tag{5.7}$$

L_{xx}, L_{xy}, L_{yy} contain the discretised cropped Gaussian smoothed second order image derivatives with a Gaussian kernel of size σ_D. The matrix describes the change of a normal to an isosurface. That way important shape information is captured from the image. Local image features can be detected based on the

Hessian matrix using its determinant or its trace. Using the determinant is advantageous in the sense that it finds features less likely to lie near contours or straight edges improving invariance towards noise or scale changes (Mikolajczyk & Schmid, 2001 and Tuytelaars & Mikolajczyk, 2008).

Bay, et al. (2008) developed a method significantly speeding up the feature detection process by coarsely approximating the different Gaussian filters. Due to necessary discretization and cropping of any image filter optimal properties are never reached. Following the promising results Lowe, 2004a had with efficiently approximating the Laplacian of Gaussian detector, Bay, et al., 2008 used simple box filters as illustrated in Figure 18 to approximate the Gaussian second order partial derivative filters (Bay, et al., 2008).

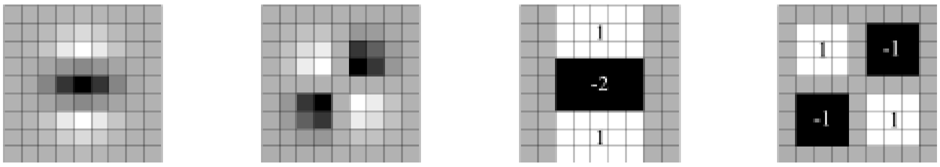

Figure 18: Discretized Gaussian filter in Lyy and Lxy direction and coarse approximation filters
(Bay, et al., 2008)

These box filters can be computed inexpensively using integral images. In an integral image each pixel represents the sum of all original image elements in the rectangular area between the origin and the pixel (see Equation (5.8)).

$$I_{\Sigma(x,y)} = \sum_{x'=0}^{x} \sum_{y'=0}^{y} I(x',y') \tag{5.8}$$

Once this representation of image values is calculated an integral over a rectangular image region of any size can be calculated with four memory accesses and three additions. The upper right and lower left corner values of the box filter in the integral image are subtracted by the lower right value and the upper left value is added to form the integral. The required processing time for these operations does not depend on the size of the calculated region. This means that for the detection of larger scaled features no image pyramids have to be calculated, the only thing that has to be changed is the size of the box filters. Surprisingly it could also be shown that the results of the approximated filters are comparable or even better than with original Gaussian kernels as shown in Figure 18 (Bay, et al., 2008).

An additional advantage of this method is the possibility to separate found features based on the sign of the trace of the Hessian matrix. It describes either light blob like features on a dark background or vice versa. This can be used to separate the sets of points in a later feature matching process further reducing required processing time and improving matching accuracy.

All these properties make the speeded up Determinant of the Hessian detector by Bay, et al., 2008 one of the best available solutions for feature detection. The unfortunate downside of the method is the existing US patents (Funayama, et al., 2007) restricting its application. Academic use is generally tolerated but commercial implementations have to acquire a license from the patent holder. To the authors knowledge there is no corresponding patent in Europe.

5.2.3.2 Laplacian of Gaussian and Differences of Gaussians

The Hessian matrix as defined in the previous section consists of second order image derivatives and can be used to detect blob like features from its determinant as done in the described Determinant of the Hessian detector or from its trace. The trace of the Hessian is identical with the Laplacian of the original intensity function. The resulting Laplacian of Gaussian (LoG) is a general edge detector finding closed contours as described by Marr & Hildreth (1980). Thresholding can be used to detect local features which, however, have the disadvantage of lying on or close to edges making them sensitive to noise. This effect can be reduced by an evaluation of the eigenvalues of the full Hessian matrix to filter out points similarly to the approach taken in the Harris operator (Section 5.2.2.1).

Lowe, 2004a developed a method to approximate the Laplacian of Gaussian via the calculation of Differences of Gaussians. This simple and effective method computes multiple versions of the image with different levels of Gaussian smoothing and pairwise subtracts the lower levels of smoothing from higher levels (Figure 19). The results resemble the Laplacian of Gaussian but can be computed with a fraction of the computational costs.

Figure 19: Subtraction of different smoothing levels approximating Laplacian of Gaussian (Tuytelaars & Mikolajczyk, 2008)

The scale of the points is defined based on the size of the used Gaussian kernels. To avoid ambiguities points found in adjacent scales are discarded using a 3x3x3 non-maxima suppression that ensures that the found feature forms a maximum within the 3x3 neighborhood of the current scale and the according neighborhoods of the higher and lower scale. To eliminate edge responses from the found local features the ratio of the eigenvalues of the Hessian is evaluated. The derivations forming the Hessian are estimated from differences of neighboring sample points (Lowe, 2004a). This feature detector is generally used in combination with Lowe's famous local feature descriptor SIFT (scale invariant feature transform, see Section 5.3.2.1).

The detector is protected by the same US patent as the SIFT descriptor (Lowe, 2004b) creating similar conditions as already mentioned for the speeded up Determinant of the Hessian detector (Section 5.2.3.1).

5.2.3.3 *CenSurE – Center Surround Extremas*

The Center Surround Extrema (CenSurE) feature detector was specifically designed for the task of Visual Odometry and Structure from Motion for off-road vehicles. The estimation of camera positions requires stable and accurate features over a large range of scales to accurately estimate movement trajectories. The available detectors generally use image pyramids to detect features at higher scales, however, the required smoothing is accompanied with a reduction in image resolution and therefore feature localization accuracy. The idea behind CenSurE is to detect features at all scales at original image resolution (Agrawal, et al., 2008).

While Lowe, 2004a approximated the Laplacian of Gaussian from differences of Gaussian, Agrawal, et al., 2008 used even simpler bi-level center surround filters multiplying the image with -1 or 1. The most accurate filter exhibiting the largest rotational invariance would be circular as shown in Figure 20 and is most computationally expensive to compute. The approximations octagon, hexagon and box have decreasing degrees of rotational invariance but can be calculated efficiently using integral images (as described in Section 5.2.3.1). The computational cost rises with the degree of complexity of the used filter.

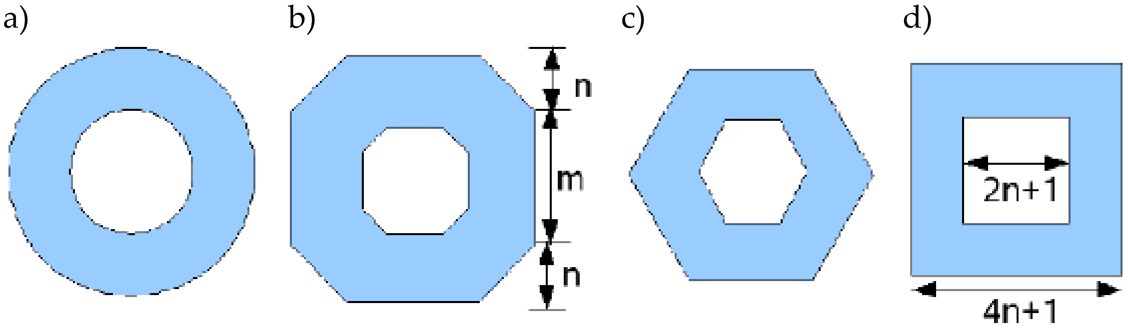

Figure 20: Different approximations of the Laplacian of Gaussian from bi-level filters a) Circular symmetric b) Octagon c) Hexagon d) Box (Agrawal, et al., 2008)

While simple box filters can be calculated easily using standard integral images, the slanted edges of octagon and hexagon filters require special integral image representations reflecting the respective angles. Equation (5.9) describes the calculation of slanted integral images. If α is set to zero rectangular integral images are calculated, the summed areas slant left if the parameter is set to a value below zero and right if it is above zero. The filters can be split into trapezoid regions consisting of filter parts with the same slant which can be calculated from the integral images requiring only three additions (Section 5.2.3.1). The hexagon can be split into two trapezoids, the octagon into three. This determines the computational cost of calculating each filter.

$$I_{\alpha(x,y)} = \sum_{y'=0}^{y} \sum_{x'=0}^{x+\alpha(y-y')} I(x', y') \tag{5.9}$$

The Harris corner measure is then used to suppress filter responses on edges and lines (equation (5.3)). It is computationally more expensive than the Hessian used in other detectors but is only computed for a small number of points above a certain threshold and was found to be better at suppressing edge responses (Agrawal, et al., 2008).

The CenSure operator is scale invariant and fast to compute. It has the additional benefit that the features are detected at different scales in the original image resolution. This means that no interpolation over the scale space is necessary and points at higher scales are better localized. Experimental results for visual odometry data-sets showed that CenSurE with an octagon filter outperformed Harris, DoH, DoG and other detectors in terms of runtime, stability as well as accuracy due to the longer and more accurate trajectories formed from features tracked over large scale changes (Agrawal, et al., 2008).

5.2.3.4 Maximally stable extremal regions

The Maximally stable extremal regions (MSER) detector finds blob like features but is strictly speaking not a blob based detector but as already contained in its name: "region based". Region based detectors are generally defined as finding homogenous image regions which may contain the above mentioned blobs but also larger and more complex structures. The main advantage of these methods is their localization accuracy as the center of larger, more complex shapes can be calculated very accurately. They are more complementary to corner based methods than to blob based methods as similar features are found.

MSER developed by Matas, et al. (2002) detects regions defined by connected boundaries resulting in an exact localization. It works best on images which can be segmented and contain homogenous regions which are well divided. This makes the detector more suitable for urban environments or generally scenes containing different well-defined objects. It is therefore prone to image blur causing separate regions to appear connected. A further disadvantage is the small number of features that is generally found in each image. The main advantage of the method is its high location accuracy and its invariance to affine image transformations (Tuytelaars & Mikolajczyk, 2008).

5.3 Feature description

To be able to find homologous points in different images of varying viewpoints the descriptors representing the local neighborhood of the feature can be used. The resulting feature vectors can be compared resulting in correspondences which can be used in numerous applications such as image matching, structure from motion, visual odometry, image panorama calculation and many more. There are multiple feature descriptors available exhibiting different degrees of invariance towards changes in the image data based on various properties such as image intensity values, image texture, gradients or color. This work will be focused on descriptors based on gray value input data. This section explains the feature description step and compares different descriptors.

5.3.1 Requirements and concepts

There are many descriptors available capturing the information contained in the intensity values of the local neighborhood of a defined feature position. The main requirements are the specified feature and a local neighborhood containing a sufficient degree of information for a distinctive description.

The easiest version of a descriptor would be a vector of those intensity values itself which could then be compared to others using a cross correlation. This results in long feature vectors and costly computation. Distribution based descriptors try to capture image properties such as texture or gradient information in histograms. This can help to reduce the complexity of the required descriptors while still capturing a maximum of the information contained in the local neighborhood. Another class of methods is formed by spatial frequency based descriptors, however, they are rarely used as spatial relations are not explicit in the frequency domain and their local use is highly complex. Another method to approximate and describe a feature is the use of differential descriptors wherein local derivatives up to a specified order describe the neighborhood (Mikolajczyk & Schmid, 2005).

The quality of the descriptor can be measured using the three parameters repeatability, matching accuracy and processing time. Repeatability describes the ratio of detected correctly matched features over the total number of correspondences. The matching accuracy is defined as the percentage of correct matches within the total amount of matches.

5.3.2 State of the art algorithms

The following section contains a description of state of the art algorithms with the best performance according to the literature and conducted analysis. The ones with the highest performance in terms of detections rates are more closely described other tested descriptors leading to poorer results are mentioned by name only including reasons why they do not work as well.

5.3.2.1 SIFT

SIFT stands for Scale Invariant Feature Transform and was developed by Lowe (2004a). Many other feature descriptors are based on similar principles but SIFT tends to achieve the highest overall matching rates over a wide range of applications. The fact that the method is not tailored to a certain task represents its main advantage. Other descriptors may be better suited in certain cases regarding computational complexity or even accuracy but SIFT is the best general purpose descriptor.

Lowe, 2004a also developed an optimized feature detector for SIFT, the previously described Difference of Gaussian (DoG) detector (Section 5.2.3.2) which delivers accurate feature locations including scale information corresponding to a certain level of Gaussian smoothing of the image. The first step of the SIFT descriptor computation is the calculation of the dominant gradient orientation around the feature at the current scale to detect its orientation. All further processing steps are executed relative to the orientation and scale of the feature.

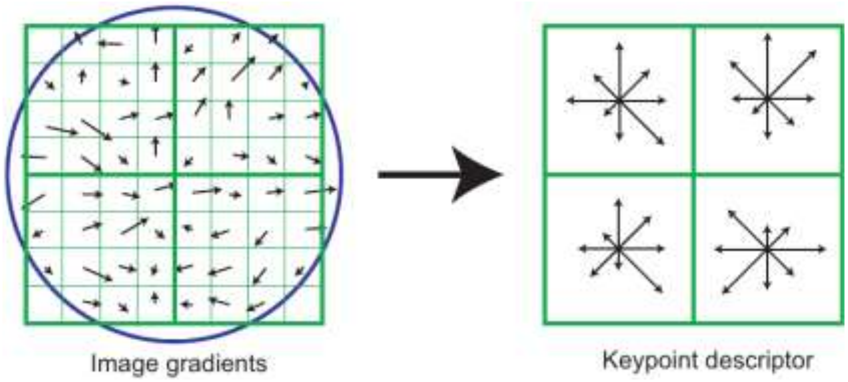

Figure 21: Simplified SIFT feature descriptor, local gradients combined into histogram of oriented gradients (Lowe, 2004a)

The descriptor is formed by local image gradients which are sampled at a defined raster around the feature. The gradient values are weighted with a Gaussian to gradually reduce the influence of gradients from the center towards the sides of the descriptor window. Sub-regions of 4x4 gradients are aggregated into histograms of orientated gradients (HoG, see Figure 21). The eight values of each histogram represent the magnitude of the sum of the gradients of a certain direction. SIFT generally uses 16x16 sampling points leading to 4x4 histograms of 8 values and a total descriptor vector length of 128 values. The final feature vector is normalized to unit length to compensate for illumination changes. Image contrast changes represented by a constant shift in intensity values is cancelled by this normalization step, constant brightness changes will not affect the descriptor at all as the gradients are calculated relatively from intensity differences (Lowe, 2004a).

The descriptor is invariant to scale and orientation changes and can compensate for changing illumination conditions by capturing the distribution of gradient magnitude and direction rather than intensity values. Compared to other descriptor such as SURF (Section 5.3.2.2) the computation of the descriptor is computationally more expensive. The same applies to the matching process due to the relatively long feature vector.

SIFT with its HoG descriptor has become the reference feature description method every new descriptor has to measure up to. There are multiple implementations available but it remains to be mentioned that the method is patented in the US (Lowe, 2004b) restricting its application. Academic use is generally tolerated but commercial implementations have to acquire a license from the patent holder. To the authors knowledge there is no corresponding patent in Europe.

5.3.2.2 SURF

The second most used feature descriptor for gray scale images is SURF (speeded up robust features). It was developed by Bay, et al., 2008 at the ETH-Zurich and is based on similar principles as the SIFT feature descriptor. The main difference is the strong focus on efficient computation and a limitation of runtime without compromising matching accuracy and repeatability compared to SIFT.

Bay, et al., 2008 developed a feature detector for his SURF descriptor based on an efficient approximation of the Determinant of the Hessian which relies on the use of box filters which can be computed inexpensively using integral images (Section 5.2.3.2). Information about the position of the feature within the scale space is also retrieved by the detector.

The descriptor is based on simple binary first order Haar Wavelet filter responses instead of image gradients using the integral images to speed up computation. The filter responses are weighted with a Gaussian and used to calculate the feature orientation based on their dominant direction.

a) b)

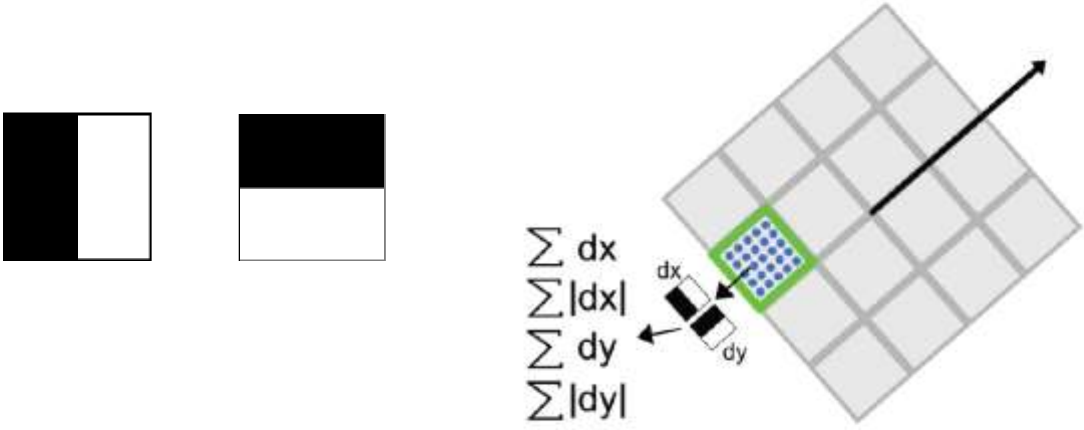

Figure 22: a) Haar Wavelet filters with weights of -1 and 1 in horizontal and vertical direction and b) the SURF descriptor derived from filter responses (Bay, et al., 2008)

The descriptor is calculated in relation to the found feature orientation splitting the descriptor window defined by the scale of the feature into four rectangular sub-regions. These are regularly sampled to calculate 5 x 5 Haar Wavelet responses (Figure 22a) in each sub-region. The sum and the absolute value of the sum of filter responses in horizontal and vertical direction relative to the feature orientation are used to form the feature descriptor (Figure 22b). This leads to 4 x 4 sub-regions each contributing 4 entries resulting in a feature vector length of 64 values (Bay, et al., 2008).

The shorter feature vector of only 64 values compared to 128 in the SIFT descriptor further reduces processing time and makes SURF generally more than five times faster than SIFT. However, the shorter descriptor leads to a reduction of information and also reduces sensitivity to image noise increasing with the feature vector length. For this reason Bay, et al. (2008) could show that their algorithm not only outperforms SIFT in terms of computational efficiency but also in certain cases in terms of accuracy and repeatability. Generally SIFT tends to be more accurate than SURF by a small margin which is made up for by the reduced runtime of the latter.

The SURF descriptor is protected by the same US patent as its detector (Funayama, et al., 2007 and Section 5.2.3.2) creating the same conditions as already mentioned for SIFT (Section 5.3.2.1).

5.3.2.3 BRISK

The next generation of feature descriptors after SIFT and SURF tried to achieve similar matching accuracy and repeatability at much lower computational costs. A big step in this direction was done by the introduction of a descriptor based on a bit string describing the results of intensity value comparison tests within a descriptor window instead of a vector of floating point values. This descriptor was developed by Calonder, et al. (2010) and is called BRIEF (binary robust independent elementary features). The use of bit strings drastically reduces the time needed to do a full matching of two feature vectors as the comparison can be done using the hamming distance defining the amount of different values of two bit strings. This comparison can be computed inexpensively on modern CPUs and makes the matching process up to an order of magnitude faster than the matching of SURF feature vectors based on the calculation of Euclidean vector distances (Calonder, et al., 2010). In an effort to improve the BRIEF descriptor to make it more distinctive and even less computationally expensive, BRISK was developed by Leutenegger, et al., 2011.

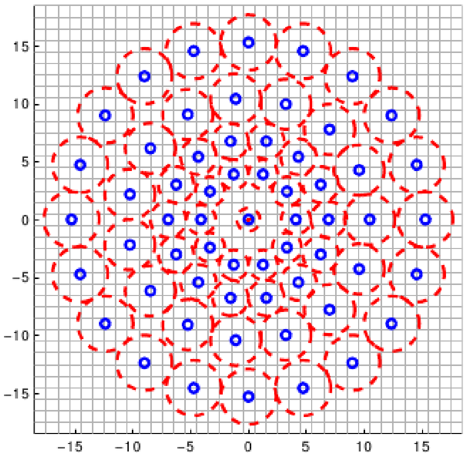

Figure 23: BRISK descriptor sampling pattern. Red circles describe Gaussian kernel sizes used to smooth intensities at the blue sampling locations (Leutenegger, et al., 2011)

BRISK makes use of a regular spaced concentric sampling pattern around a given feature position (Figure 23). The pattern resembles the sampling of the DAISY descriptor (Tola, et al., 2010). Image intensities at the sampling positions are smoothed with a Gaussian of σ proportional to the distance to next point. Pairwise combinations of all points within the sampling pattern form two classes, short distance and long distance combinations separated by an empirically determined threshold proportional to the scale of the feature (Tola, et al., 2010 Leutenegger, et al., 2011). A local orientation of the feature is calculated estimating the overall characteristic pattern orientation from estimated gradients of the long distance combinations only. As for all rotationally invariant descriptors the extraction of the descriptor values is done in relation to the found orientation. The descriptor itself consists of comparisons of the smoothed intensity values of the short distance combinations. The vector is set to one or zero depending on whether the first or second sample intensity value is larger. That way a bitstring of 512 values or 64 Bytes is formed (Leutenegger, et al., 2011).

These vectors can be matched calculating their Hamming distance with a simple XOR operation and a following bitcount. The amount of same values is the measure of similarity. Leutenegger, et al., 2011 report that the matching accuracy as well as the repeatability is comparable to SIFT and SURF while being over an order of magnitude faster in feature description as well as matching.

5.3.2.4 FREAK

A further enhancement of the concept of bit string based descriptors was developed by Alahi, et al. (2012) and is called Fast Retina Keypoint (FREAK). According to the authors the sampling pattern of the descriptor was inspired by the human visual system.

The circular sampling pattern of the descriptor is similar to BRISK with the main difference of an exponential increase in the distance of sampling points from the center similar to the amount of photoreceptor ganglion cells in the human retina. Another important difference to BRISK is the overlap of receptive fields (Figure 24, left). The 512 sampling point intensity comparison combinations used to generate the descriptor were found from the analysis of training data. The combinations were segmented into four main patterns ranging from coarse to fine (Figure 24, right). The first quarter of the descriptor consisting of a 16 Byte bitstring involves combinations furthest from the center. This can be used to speed up the consecutive matching step even further. Due to hardware implementations on modern CPUs 16 Byte can be compared in one parallel command. More than 90% of candidates can already be discarded by a comparison of only the first 16 Byte of the descriptor leading to a substantial reduction of necessary matching operations (Alahi, et al., 2012).

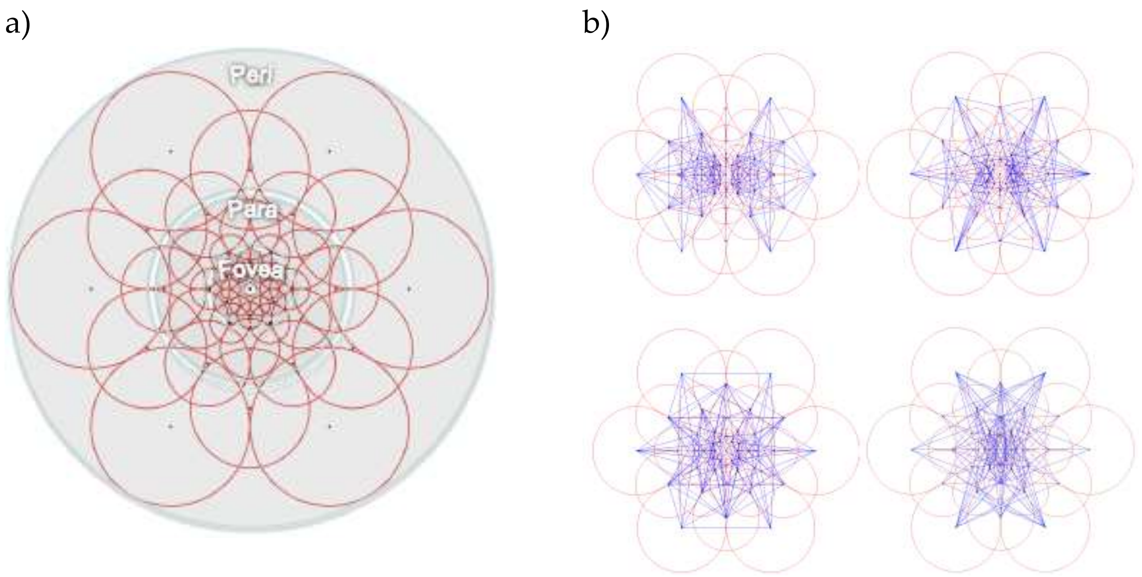

Figure 24: a) FREAK descriptor sampling pattern with corresponding Gaussian smoothing kernel in Analogy to retinal photoreceptor cells and b) intensity combinations of sampling points leading to highest descriptor variance (Alahi, et al., 2012)

The feature extraction as well as the matching step is around 30% faster than BRISK which in term is more than an order of magnitude faster than SURF. According to the authors and external evaluations matching accuracy and repeatability is higher than for BRISK, SIFT and SURF in most officially available test data-set (Alahi, et al., 2012).

5.3.2.5 Others

Other recently developed descriptors that score well in one or all of the categories matching accuracy, repeatability and computation time will be mentioned shortly in the following section for the sake of completeness.

ORB (oriented robust BRIEF) uses the BRIEF descriptor relative to a given orientation increasing descriptor robustness by not choosing sampling point combinations arbitrarily but calculating optimal comparisons leading to the highest descriptor variance from training data (Rublee, et al., 2011).

In contrast to the new binary descriptors targeted at efficient computation a different class of methods was introduced focusing on matching accuracy and repeatability. These are LIOP (Local Intensity Order Pattern for Feature Description, Wang, et al., 2011), MROGH (Multisupport Region Order-Based Gradient Histogram, Fan, et al., 2011) and its extension MRRID (Multisupport Region Rotation and Intensity Monotonic Invariant Descriptor, Fan, et al., 2012) which was introduced to handle large illumination changes. These descriptors are intrinsically rotation invariant without having to be extracted relative to a feature orientation and are highly discriminative. This way they avoid the main source of error of other feature descriptors lying in the estimation of feature orientation. The main disadvantage of these descriptors is their high required processing time which is in the range of an order of magnitude higher than SIFT for LIOP which is the fastest of these methods. LIOP is in term around 5 times faster than MROGH and 2 times faster than MRRID (Miksik & Mikolajczyk, 2012). This means that for the application at hand these high level descriptors are computationally far too expensive and will be discarded in favor of modular post-processing for mismatch detection and accuracy refinement.

5.4 Feature matching

After the detection of local image features and the extraction of feature vectors distinctively describing the features' local neighborhood the final step is the matching of these vectors to find correspondences. The following section will describe the different possible methods and their implication on matching times and accuracy.

5.4.1 Brute force

Computing and comparing the Euclidian distance for floating point number descriptors or the Hamming distance for binary descriptors of each feature vector extracted from search data to each feature vector of reference data is referred to as brute force matching. This exhaustive search approach covers all possible feature combinations without any restrictions and results in a quadratic relation of computational effort and vector length.

A global thresholding of the resulting Euclidian or Hamming distances does not perform well in finding correspondences as some features show a much higher variation than others. Lowe (2004a) proposed the use of the concept of best and second-best matches without a global threshold on the Euclidian distance. The distance of the best match has to be significantly lower than the second-best match. This ratio can be used in the matching process to accept a higher number of matches at the cost of increased false matches or to dismiss correct matches with the benefit of lowering the amount of incorrect ones. Lowe (2004a) found that a ratio of 0.8 eliminates 90% of mismatches and only 5% of correct matches. This concept can equally be applied to binary descriptors matched based on their Hamming distance.

5.4.2 Nearest neighbor approximation

The nearest neighbor matching of large feature vectors resulting in problems of high dimensional space generally is the most time consuming part in the search for corresponding features. Exact nearest neighbor indexing algorithms do not provide a speedup over linear exhaustive search over around 10 dimensional spaces (LOWE, 2004A). This means that approximate algorithms have to be applied which can offer a large reduction of computational effort while still reaching accurate solutions.

These algorithms are based on k-d trees representing a structure that partitions data in a k-dimensional space (Freidman, et al., 1977). Similarity search operations can be executed very efficiently following the branches of the tree. According to Muja & Lowe (2009) the two algorithms best suited for the application of feature vector matching are searching hierarchical k-means trees with a priority search order and the use of multiple randomized k-d trees. Compared to linear exhaustive search these methods can speed up the matching process up to several orders of magnitude (Muja & Lowe, 2009).

5.4.3 Geometrical restrictions

Matching based on geometrical restrictions in the scope of this thesis refers to the known exterior orientation which can be used to construct epipolar lines which in term restrict the matching process to a small matching window. This greatly decreases the runtime requirements of a brute force matching and increases accuracy by restricting the search space.

In the case of IATS measurements the orientation of the images is available. This allows the calculation of epipolar lines in stereo-images (Hartley & Zisserman, 2000, p. 240) which can be further clipped by the approximate object distance to a rectangular matching window to drastically reduce the amount of features that have to be compared. This changes the computational complexity of the brute force matching process from quadratic to constant.

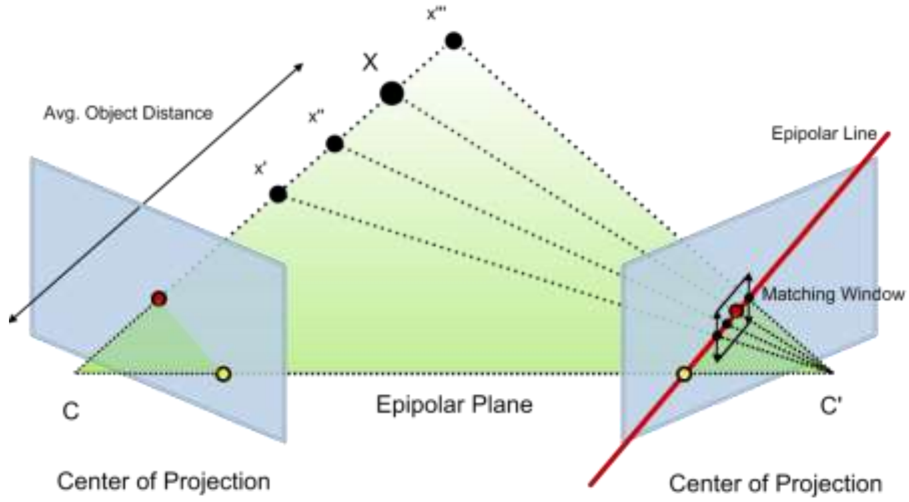

Figure 25: Epipolar constrained matching (Reiterer, et al., 2010)

Figure 25 illustrates the geometric constraints leading to the described matching window. The achievable speed-up can be described using some example figures: two lists of feature vectors of 25 thousand points each, require 625 million vector comparisons. Using the calculated constraints this process can be reduced to about 1.25 million comparisons, based on the assumption of an even point distribution in images of 2560x1920 pixel and a matching window of 200x50 pixels, resulting in a matching time reduction of over two orders of magnitude (Reiterer, et al., 2010).

5.5 Conclusion

According to literature the presented algorithms involved in the feature detection, description and matching process all have their individual strength and weaknesses. The requirements and conditions set by different applications may lead to different sets of algorithms that should be chosen to process the data. Furthermore, some described methods such as the epipolar restricted matching may not be applicable because of missing relative camera orientation in cases where processed data does not stem from IATS devices. Another more specific problem would be a commercial use of the developed software in the USA. This leads to a restriction of algorithms protected by US patents such as SIFT and its DoG detector or SURF with the fast DoH detector.

Algorithm performance as specified in the original papers as well as in different comparative survey studies was always tested under similar conditions for the sake of comparability. However, these test settings are generally not representative for the applications of geo-monitoring as studied in the work at hand. Also new methods are always described as outperforming state of the art algorithms in various ways. In objective comparison this generally only holds for carefully chosen sets of test data and accordingly optimized sets of processing parameters. Therefore an independent comparison of methods with specifically chosen data-sets representing the task of geo-monitoring is absolutely essential to be able to make an explicit statement about which algorithm to use for what kind of application including the accuracy and runtime that is to be expected.

The above described methods for feature point matching lead to results of varying degrees of sub-pixel accuracy. However, no mutual stereo information is used to refine the feature correspondence. The matching process just assigns local features in both images to their respective correspondences. This allows very fast processing compared to dense methods and optimally leads to stable features that can be tracked over time and larger degrees of various distortions in the images. If the matching accuracy should be increased an additional stereomatching process with the found feature positions as initial values has to be applied. This will be described in the following section.

Mismatches that may be present after the detection of homologous local features may still be present after applying the described methods. Strategies how to deal with this problem were implemented and are also described in Section 6.

6 Refinement and temporal tracking of homologous non-signalized points

After a successful detection of homologous image features potential mismatches have to be detected and excluded. If allowed by the time constraints of the current application the sub-pixel accuracy of the detected points can be refined. The next step of processing is the temporal tracking of homologous features. Different strategies of how this can be done will be described.

6.1 Detection of mismatches

The detection of mismatches, the so called outlier detection is an important part of the feature matching process, especially in the application at hand false positives should be reduced to zero wherever possible. This can be achieved using a very robust feature descriptor and a suitable feature vector matching method. However, in most cases the results will have to be checked in a post processing step dealing with outliers. The term "outlier" is here used as a synonym for mismatch. In order to decide what a "correct match" and what a "mismatch" is some kind of model has to be evaluated. Each measurement is either in agreement with this model or lies above a certain error threshold outside of the model. Hence, correct measurements are called "inliers" whereas mismatches are called "outliers".

6.1.1 Global geometrical restrictions

An intuitive first approach to filter outliers is the use of a known exterior orientation and the geometrical restrictions that come with it. This means that for a stereo system the fundamental matrix can be used to evaluate whether detected homologous points lie on common epipolar lines.

There are multiple possibilities to construct the epipolar lines from image orientations. One possibility is the direct use of the P matrices commonly used to describe camera orientations (Section 2.3.2). The projection of the center of camera one (described by P) into the second image (described by P') is defined by $P'C$ and leads to the epipole in the second image. The projection of an arbitrary image point x into the second image can be described with $x' = P'X = P'P^{+}x$, where P^{+} is the pseudoinverse of P. The epipolar line in the second image connects the epipole and the projected point with $l' = [P'C]_{\times} (P'P^{+})\, x = Fx$, where F is the fundamental matrix (Equation (6.1) and Hartley & Zisserman, 2000, p. 244).

$$F = [P'C]_{\times}(P'P^{+}) \qquad (6.1)$$

For the sake of completeness the formula to calculate F without the use of P matrices from the individual orientation components of two intrinsic calibration matrices, two rotation matrices and a base is given in Kraus (2007, p. 202).

If the image orientation and calibration is not available or has to be refined the fundamental matrix can be calculated from point correspondences either using the seven point algorithm or the normalized eight point algorithm (Hartley & Zisserman, 2000, pp. 281-282) according to the amount of available correspondences. These algorithms are generally applied with a robust estimation method such as RANSAC (**RAN**dom **SA**mple **C**onsensus, Fischler & Bolles, 1981). Seven (or eight) points are chosen from the set of correspondences to calculate the fundamental matrix using the respective algorithm. The result is used to evaluate how many entries in the set are within a certain outlier distance to the calculated model. If the amount of inliers is smaller than a certain threshold then the process is repeated if it is larger than the threshold or a certain amount of iterations have been evaluated it is terminated and the final model is estimated again using all inlier correspondences (Hartley & Zisserman, 2000, pp. 117-118). This method can retrieve a precise estimate of the fundamental matrix but is prone to errors stemming from an irregular distribution of correspondences caused by large clusters of matches in texture rich sub-regions of the images.

Figure 26 shows epipolar lines drawn into an IATS stereo-image pair together with a couple of point matches to illustrate this concept of mismatch detection. Epipolar lines go through the projection of the camera center of the partner image, the so called epipole and limit the position of the corresponding point to one dimension (Figure 25). Homologous stereo-points lie somewhere along the corresponding epipolar line, matches violating this condition can be excluded.

As an extension of outlier filtering based on epipolar geometry a method was developed in this thesis to restrict matches in their position on the epipolar line. Therefore the mean position on the epipolar line of the K nearest neighbors is calculated and compared to the current match. The maximum deviation can be given as a fixed threshold factor, based on the intensity entropy of the feature neighborhood, or based on the variance of the epipolar line positions of the nearest neighbors. This restricts the generation of local spikes of 3D coordinates calculated from the corresponding features.

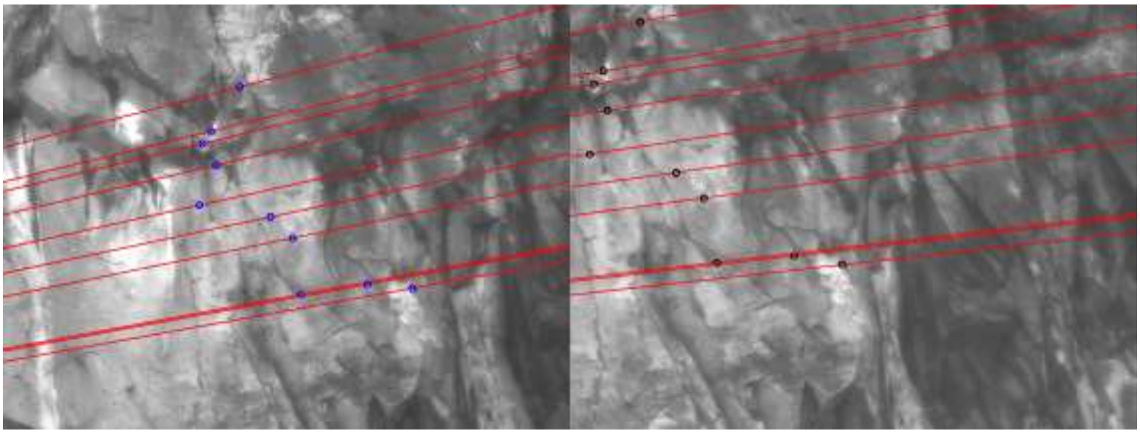

Figure 26: Concept of match evaluation based on epipolar lines illustrated on IATS stereo-images

There is also a global approach to geometry based matching which only relies on points and not on local features using an iterative closest point (ICP) algorithm (Besl & McKay, 1992) to register two sets of points (i.e. find the relative pose transformation) based on the iterative minimization of square errors between the correspondences. However, like for other gradient descent techniques a suitable initial approximation has to be present for the method not to converge into the first local minimum without delivering a useful solution. However, the default relative pose transformation model generally applied in the ICP algorithm is a similarity or an affine transformation. For this reason the algorithm is more effective for the registration of multiple 3D point clouds and less for 2D mismatch detection in a stereo set-up which would require a projective transformation model.

6.1.2 Local geometrical restrictions

Global geometrical restrictions work well in case of stereo-image matching where the poses of the cameras serve as a model which can be estimated and used. However, the application of geo-monitoring also requires techniques to reliably detect mismatches in correspondences over multiple measurement epochs within which occurred local deformations violate global geometrical models. Filtering correspondences using these models would result in the loss of information about the deformation which is a crucial part of the measurement. This problem can be solved with local geometrical restrictions or a hybrid approach wherein the global model sets a loose boundary of outliers whereas local restrictions are used to validate the correspondences.

Local geometrical restrictions in this case refer to so called "meta-descriptors" these are formed using a combination of multiple adjacent detected homologous points. This set of points forms a pattern which can be evaluated as a local model to detect outliers. Such an approach was developed by Liu, et al. (2012) for the purpose of aerial image registration. The method is called Restricted Spatial Order Constraints (RSOC) and was developed to automatically register aerial images with low or repeated texture and large affine transformations (Liu, et al., 2012). It can be used to evaluate a set of correspondences to find mismatches by creating a meta-descriptor from the angle sequence of the K nearest neighbors around each point as shown in Figure 27.

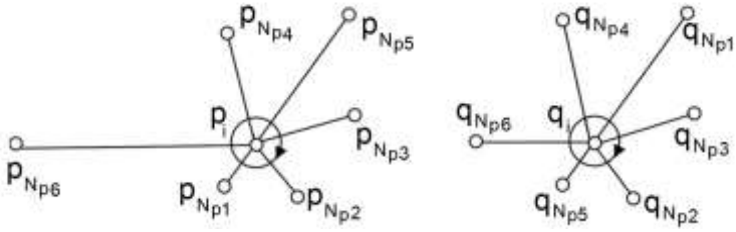

Figure 27: Local feature meta-descriptor consisting of the angular spatial order of p_i's K nearest neighbors and the angular order of the respective correspondences (Liu, et al., 2012)

The descriptor is built from the angular spatial order of each point's nearest neighbors counted clockwise. The respective descriptor formed by the matched point in the second image surrounded by the correspondences of the nearest neighbors of the original point in the first image can represent local affine distortions without ever changing the angular sequence of the descriptor. Assuming two sets of points P, Q and ordered angle values to the K nearest neighbors given as S, the feature point descriptor of the first image may be denoted by $O(p_i) = \{ s_1, s_2, \dots s_K\} = \{N_{p2}, N_{p1}, N_{p6,} N_{p4}, N_{p5}, N_{p3} \}$, the descriptor of the correspondence as $O'(q_i) = \{ s'_1, s'_2, \dots s'_K\} = \{N_{q2}, N_{q1}, N_{q4}, N_{q5}, N_{q3\,,} N_{q6,}\}$.

The two descriptors are rotated so that both start with their lowest lexical entry and are then compared. If the match as well as the correspondences of all nearest neighbors is correct the rotated descriptors should be identical. The count of differences between the descriptors represents the likelihood that the match is incorrect. The rotated descriptors of the above given example would be { 1 6 4 5 3 2 } and { 1 4 5 3 6 2 } and show such a potential mismatch.

After this evaluation a validation cross check is performed by repeating the procedure with descriptors formed by the match in the second image with its local K nearest neighbors and its counterpart in the first image formed by the correspondences of the descriptor points in the second image.

In deviation from the original formulation as given in Liu, et al., 2012 an adapted version of the algorithm was developed within this thesis. The original implementation was specifically targeted at aerial images whereas the adapted algorithm is targeted at the close range photogrammetric application of geo-monitoring. To the author's knowledge this represents the first application of this approach in close range photogrammetry. An error measure of point index penalties was introduced counting differences of all descriptor comparisons. That way a threshold value of maximum penalty count can be used to validate points. This threshold is best set to a value of two due to the twofold cross check matching process. This way points with only one penalty count will not be automatically rejected restricting the amount of discarded correct matches.

To further remove outliers the median distance between the center and the nearest neighbor points on the one hand and among the respective correspondences is calculated. If it deviates by more than a factor of two times a defined scale factor the current point is identified as a gross outlier and discarded.

a)

b)

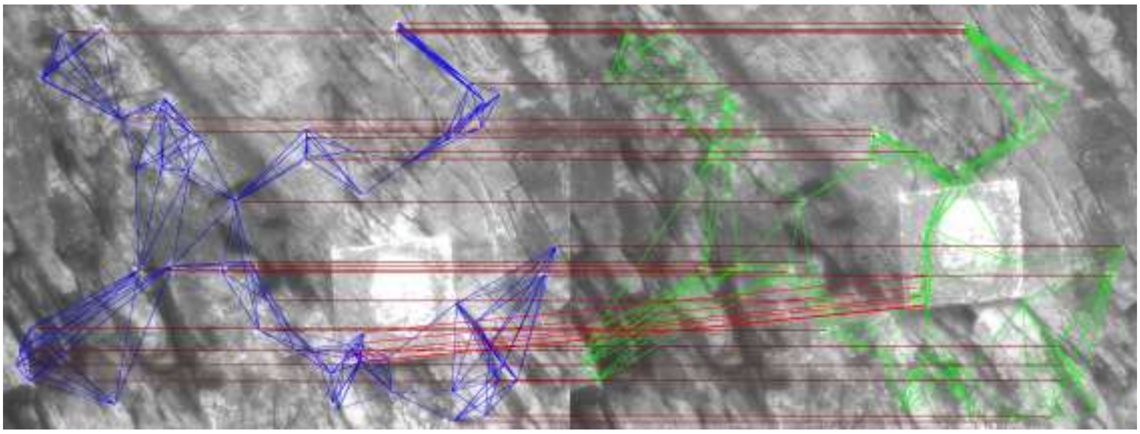

Figure 28: a) Matches on data-set with local deformation and ~25% outliers and b) result of modified RSOC algorithm. Left descriptors formed by six nearest neighbors (blue), cross-check right descriptors (green) - respective descriptors formed by correspondences omitted for the sake of clarity, filtered matches of stable area (dark red) and deformation (light red).

Where global geometry based outlier filtering would exclude correspondences in the deformed area, RSOC maintained those matches by evaluating the local correspondence neighborhood. The descriptor formed by each point connected to its nearest neighbors is shown in blue and green respectively. The descriptors formed by the associated correspondences of these nearest neighbors are omitted in the figure for the sake of clarity. All outliers could be removed whereas matches describing the deformation where retained.

Another local method that was developed and implemented to get rid of coarse mismatches is a filter based on the mean disparity direction of the K nearest neighbors of a match. The allowed deviation of the mean disparity angles of neighboring matches can be given as a fixed threshold or based on the variance of the neighboring angle values.

6.2 Least squares sub-pixel accuracy refinement

The different point detection methods all take the local intensity pattern environment of the detected points into account. However, this is done in varying degrees of accuracy. As described in Section 5.2 the used feature detection methods all apply various degrees of sub-pixel interpolation, however, the process does not use mutual intensity correlation information from the stereo-partner to refine the position. This can be done in the hereby described refinement process based on a least squares matching (LSM). The necessary predictions can be retrieved from a previous feature matching process.

The LSM method was introduced by Grün (1985) and is based on the principle of iteratively transforming a search window of a defined size around a predicted match position to minimize the difference to a reference window around the predicted correspondence position. The optimal transformation is found by minimizing sums of squared intensity differences between search and reference window. The transformations that are applied to the search window used to approximate perspective distortions between image pose are shift, scale, rotation and skew. According to accuracy requirements and processing time constraints two parameters of a shift, four parameters of a similarity- or six parameters of an affine-transformation can be adjusted. Bethmann & Luhmann (2011) extended the model with more complex polynomial functions of up to 12 parameters to accurately match smoothly rounded surfaces.

An affine transformation can approximate perspective distortions on a local scale assuming that the transformation is consistent within the search and reference windows. In addition to the geometric transformations two radiometric parameters can also be approximated and adjusted. One is a constant shift in intensity values and the other is a multiplication factor of intensity values stemming from brightness and contrast changes in the images.

The resulting sub-pixel refined match results heavily depend on the image texture within the matching windows but can reach accuracies better than 0.1 pixel. To the knowledge of the author LSM represents the most accurate feature point matching method available. Local invariant features which can be very well localized form excellent predictions for an LSM refinement which makes the method such a valuable component of the developed processing chain. The downside, however, is long computation time due to the complex iterative process.

Different models of LSM where formulated by Grün (1985), Kraus (2007, pp. 326-328) or Bethmann & Luhmann (2011). The standard method is well known and widely used in photogrammetry and was already applied in the related application of deformation vector generation for glaciers from temporal series of satellite imagery (Debella-Gilo & Kääb, 2012a). A dense matching approach was applied in contrast to the application at hand where LSM is used to refine matching accuracy of locally well-defined image features which generally leads to higher accuracies.

In the following the original method which was implemented and evaluated by the author will be described in more detail. Extensions of the method towards color could be neglected as only one channel gray value images are used in this work.

The intensity values of the image patches within the reference and the search window can be represented as the discrete two-dimensional functions $f(x,y)$ and $g(x',y')$. Due to the presence of noise the ideal correlation condition of $f(x,y) = g(x',y')$ can generally not be established without adding image noise effects $e(x,y)$. This results in the model: $f(x,y) - e(x,y) = g(x',y')$.

Depending on the parameters left to adjust the refinement process can deal with a different amount of geometric transformations in the image. A first approach only adjusts translations by the two shift parameters $\Delta x', \Delta y'$ which are added to the initial position x', y'. This can be successful in homogenous image areas on a rather small scale as in this case the extracted image patches may be similar.

In order to achieve better matching results the above mentioned approach can be extended to an affine solution. The transformation is extended to 6 parameters to allow the approximation of the best fitting affine transformation between the image patches. Since the local extent of the compared patches is rather small an affine transformation accurately enough approximates the real projective transformation between the images, in some cases the affine transformation can even be replaced by simple translations without a significant loss of accuracy. The radiometric transformations between the images can be modeled via with two linear parameters representing a gray value shift and offset.

The system contains eight unknowns, divided in six geometrical and two radial parameters and needs at least as many pixels to compare within the matching windows. However, as generally larger window sizes of seven, nine or more pixels side-length are chosen the resulting systems are highly over determined.

The unknowns are combined into the vector $x^T = \{da_{10}, \ldots, da_{22}, r_0, r_s\}$, with the design matrix A containing their coefficients. The image intensity differences *f(x,y) - g(x′,y′)* are kept in the vector l. This leading to the observation equation (Grün, 1985):

$$l - e(x, y) = Ax. \tag{6.2}$$

The equation can be evaluated using a least squares estimation.

$$x = (A^T A)^{-1} A^T l \tag{6.3}$$

The vector $v = Ax - l$ can be used to keep track of the differences in intensity values of the current solution. The found parameters are used as new predictions and fed into a new iteration of calculations. Intensity values at *g(x′,y′)* will generally not be integer pixel values after or even in the first iteration. This means that the used values have to be interpolated from their local neighborhood. Evaluations have shown that bilinear interpolation is the best method for LSM considering the tradeoff between accuracy and computational effort (Matsuoka, et al., 2008). It is therefore also used in the implementation developed by the author.

The iterative process is ended unsuccessfully after a maximum amount of iterations is reached or successfully if the criterion of convergence is met. In the developed implementation the criterion given by Kraus (1996, p. 80) is used with the proposed value of $\varepsilon = 0.0001$.

$$l^T l - x^T x < \varepsilon \, l^T l \tag{6.4}$$

Template windows are always rectangular and have an uneven number of pixels as side-length so as to have a defined center pixel. The system is calculated in a local template window coordinate system due to numerical reasons. Resulting refinements of the predicted match positions are added to the initial values after each iteration. LSM generally uses integer value prediction positions in the reference image. Due to the automated feature matching process applied in this work to retrieve initial predictions the homologous points typically have sub-pixel positions. Therefore an adaption of the standard method was developed by the author which uses the retrieved geometric transformation parameters after each iteration to calculate the new feature position in the search window not from the reference windows center point but from a position shifted by the initial offset to the closest integer value. Start values for the affine transformation are generally set to an identity matrix but can also be initialized with a global affine transformation estimated from a larger set of matched points. The same holds for the initialization of the radiometric parameters.

As an alternative method to the convergence criterion and the intensity value residual vector, to evaluate the accuracy of a refined position a back-matching approach was implemented. In a first operation the search image position of a correspondence to a certain feature in the reference image is refined with LSM, in a second run the found search image position is used to match back into the reference image switching the roles of both template windows. That way a refined position in the original reference image can be retrieved. The distance from the initial feature position in the reference image to the final position after matching into the search window and back into the reference, measured in pixels is called "back-matching distance" and serves as a consistency measure. The clear drawback of this method is the twofold application of a computationally expensive matching process. However, the back-matching typically converges after the first couple of iterations due to the almost perfect predictions of positions as well as radiometric and geometric transformation parameters from the first matching run.

6.3 Temporal tracking of homologous points

An important part of this work is the temporal tracking of calculated homologous points over time in multiple measurement epochs. There are different strategies of how to implement this temporal tracking. These strategies are further described including their advantages and disadvantages. The contribution of this processing step to the overall processing runtime also has to be considered.

There are multiple examples for the calculation of deformation vectors derived from photogrammetric methods in literature. Ladstädter (2004) developed a monitoring system for aerial images of rock glaciers to retrieve deformation information. It is based on a temporal series of "pseudo-orthophotos" derived from a rectification of the original aerial images with a coarse terrain model evaluated using an "Multi-Photo-Constrained-Matching-Algorithm" based on point prediction and LSM. The 3D points are triangulated using the orientation of the original perspective images. Other approaches only cover the mono image case where disparities are calculated over a series of images consisting of multiple measurement epochs. This way only pixel value deformations can be calculated. In an extension of this approach the texture information from registered Ortho images with corresponding digital elevation models (DEM) is used. Measurements of the deformation are then available using the known pixel size in units of the global coordinate system. The 3D information of both ends of the vectors can be retrieved from the globally oriented DEM. This method was extensively evaluated for the measurement of surface displacements on mass movements especially in the case of glacier creep and rock falls by Debella-Gilo & Kääb (2011) using NCC in combination with image up-sampling and Debella-Gilo & Kääb (2012a) using LSM for dense matching. The NCC based method was improved in Debella-Gilo & Kääb (2012b) by adaptively choosing the template matching windows according to the signal to noise ratio in the underlying texture. This represents a simple scale space analysis which is also an integral part of the detection of local features (Section 5.2) as used in this work. For cases where dense matching results are necessary and processing time or computational effort is unrestricted Redpath, et al. (2013) could show that NCC in combination with an up-sampling of images by a factor of 10 leads to high sub-pixel accurate matching results for the same set of applications. Further investigations were carried out by Heid & Kääb (2012) comparing NCC with different frequency based matching methods to derive glacier surface displacements, resulting in an advantage for matching in the spatial domain.

The above described approach of matching multiple globally registered Ortho images requires only one matching step per measurement epoch. In the application at hand the tracking of homologous non-signalized points, however, requires stereomatching in each epoch as well as a second step tracking the detected features over time. This increases complexity as only temporally connected points with valid stereo-correspondences in their respective epochs form required 3D deformation vectors. One method of tracking points over time is based on feature point matching as described in Section 5.4, re-using feature descriptors already calculated for stereomatching. The other is based on a least squares matching or another correlation based refinement method. The chosen strategy depends on the amount of movement that is to be expected in between measurement epochs. In the following both strategies are described in detail.

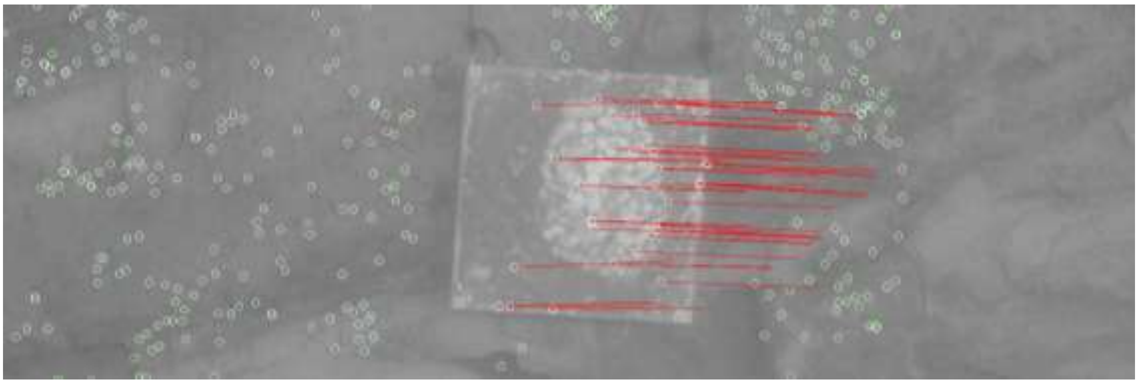

Figure 29: Deformation vectors of moved target tracked using descriptor matching projected into left image of first epoch exaggerated by a factor of 10

The two following strategies to track homologous points over multiple measurement epochs were implemented in the processing chain developed in this work. Strategy one is based on a second application of descriptor based feature matching over time (Figure 29). To avoid mismatches the length of the deformation vectors can be restricted to a certain distance around the projection of the homologous point in a consecutive measurement epoch. This process can be applied to only one side of the stereo-images in each epoch or to both. Further the developed RSOC meta-descriptor (Section 6.1.2) can be applied to detect and exclude potential mismatches. If sufficient computational power or runtime is available the sub-pixel results can be refined using LSM (Section 6.2). The workflow for this first strategy can be described as follows.

- Execute feature based stereomatching in every measurement epoch
- Save feature descriptors of homologous points
- Project 3D coordinates of points from previous epoch into the images of consecutive epoch
- Execute descriptor matching of left images over epochs restricted to a maximum deformation around the projected points
- Execute descriptor matching of right images over epochs (optional)

- Compute RSOC meta-descriptors to detect mismatches (optional)
- Refine matching results with LSM (optional)

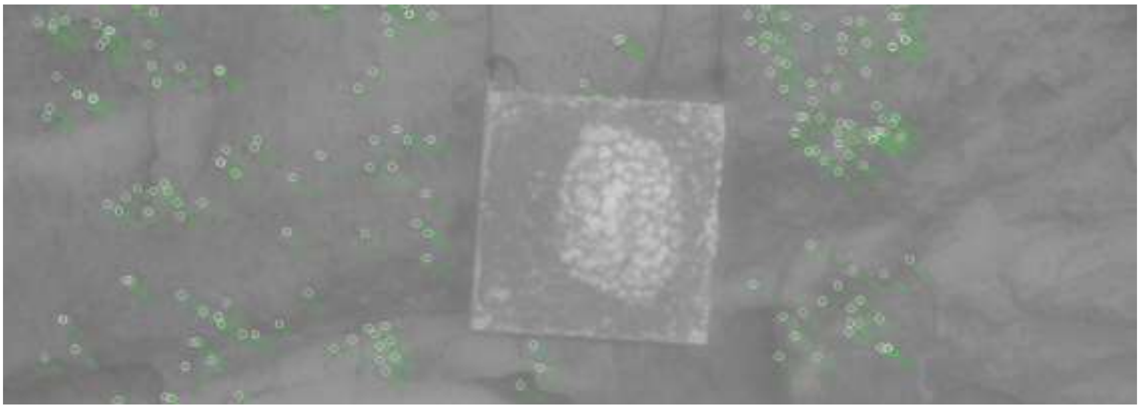

Figure 30: Artificially introduced deformation tracked with LSM projected into left image of first epoch exaggerated by a factor of 10

Whereas this approach can be used to detect larger und unpredictable deformation, the second developed strategy for feature tracking over time is based on LSM and can be used to reliably follow predictable small scale deformation in high accuracy. Figure 30 shows the application of this method on an artificially deformed data-set, larger deformations such as the moved target cannot be detected. The second strategy consists of the following steps.

- Execute feature based stereomatching in every measurement epoch
- Save feature descriptors of homologous points
- Project 3D coordinates of points from previous epoch into the images of consecutive epoch
- Add offset to account for predicted deformation movement (optional)
- Execute LSM to detect sub-pixel deformation

Due to the complex multi-image matching involved in the generation of 3D deformation vectors which require non-signalized homologous points to be detected in four images to connect two measurement epochs, the amount of vectors that can be tracked over multiple epochs is drastically reduced with each epoch. Therefore all found temporal correspondences of homologous points are kept in each epoch making up for tracks lost because of changed illumination conditions are real rock fall events. The occurred deformation can be derived for a whole time series by connecting the movements of neighboring deformation vectors within certain proximity.

6.4 Calculation of deformation vectors

The calculation of the deformation vectors is the final step of evaluation done in the herein presented work. The homologous stereo-points tracked over multiple time series are used to calculate 3D deformation vectors with spatial forward intersection using the camera calibration parameters. Deformations are visualized in 2D as an overlay on the original images as well as exported to 3D in the VRML format. This allows an expert user to interpret on-going deformations.

The calculation of 3D coordinates from rays in object space derived from image positions and intrinsic as well as extrinsic camera calibration can be done with a spatial forward intersection or triangulation. Each point has to be observed from at least two positions. Due to errors in camera calibration as well as in the corresponding 2D feature positions the rays will generally not intersect. This means there may be no 3D point X that exactly satisfies $x = PX$, $x' = P'X$ for the measured image points x, x' with the camera orientations P and P' (Hartley & Zisserman, 2000, p. 310). Deviations caused by lens distortion effects as calculated in the calibration process (Section 2.3.3) are also taken into account at this stage as increments to the measured image coordinates.

There are different possibilities to calculate the approximate 3D point position. The first one is the geometric solution derived from the calculation of the midpoint on the line orthogonal to both rays representing the shortest distance. The calculation is based on the orthogonality condition (equation (6.5)) of the two rays L, L' through the camera centers C, C' in the direction defined by the angle values of the measured image coordinates denoted as u, u' with the measurement errors d, d'. The resulting linear equations can be solved for the error increments. Averaging leads to the position of the point X.

$$\begin{gathered} L = C + du, \quad L' = C' + d'u' \\ (L - L')\,u = 0, \quad (L - L')\,u' = 0 \end{gathered} \tag{6.5}$$

The algebraic solution is based on the collinearity conditions. The projection equations of the measured point positions in the images can be combined into a form of $AX = 0$. The scale factor of homogenous 2D coordinates is eliminated by a cross product. This leads to three remaining equations for each image position. Two of them are linearly independent. Combining the four equations from two images the homogenous system of equations $AX = 0$ can be established with p^{iT}, p'^{iT} denoting the individual lines of the camera orientation matrices P and P' (Hartley & Zisserman, 2000, p. 312).

$$A = \left[x\, p^{3T} - p^{1T}, \quad y\, p^{3T} - p^{2T}, \quad x'\, p'^{3T} - p'^{1T}, \quad y'\, p'^{3T} - p'^{2T} \right]^T \tag{6.6}$$

The homogenous solution can be calculated as the unit singular vector with the smallest singular value of A using singular value decomposition (SVD). Another option is setting $X = (X, Y, Z, 1)^T$ to reduce the set of homogenous equations to four inhomogeneous equations with three unknowns. The disadvantage of this method is the condition that X does not lie at infinity (Hartley & Zisserman, 2000, p. 313).

If more detailed information about the intersection errors is necessary the intersection can be calculated as a constrained adjustment minimising the distance between 3D rays using the Gauß-Helmert model (Niemeier, 2008, p. 176). The unknowns $\vec{\mathrm{x}}_\mathrm{N} = [\mathrm{x_N}, \mathrm{y_N}, \mathrm{z_N}]^\mathrm{T}$ describe the minimum distance between the rays $\vec{\mathrm{r}}_i$ going through corresponding camera centers $\vec{\mathrm{x}}_i$, d_i are the unknown distances to $\vec{\mathrm{x}}_\mathrm{N}$ (Equations (6.7)). The observations are coordinates of $\vec{\mathrm{x}}_i$ and the horizontal and vertical angle measurements Hz_i and V_i.

$$\vec{\mathrm{x}}_i + \vec{\mathrm{r}}_i - \vec{\mathrm{x}}_\mathrm{N} = 0 \qquad \vec{r}_i = \begin{pmatrix} d_i \, sin\, V_i \; cos\, Hz_i \\ d_i \, sin\, V_i \; sin\, Hz_i \\ d_i \, cos\, V_i \end{pmatrix} \tag{6.7}$$

With the Gauß-Helmert model defined in (Niemeier, 2008, p. 176; Equation (6.8)) a covariance matrix (Equation (6.9)) can be derived.

$$Bv + A\hat{x} + w = 0 \tag{6.8}$$

$$Q_{xx} = (A^T (B Q_{ll} B^T)^{-1} A)^{-1} \tag{6.9}$$

Based on this covariance matrix error ellipses can be calculated with the matrix's eigenvalues defining the semi-major and semi-minor axes and its eigenvectors defining the error ellipses' direction.

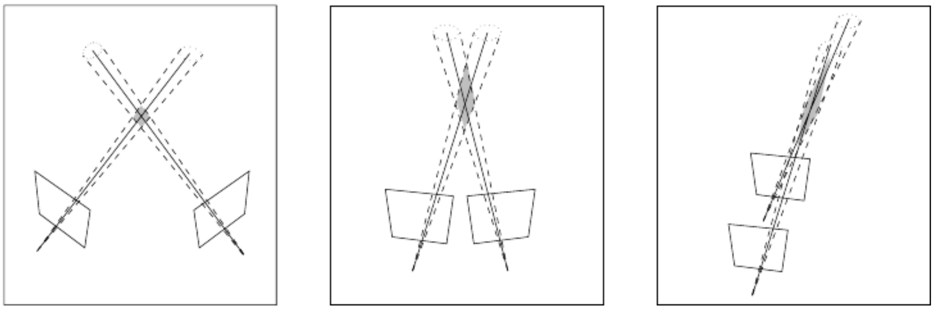

Figure 31: Uncertainty of triangulation, glancing angles lead to reduced accuracy of localization, indicated by shaded regions (Hartley & Zisserman, 2000, p. 321)

As mentioned in Section 3.1 due to imperfections in camera orientation as well as in homologous point measurements the accuracy of the reconstruction depends on the geometry of the intersected angles. Figure 31 illustrates this principle for different configurations.

7 Method evaluation on synthetic test data

The feature detection, description and refinement algorithms described in Section 4 and 6 were evaluated on artificially generated synthetic test data-sets with known ground truth to determine strength and weaknesses of the individual methods.

7.1 Generation of synthetic "ground truth" data

In order to evaluate the different algorithms test data-sets were chosen and generated representing different challenges for the feature detection, description and matching algorithms that are described later in this section. The data-sets consist of IATS images captured at different sites in Austria, Switzerland and Germany. The chosen data-sets show homogenous rock-walls, inhomogeneous clutter and debris as well as mixed regions. Further data-sets of increasing viewpoint change were evaluated as well as images containing illumination changes as this is one major issue to be covered.

Data-sets to evaluate feature matching methods generally contain planar structures related to the stereo-partner with a reference homography which can be used to calculate reference feature positions. Due to the unstructured environment encountered in geo-monitoring this approach could not be used. Instead synthetic reference ground truth data was generated. Highly accurate matching results were calculated using a state of the art dense matching of stereo-partners within the individual data-sets using the Semi-global matching method developed by Hirschmuller (2006). The resulting disparity maps where interpolated to fill holes caused by a failed stereo correlation or missing information due to occlusions. Using the first stereo-partner and the interpolated dense disparity maps describing the image relation to the second stereo-partner an artificial version of the second image was created. The texture of the first image was warped into the geometry of the second using lanczos based resampling with a kernel size of eight pixels. The lanczos filter was developed by Duchon (1979) and represents one of many finitely supported approximations of the sinc-function which is in theory considered as the ideal interpolation method. The lanczos kernel is defined as the normalized sinc function multiplied by sinc(x/a) for the interval $-a \leq x \leq a$.

The individual feature detection and matching algorithms can be evaluated on the real left image and the artificially generated right image. From the disparity maps used to generated the synthetic image and according sub-pixel interpolation the stereo-correspondence of each pixel is known and can be used to evaluate the applied feature detection and matching methods in terms of repeatability, localization accuracy and runtime performance.

Ten synthetic test data-sets where generated to test different properties of local feature detectors and descriptor as described above. The data-sets including a rationale for choosing the individual images are given. Left stereo-partners consist of original IATS images, right stereo-partners were synthetically generated from interpolated dense matching results as described above. The exact position of each pixel correspondence is therefore known and can later be used to evaluate an applied feature matching process. The following table lists the origin of the individual data-sets.

Table 1: Generated synthetic stereo data-sets

ID	Location	Test case	Viewpoint change [gon]
01	Quarry Untermurbach, close to Munich	Uneven intensity distribution, cast shadows	41.1
02	Pasterze glacier, Austria	Homogenous rock wall	5
03		Homogenous rock wall and in-homogenous debris	4.3
04		Homogenous rock wall and in-homogenous debris	4.4
05		In-homogenous debris	5.5
06	Weinzödl, Graz	Rock wall with cracks	18
07	Clay-pit Pellheim, close to Munich	Low texture clay wall	18.2
08	Illgraben, Switzerland	Scale change	15.9
09	Brick wall, Munich	Increasing viewpoint change	22
10			47.5

The first test data-set stems from IATS images acquired in a quarry close to Munich. The total viewpoint change between the two images is around 41.1 gon. The images were chosen to test the algorithms' performance in the presence of unevenly distributed intensities caused by cast shadows and the potential positive influence of pre-processing methods.

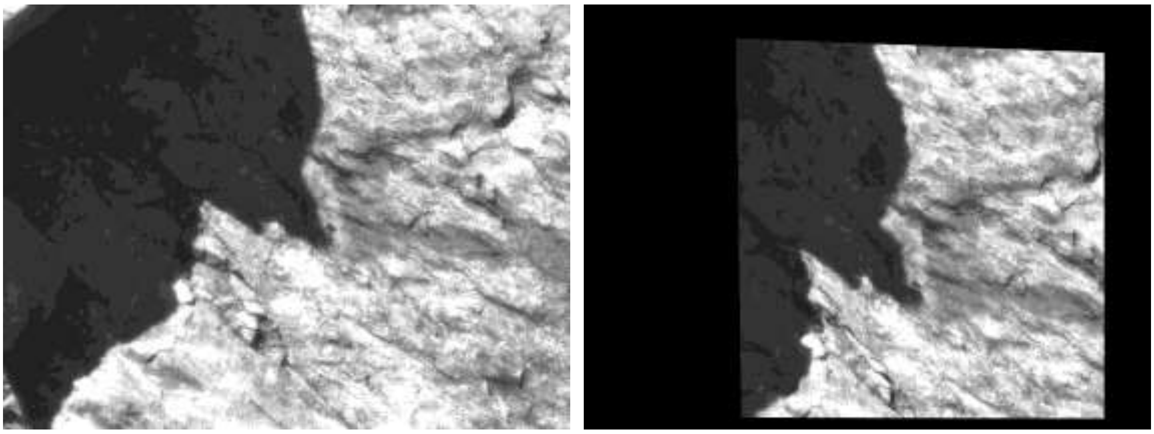

Figure 32: Synthetic stereo data-set 1, cast shadows – viewpoint change ~41.1 gon

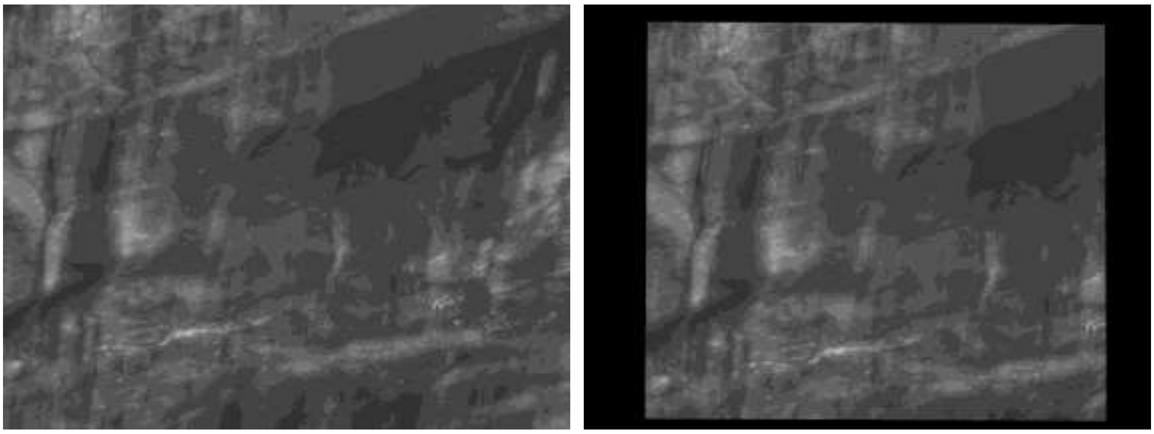

Figure 33: Synthetic stereo data-set 2, homogenous rock wall – viewpoint change ~5 gon

The second IATS data-set (Figure 33) shows a homogeneous rock wall from a distance of about one kilometre and a viewpoint change of only around 5 gon. The images where acquired at a measurement campaign at the Austrian glacier Pasterze. Rationale for this data-set is the test of algorithm performance on homogenous rock areas.

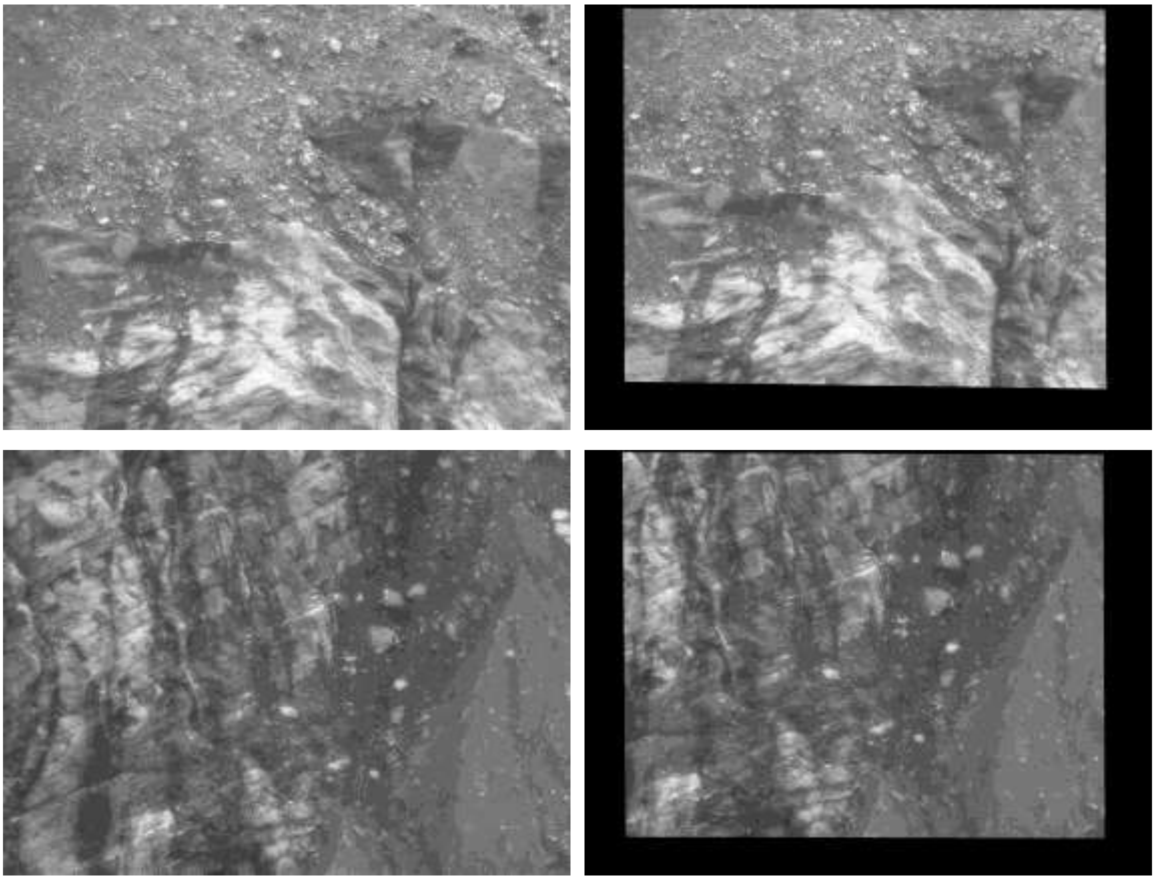

Figure 34: Synthetic stereo data-set 3 and 4, homogenous rock wall mixed with in-homogenous debris – viewpoint change ~4.3 / 4.4 gon respectively

The third and fourth chosen data-sets (Figure 34) were also acquired at the Pasterze and illustrate a common form of terrain encountered in applications of geo-monitoring, areas consisting of homogenous rock walls mixed with areas of in-homogenous debris.

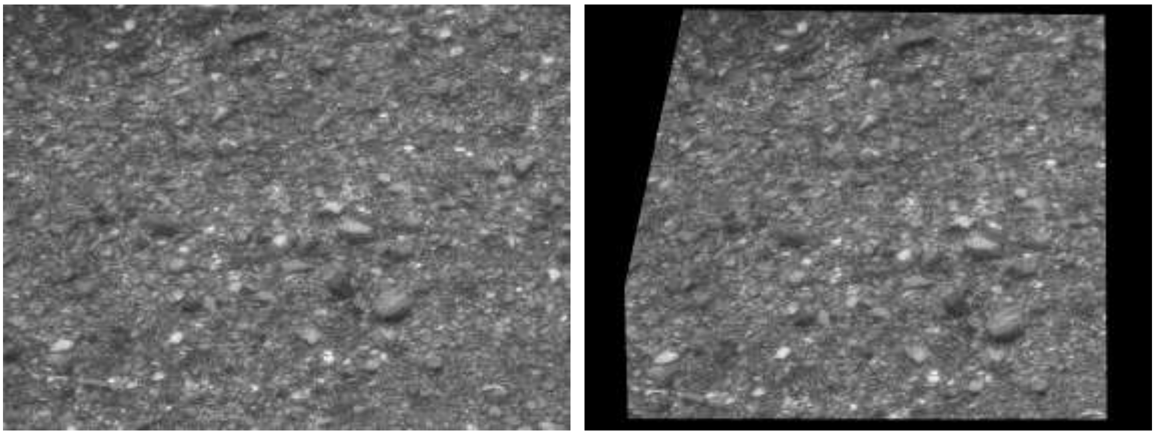

Figure 35: Synthetic stereo data-set 5, in-homogenous debris – viewpoint change ~5.5 gon

Test set five (Figure 35) consists of in-homogenous debris and rounds up the narrow viewpoint test case designed to evaluate complementary properties of the different methods. The response of the different local feature detectors and descriptors to changing environment was evaluated with these data-sets.

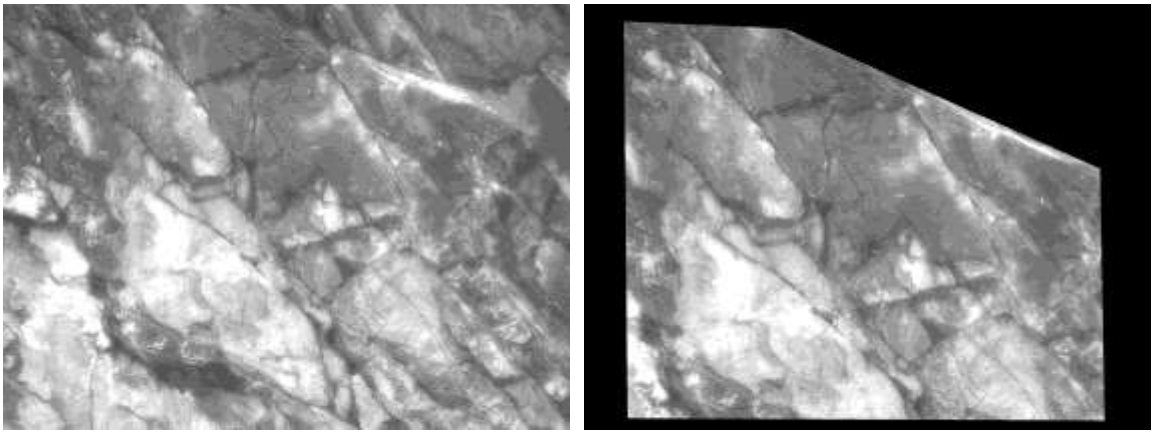

Figure 36: Synthetic stereo data-set 6, rock wall with cracks – viewpoint change ~18 gon

The data-sets six and seven consist of very different objects monitored from a distance of around 260 meters each in a typical IATS measurement configuration with a stereo viewpoint change of around 18 gon. Figure 36 shows a rock wall with multiple cracks in a climbing crag in Graz Weinzödl which has been known for repeated rock fall events. Figure 37 shows a lowly textured but nonetheless instable clay wall at a clay-pit in Pellheim close to Munich.

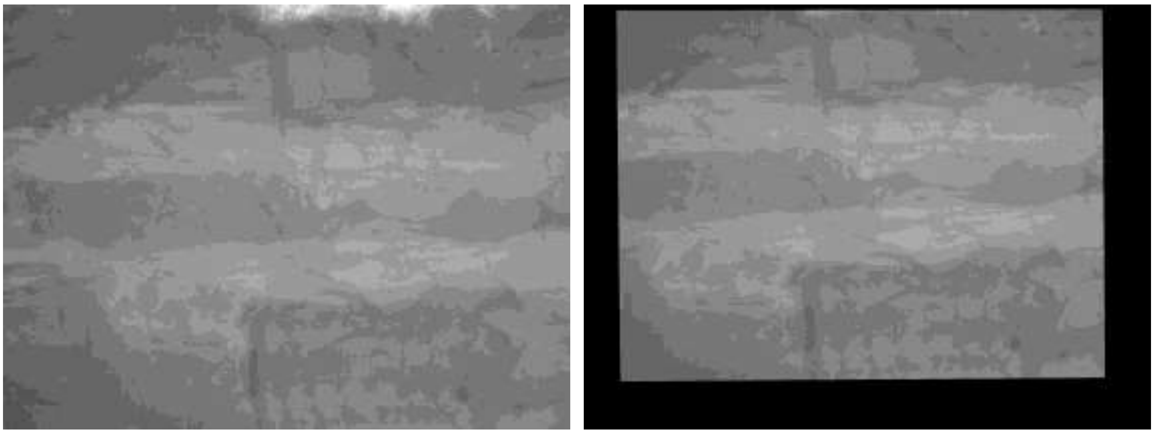

Figure 37: Synthetic stereo data-set 7, low textured clay wall – viewpoint change ~18.2 gon

Data-set eight illustrates a challenging measurement configuration wherein the stereo base line is not parallel to the measured wall but in the measurement direction. As the images are taken with upward looking cameras this results in a scale difference as well as a larger geometrical distortion as can be seen in Figure 38. The images where acquired at a measurement campaign in Illgraben, Switzerland.

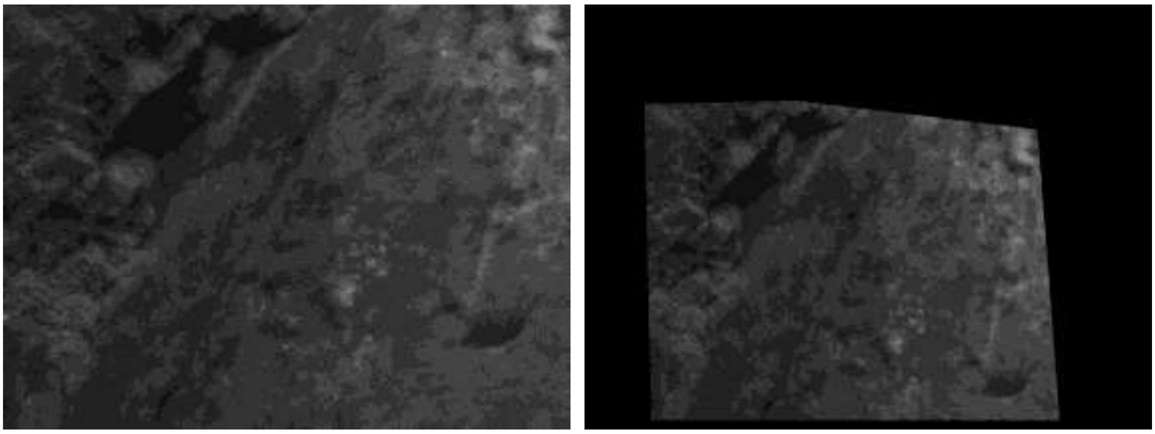

Figure 38: Synthetic stereo data-set 8, rock wall with scale difference – viewpoint change ~15.9 gon

In order to evaluate the effect of viewpoint changes in a controlled environment the data-sets nine and ten were created which consist of IATS images of a brick wall in Munich taken from three different positions. Therefore a data-set with a viewpoint change of ~22 gon and one with a viewpoint change of ~47.5 gon showing the same area and texture is available.

Figure 39: Synthetic stereo data-set 9 and 10, brick wall with increasing viewpoint change of ~22 / 47.5 gon respectively

Figure 39 shows the brick wall images forming two stereo configurations with increasing viewpoint change. This data-set was used to evaluate the influence of the resulting geometrical distortion on the feature matching process.

7.2 Local feature detectors performance evaluation

Plenty of studies evaluating and comparing the performance of local feature detectors have been published (Mikolajczyk & Schmid, 2001, 2002, 2004, 2008, 2012, Triggs, 2004, Jazayeri & Fraser, 2008, Gauglitz, et al., 2011). In addition almost every publication about a newly introduced detector has a section comparing it to other state of the art methods (e.g. Rosten & Drummond, 2006). Publicly available test data-sets were created against which new developments can be tested (Scharstein & Szeliski, 2002). The evaluation data-sets generally consist of different sequences to illustrate detector invariance towards certain distortions in the image such as increasing in plane rotation, affine transformation, scale change, Gaussian blur or jpg compression artefacts. Most of these sequences consist of man-made structures and cover only one viewpoint.

In the current work available evaluations were used as a base decision criterion which detectors to implement and analyze further. In deviation from existing analysis the task at hand places a whole new set of requirements which cannot be completely fulfilled by any single algorithm. This required an extensive comparative evaluation of feature detection methods on data-sets specifically selected to illustrate the conditions to be considered in the process of image-based geo-monitoring from IATS image data. In the following the conducted evaluation of local feature detector performance is described.

A performance evaluation of implementations of previously mentioned algorithms was carried out. They were analyzed regarding their localization accuracy and repeatability under applied transformations in the image domain as well as regarding their individual runtime behavior.

7.2.1 Implementation details

The algorithms described in Section 5.2 with the highest performance according to literature were implemented and tested. For this reason a unique command line based test tool was developed that includes image pre-processing (Section 4.2) such as histogram equalization and (contrast limited) adaptive histogram equalization (Section 4.3), local feature detection and description using various different algorithms, feature vector matching using different techniques and post processing to detect mismatches and refine feature positions (refer to Appendix H for a full list of the available functionality). The tool was developed as a 32bit C++ application under Microsoft Windows 7 using the development environment Visual Studio 2008. As a re-implementation of all used methods from scratch would not have been feasible nor have led to reproducible results the widely used computer vision library OpenCV (Bradski, 2000) was integrated on source code basis allowing modifications and detailed evaluations of all integrated functionalities. Further a well-documented, high performance improved implementation of the SURF algorithm including its fast DoH detector available as source code in the OpenSURF library (Evans, 2009) was adapted and integrated.

A set of best performing corner-based as well as blob-based detectors according to literature were evaluated. The analyzed cornerbased detectors consist of the recent high speed detector FAST which makes use of machine learning techniques (Section 5.2.2.3), two of its multi-scale adapted derivatives ORB and BRISK, the well-known photogrammetric method HARRIS (Section 5.2.2.1) and a simple multi scale adapted version of the latter which will be called Pyramid HARRIS. The original implementations of FAST, ORB and BRISK were developed and implemented by the respective authors already in the used OpenCV framework. The OpenCV HARRIS implementation is a re-implementation of the original method with a strong focus on processing speed which makes it perfectly suitable for the application at hand. As the standard HARRIS operator only finds features on a single scale an enhancement of the method was implemented by the author of this work that creates a scale space by generating Gaussian image pyramids and applies the HARRIS corner detector on each level. This detector will be called Pyr(amid) HARRIS. However, this simple approach does not contain non-maxima suppression of features present on adjacent scales. The blob based methods DoG, DoH, CenSurE and MSER were also evaluated. DoG (Section 5.2.3.2) is part of the SIFT algorithm (Section 5.3.2.1 and Lowe, 2004a) of which a closed source reference application exists in the form of a command line tool executable. As it is impossible to separate the detection from the description step within this tool it was tested with a high threshold leading to zero detected features and therefore no feature description. The runtime was therefore solely used for image IO operations (which are negligible) and feature detection. That way it could be determined that Lowe's original implementation runs four to five times slower than the one contained in OpenCV. Due to lower runtime and the additional benefit of a direct interfacing with the developed point matching tool on source code basis the DoG implementation from OpenCV was used. The CenSurE (Section 5.2.3.3) detector is also present in OpenCV with some minor improvements under the name STAR, as is the region-based detector MSER (Section 5.2.3.4). The fast DoH detector (Section 5.2.3.1) which is part of the SURF algorithm (Section 5.3.2.2 and Bay, et al., 2008) was integrated in the evaluation tool based on the afore mentioned OpenSURF library.

7.2.2 Accuracy and runtime evaluation

The ten synthetic test data-sets were used to evaluate the performance of the different algorithms. The three evaluated criteria are repeatability, localization accuracy and computational complexity measured in algorithm runtime. The repeatability score is calculated as described in Mikolajczyk & Schmid (2004) as the ratio between point correspondences and the lower number of points detected in one of the images. Only features present in both images are taken into account whereas the feature sets are restricted to the minimum amount of features found in both images. The localization accuracy is the error in relative location of corresponding features according to the reference measured in pixels. Two features are considered as corresponding if the localization accuracy is higher than 1.5 pixels.

7.2.2.1 Local feature detection evaluation overview

The following tables give an overview over the full evaluation of detectors. The Data-set ID links to the individual generated ground truth data-sets as introduced in Table 1. For each applied detector the amount of found features and the detector runtime is shown. All data-sets where evaluated in their original form as well as after contrast enhancement with adaptive histogram equalization. The amount of features represents points found in the first image restricted by the condition that they are also theoretically visible in the stereo-partner image, further reduced to the minimum of remaining features in both images. This means that a total of 1000 features found in the first image might be reduced to 800 visible in both images. If however only 500 features were found in the second image the 800 points, are ordered according to their filter response and restricted to the 500 strongest features. On the other hand if 1000 features were to be found in the right image they are accordingly also restricted to 800.

Detector runtime was measured directly in the individual detection functions not including data IO operations or image pre-processing. For each data-set each analyzed algorithm was executed 100 times leading to averaged system runtimes which are shown in Tables 18 - 21. Due to restricted space and for the sake of clarity calculated empirical standard deviations of the runtime were omitted in the tables. Average values of these runtime deviations over all data-sets measured in seconds are shown in the aggregation of detector runtime performance in Table 2.

Every tested algorithm is configurable using a set of different parameters. Due to the exponential amount of possible configurations it is not possible to evaluate an extensive comparison of all detectors with a full set of different parameters. Therefore the main common parameters such as size of the scale space which might be given as amount of image pyramids to be evaluated were kept constant over all algorithms. Further every method has one main filter response threshold which was carefully adjusted for every detector on every individual data-set to lead to similar amounts of features. Due to the different implementations, detection approaches and responses to image texture this can only ever lead to a similar range with a larger degree of variation. The alternative strategy of calculating a maximum amount of present features with each method and then restricting those to the same value would have seriously distorted the results of the runtime evaluation and was therefore discarded.

As runtime is one of the key parameters that was evaluated the above mentioned timing of all detectors averaged over all data-sets is shown in Table 2. The algorithms are ranked in ascending order and a comparison is done by showing the runtime in percentage of the slowest method which turned out to be the DoG implementation of the SIFT algorithm. The FAST method remains true to its name proving to be the fastest available feature detector being around 480 times faster than DoG followed by the corner based methods ORB and BRISK which take around three to five percent of the slowest method's runtime. The similar blob based methods DoH and STAR come third with runtimes of around 14 to 18 percent of DoG followed by HARRIS and MSER which are still showing a speedup factor of two. To set the absolute runtime values into perspective; the evaluation was carried out on a Laptop with an Intel i7-3520M CPU and 8 GB of RAM.

Table 2: Aggregation of detector runtime evaluation

Detector	Runtime [sec]		Percentage of slowest
FAST	0.0052	±0.0003	0.26 %
ORB	0.0690	±0.0032	3.46 %
BRISK	0.1002	±0.0040	5.03 %
STAR	0.2749	±0.0047	13.80 %
DoH	0.3504	±0.0168	17.59 %
HARRIS	0.7007	±0.0092	35.17 %
PyrHARRIS	0.9442	±0.0099	47.39 %
MSER	0.9556	±0.0154	47.97 %
DoG	**1.9923**	**±0.0383**	**100.00 %**

A comprehensive evaluation of all local feature detection methods on all synthetic stereo test data-sets as introduced in Section 7.1 is listed in the Tables 18 - 21 which were placed in Appendix A. Evaluation results are presented separately for corner based and blob based methods showing processing results with and without applied adaptive histogram equalization.

Table 18 shows the results of the analyzed corner based methods on all data-sets, whereas values derived with the blob based methods DoH, DoG and STAR as well as with the blob and region based method MSER are given in Table 19.

Table 20 again illustrates the results of corner based methods with the main difference of applied pre-processing of the images with adaptive histogram equalization before the feature detection process.

Table 21 illustrates the respective values for the blob and region based methods. As initially mentioned all detectors were able to find larger amounts of features with a decreased variation between the individual methods.

As is visible in the tables the average amount of detected features present in both images is around 1776 without image pre-processing with an empirical standard deviation of around 880 points. For the data-sets evaluated with adaptive histogram equalization for image pre-processing leading to more evenly distributed intensity and contrast conditions an average of around 2186 ± ~674 features where detected with all methods.

7.2.2.2 *Visual evaluation of detected feature distribution on different data-sets*

The first synthetic data-set as shown in Figure 32 includes cast shadow regions posing a great challenge for all tested local feature detectors. As all of these algorithms rely on a certain global filter response threshold which if surpassed on a specific image location marks it as local feature to be used, uneven intensity distribution in the images leads to the identification of features only in certain parts of the images. Data-set one was created to illustrate this problem and to show the positive effects appropriate pre-processing can have on local feature detectors.

Figure 40 shows the resulting features of all tested algorithms on the left image of the cast shadows stereo data-set. The features were restricted to those also visible in the stereo-partner (Figure 32); therefore no features are present around the borders of the image. Local features are marked by a color coded cross corresponding to the color of the name of the method as shown in the legend in the upper left of the image. It is clearly visible that all detectors unanimously ignored the shadowed region in favor of the better illuminated area with significantly higher contrast. Another thing to be noted is the apparent lack of red FAST features. This is due to ORB and BRISK keypoints which basically rely on the same underlying detection technique as FAST and may therefore overwrite these features in the output images.

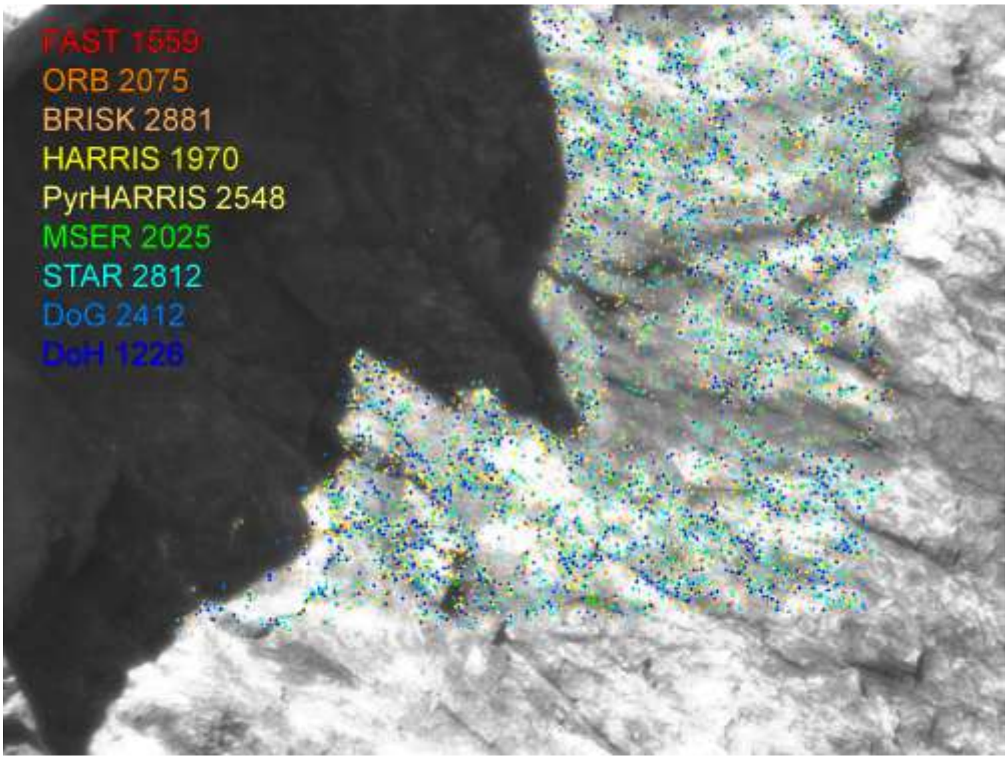

Figure 40: Application of local feature detectors on cast shadows data-set, amount of features restricted to those visible in the synthetic stereo-partner shown in color coded legend

Figure 41 present the same data-set again with an additional pre-processing step of adaptive histogram equalization (as described in Section 4.3) applied before the feature detection process. The resulting homogenous intensity distribution after the adaptive histogram equalization causes higher filter responses on the shadow region. All detectors improved feature detection rates in the enhanced region. Noticeable is the difference between blob-based (blue) and corner-based (red to yellow) methods. Improved shadow regions exhibit a higher proportion of found features from corner based methods than in the originally well illuminated image section. This leads to the conclusion that in case of low contrast areas improved by pre-processing methods corner based detectors exhibit complementary properties to blob based methods which show stronger responses in areas of higher contrast. This may be caused by their high susceptibility to image noise which is amplified through contrast enhancements.

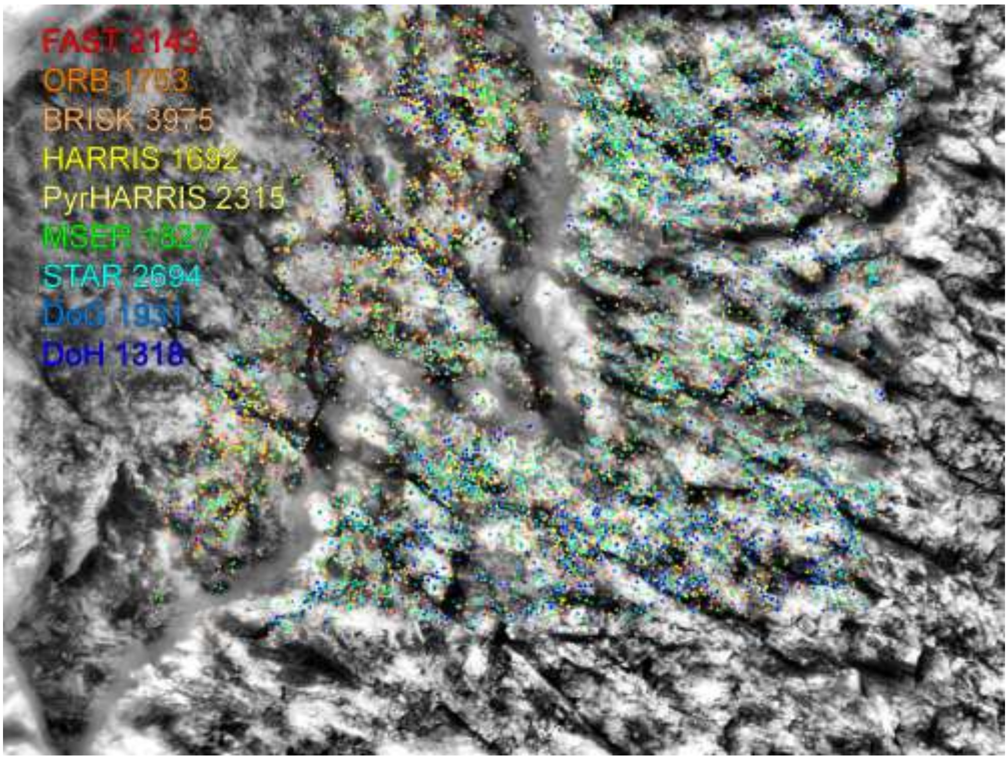

Figure 41: Application of local feature detectors on cast shadows data-set after adaptive histogram equalization, amount of features restricted to those visible in stereo-partner shown in legend

To improve the general outline and readability of this work the illustrations referred to in the following paragraphs were placed in the Appendix, B Visual evaluation of detected feature distribution.

Data-sets number two to five were chosen because of their distinctively different image texture while showing similar narrow stereo viewpoint differences of only 4.3 to 5.5 gon. Figure 83 illustrates the application of the local feature detectors on the homogenous rock wall of data-set two before and after adaptive histogram equalization. The amount of features in previously low textured regions is increased in the enhanced data-set. However, many of the previously strongest and therefore most repeatable features were replaced by new ones.

Figure 84 presents the evaluation of data-set three containing the top of a rock wall covered with loose debris. A majority of features is found in the debris region. This effect is amplified by the contrast enhancement.

Figure 85 contains the evaluation of data-set number four, depicting a homogenous rock wall mixed with a flow of debris. A shift of detected features from the rock wall to the debris is visible after contrast enhancement.

The feature detection process on the in-homogenous debris of data-set five is shown in Figure 86. Due to the already rather uniform intensity distribution in the original image the contrast enhancement does not greatly change the distribution or the amount of found features. The size of the illustrated region was reduced; the legend however still refers to the number of features found in the full uncropped image. This is also the case for all further images labeled as cropped.

The evaluation of data-set number six showing a rock wall covered with cracks of various sizes is given in Figure 87. As was to be expected a majority of the features are found along cracks. The contrast enhancement amplifies fine grained structures which are found as strong features mainly by corner based methods but may be of questionable repeatability. Data-set number seven as evaluated in Figure 88 exhibits very low contrast. This leads to the only case where the three detectors ORB, HARRIS and MSER could not retrieve enough features even though their response thresholds were significantly reduced. Detection rates of most other detectors were also below average. The applied contrast enhancement increased detection rates and lead to a satisfactory coverage.

7.2.2.3 *Evaluation of localization accuracy and repeatability of local feature detectors*

The following charts show the respective detector localization accuracy and repeatability within the narrow viewpoint data-sets two to five. As above all evaluations where carried out on the original images and on versions enhanced with adaptive histogram equalization.

The evaluation of detector localization accuracy as illustrated in Figure 42 shows that all detectors are able to reach differing levels of sub-pixel accuracy. The only detector with significantly better performance in these data-sets is DoG which is also the one with the longest runtime (Table 2). Empirical standard deviations of all methods are illustrated in the chart using error bars and lie in a range between ~0.21 pixel for DoG and ~0.4 pixel for BRISK. An interesting aspect of this evaluation is the fact that the localization accuracy curves of all detectors show a rather similar behavior for the individual data-sets. This may indicate that the complementarity of the individual methods is lower than expected. The ranking of detector performance in terms of localization accuracy is also quite stable over the different data-sets with the obviously superior performance of DoG and the generally worst results from BRISK. However, due to the fact that only those detectors with the highest performance stated in literature were evaluated the results all lie in a similar range with no method failing completely. Further, intensity based pre-processing does not significantly reduce the overall detector performances in terms of localization.

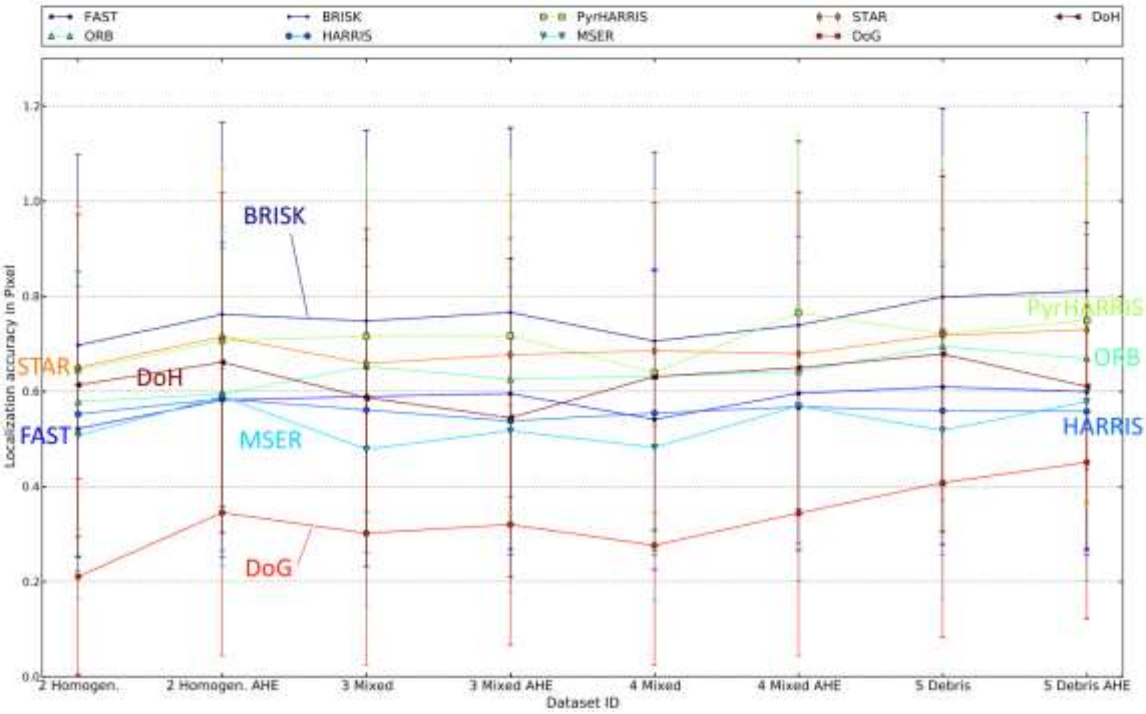

Figure 42: Narrow viewpoint data-sets detector localization accuracy

The detector repeatability scores as defined in beginning of Section 7.2.2 are shown in Figure 43. It clearly shows that the image pre-processing reduces detector repeatability for all methods. This can be explained by the fact that lower contrast features are enhanced and may get higher filter responses than the strongest features in the original images which would have been more stable. In the narrow viewpoint data-sets the BRISK features have the lowest repeatability whereas on the other end HARRIS corners, followed by MSER regions and blobs detected with DoG or STAR show the highest scores.

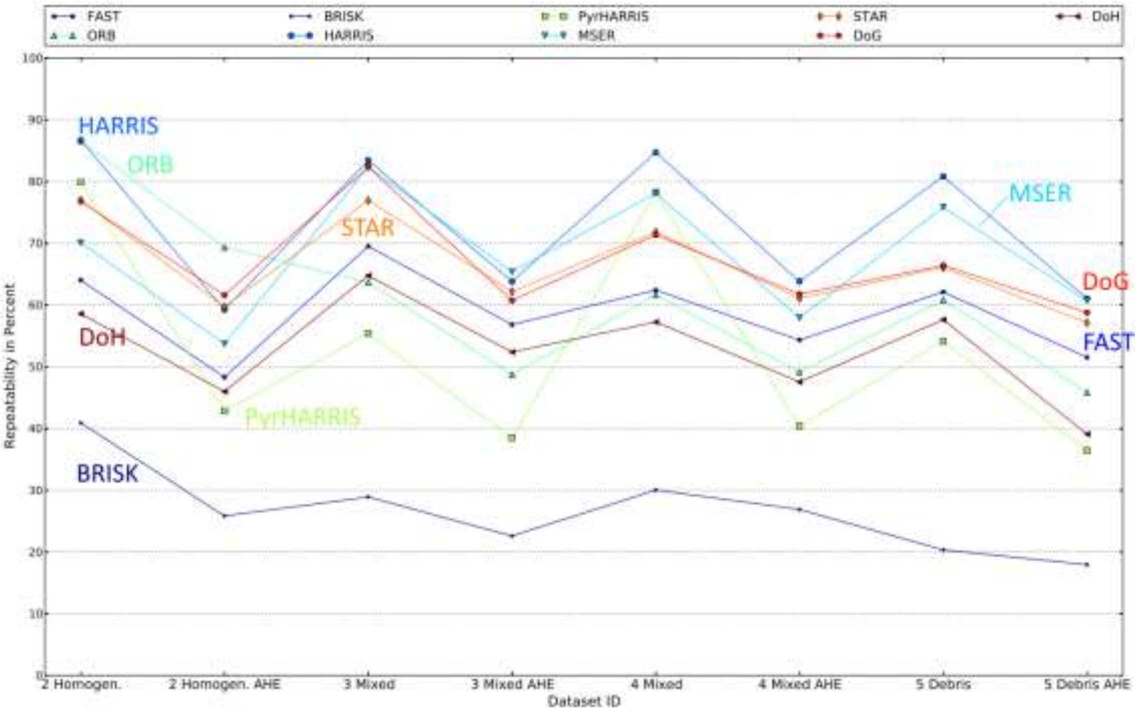

Figure 43: Narrow viewpoint data-sets detector repeatability

To improve continuous readability the remaining evaluation charts of other data-sets were placed in Appendix, C Feature detector performance evaluation charts.

Results of the evaluation of the data-sets illustrating illumination changes in a cast shadow area with a viewpoint change of 41.1 gon (data-set ID 1), a cracked rock wall with a viewpoint change of 18 gon (ID 6), a low textured clay wall with a viewpoint change of 18.2 gon (ID 7) and scale change with a viewpoint change of 15.9 gon (ID 8) are shown in Figure 89.

Throughout the different aspects evaluated within the mentioned data-sets the overall ranking of detectors with DoG among the best and BRISK among the worst performing methods remains unchanged compared to the evaluation of the narrow viewpoint data-sets. However, the advantage of DoG over the following methods is lower and only significantly better in the challenging scale change data-set. So as not to distort results, in the low textured data-set number seven (without pre-processing) ORB and HARRIS detectors where taken out of the localization as well as the repeatability chart due to the low amount of features found (Figure 88). The localization accuracies and the respective empirical standard deviations (from ~0.27 pixel for Pyramid HARRIS to ~0.42 pixel for BRISK) lie in a similar range than in the narrow viewpoint data-sets although the viewpoint changes are much higher.

The repeatability of features in these data-sets as illustrated in Figure 90 again shows the same pattern regarding lower performance after contrast enhancement. The ranking is again similar with MSER, DoG and HARRIS at the top end and BRISK features with the lowest repeatability.

The data-sets nine and ten illustrate an increasing viewpoint change on a brick wall (Figure 39) before and after adaptive histogram equalization to evaluate influences on detector performance.

The overall detector localization accuracy performance is slightly worse than for narrower viewpoints. DoG is not able to outperform the other descriptors as in the narrow viewpoint case. This is probably caused by the many well defined corners and edges on which the corner based detectors HARRIS and FAST can achieve higher localization accuracy. Interestingly the performance of all descriptors stays almost constant between the two stations although a viewpoint difference of 25.5 gon is present. Empirical standard deviations as also illustrated in the chart range from ~0.31 pixel for HARRIS to ~0.4 pixel for BRISK.

Figure 92 shows repeatability scores for the brick wall data-set. The corner based detector ORB shows the highest rates followed by the HARRIS and DoG. The already mentioned behavior of lower repeatability scores after adaptive histogram equalization is visible as well.

7.2.3 Feature detection conclusion

The chosen feature detection algorithms where evaluated with respect to their runtime, localization accuracy and repeatability. The definition of these concepts is given Section 7.2.2.1. The algorithm with the lowest runtime was FAST which is 480 times faster than the slowest method DoG with ORB and BRISK also being around 29 and 20 times faster. In absolute values this corresponds to 0.0052 ±0.0003 seconds for FAST and 1.9923 ±0.0383 seconds for DoG.

In the case of the narrow viewpoint dataset the slowest method DoG outperforms its competitors in terms of localization accuracy, whereas BRISK leads to the worst results (Figure 42). In concrete numbers DoG reaches accuracy value in the range of around 0.3 ±0.21 pixel while BRISK leads to results in the range of around 0.7 ±0.4 pixel. The application of adaptive histogram equalization does not significantly influence the localization accuracy. The opposite is true for the repeatability where a negative influence of the pre-processing method can be shown (Figure 43). HARRIS features show the highest repeatability, closely followed by MSER. DoG and STAR based features reach similar results and are ranked third. BRISK performs significantly worse than the other detectors also with regard to repeatability. HARRIS reaches repeatability rates of around 80-85% without pre-processing and around 65% after adaptive histogram equalization whereas BRISK lies in a range below 30% repeatability.

Over all evaluated data-sets the ranking with DoG among the best and BRISK among the worst performing methods stays constant. Differences occur if many well defined corners and edges are present in the data which gives an advantage to the methods HARRIS and FAST and if more scale invariance is required in which case DoG performs best.

The FAST detector method is recommended in the presence of hard time constraints due to its low runtime and good results on well-defined corners if no scale invariance is required. If the processing time is less of an issue DoG is recommended. However, it has to be noted that it is covered by software patents in the US (Lowe, 2004b).

7.3 Local feature descriptor performance evaluation

After local feature description became popular with the publication of SIFT by Lowe (2004a) multiple studies evaluating and comparing the performance of different methods were published (e.g. Mikolajczyk & Schmid, 2005; Gauglitz, et al., 2011 or Miksik & Mikolajczyk, 2012). In addition every publication about a newly introduced descriptor has a section comparing it to other state of the art methods (Bay, et al., 2008; Rublee, et al., 2011; Leutenegger, et al., 2011; Wang, et al., 2011; Fan, et al., 2011; Alahi, et al., 2012). Publicly available test data-sets were created against which new developments can be tested (Scharstein & Szeliski, 2002). The evaluation data-sets generally consist of different sequences to illustrate algorithm invariance towards certain distortions in the image such as increasing in plane rotation, affine transformation, scale change, Gaussian blur or jpg compression artefacts. Most of these sequences consist of man-made structures and cover only one viewpoint.

As mentioned in Section 7.2 literature studies were used as a base decision criterion which descriptors to implement and analyze further in the application of image-based geo-monitoring from IATS data. In the following the conducted evaluation of local feature descriptor performance is described. The evaluation is closely interlinked with the results of the previous step the feature detection. It is also evaluated which descriptors are best suited to describe certain kinds of regions such as homogenous or inhomogeneous areas, and what the effect of the used feature detector on the performance of the descriptor is.

The performance of the descriptors is rated in terms of recall and correct vs. incorrect match ratio values as defined in Mikolajczyk & Schmid (2005) as well as regarding their individual runtime behavior.

7.3.1 Implementation details

The feature description algorithms described in Section 5.3 with the highest performance according to literature were implemented and tested. The methods where integrated in the same command line based test tool already described in Section 7.2.1. The tool was developed as a 32bit C++ application under Microsoft Windows 7 using the development environment Visual Studio 2008. Wherever possible the original implementations of descriptors where used. Therefore the OpenCV (Bradski, 2000) library was integrated on source code basis allowing modifications and detailed evaluations of all integrated functionalities. Further a well-documented, high performance improved implementation of the SURF algorithm available as source code in the OpenSURF library (Evans, 2009) was adapted and integrated. The latter was necessary as the source code of the original implementation of SURF was not released by the authors. The evaluated algorithms BRISK (Leutenegger, et al., 2011; Section 5.3.2.3) and FREAK (Alahi, et al., 2012; Section 5.3.2.4) where developed using the OpenCV library. The original implementations are therefore available and used as source code. The source code of the original SIFT algorithm is also not published therefore the implementation available within OpenCV was used instead.

7.3.2 Accuracy and runtime evaluation

The evaluation of feature descriptors was carried out based on the results retrieved in the evaluation of feature detectors. The analyzed descriptors were evaluated in combination with a feature matching process with the generated correspondences. The goal of this evaluation is the identification of detector, descriptor pairs that lead to the best overall results and the best possible results for certain classes of monitored objects illustrated by the chosen evaluation data-sets. Correctly and incorrectly matched features are displayed as a fraction of the theoretically maximal possible feature correspondences. The matching accuracies are displayed as a deviation of the reference positions in pixel. The following section contains an evaluation of required computational runtimes (Table 3) and the ratio of correct and incorrect matches and the respective accuracies illustrated for each detector, descriptor combination on each individual data-set. These illustrations were generated for features computed with and without adaptive histogram equalization pre-processing and with and without application of outlier filtering to improve the quality of calculated matches. Due to the large amount of data-sets evaluated with different options only a subset of those, offering most valuable comparisons are included in the following section. Figures 93 - 125 illustrate in depth analysis of the descriptor behavior on all individual data-sets allowing an evaluation which method is best suited for which kind of terrain. Figures 46 - 49 towards the end of this chapter contain a compressed overview wherein a mean of the results of all evaluated data-sets was generated in order to show all detector-descriptor combinations in one single chart.

As runtime is one of the key parameters that was evaluated it was measured directly in the individual feature description functions not including data IO operations or image pre-processing. Average values of these runtime deviations over all data-sets measured in nanoseconds are shown in the aggregation of descriptor runtime performance in Table 3. The algorithms are ranked in ascending order and a comparison is done by showing the amount of time needed to calculate each descriptor for one single feature. To set the absolute runtime values into perspective; the evaluation was carried out on a Laptop with an Intel i7-3520M CPU and 8 GB of RAM.

Four different feature descriptors were chosen and applied to the features derived in Section 7.2. The binary descriptors BRISK and FREAK (Sections 5.3.2.3 and 5.3.2.4) and the floating point number descriptors SIFT and SURF (Sections 5.3.2.1 and 5.3.2.2). The evaluation was therefore carried out on the same data-sets and was done on the features computed with and without adaptive histogram equalization (AHE). The same criteria for feature correspondence as in the feature detector evaluation (Section 7.2) were used. A value of maximally possible matches was computed based on the detector repeatability scores as calculated in Section 7.2.2.3. Matching scores are further expressed as a ratio of maximally possible correct matches and computed matches for each detector, descriptor and matcher combination. The amount of possible matches was retrieved from the total amount of features (tables in Section 7.2.2.1), multiplied with the repeatability score of the individual detector as given in Section 7.2.2.3. To generate the graphs the descriptors were matched with the brute force method (Section 5.4.1) using the Euclidian distance for the floating point number descriptors SIFT and SURF and the Hamming distance for the binary descriptors BRISK and FREAK. Further the best to second best match ratio check introduced by Lowe, 2004a (Section 5.4.1) was applied with a minimum distance ratio of 0.65 (as used in Miksik & Mikolajczyk, 2012). To give a representative image of descriptor performance a minimum threshold of ten total correspondences was chosen below which the descriptor results on a certain data-set are omitted in the individual charts (e.g. see Figure 109).

7.3.2.1 Descriptor timing evaluation

The following table shows the aggregated timing of the individual feature descriptors averaged over all executed processing runs as a percentage of the slowest descriptor. The values are given in nanoseconds necessary to describe one feature. The large standard deviations stem from the description of features of different scales. Values are as expected and in accordance with Leutenegger, et al., 2011 and Miksik & Mikolajczyk, 2012. SIFT is the slowest method, followed by SURF and the far faster methods BRISK and FREAK.

Table 3: Aggregation of descriptor runtime evaluation

Detector	Runtime [nano sec]		Percentage of slowest
BRISK	22.44	±12.65	0.60 %
FREAK	45.40	±33.42	1.21 %
SURF	1003.30	±958.74	26.70 %
SIFT	**3757.11**	**±2099.97**	**100.00 %**

The evaluation of time needed for matching was done in a similar manner. The following times describe the nanoseconds needed to compare one pair of descriptors. SIFT takes longest due to the longer descriptor vector compared to SURF. The matching of BRISK and FREAK is fastest due to the simple comparison of the binary descriptors (Table 4).

Table 4: Aggregation of descriptor matching evaluation

Detector	Runtime [nano sec]		Percentage of slowest
BRISK	5.24	±1.91	21.19 %
FREAK	5.28	±2.01	21.34 %
SURF	15.52	±2.71	62.74 %
SIFT	**24.73**	**±1.82**	**100.00 %**

The applied filters, namely the geometry based filter estimating a refined fundamental matrix from calculated correspondences and the disparity length and angle filter comparing a match with its K nearest neighbors were also timed. The geometry based filter takes 0.0624 ±0.0108 milliseconds whereas the disparity length and angle filter needs 0.1498 ±0.0269 milliseconds per evaluated match.

7.3.2.2 Descriptor matching evaluation

Similarly to the previous sections only a sub-set of exemplary evaluation charts is illustrated within the test whereas others were moved to Appendix D Feature descriptor performance evaluation charts to facilitate continuous reading.

Descriptors applied to BRISK features

Figure 93 shows the application of the four chosen descriptors, BRISK, FREAK, SIFT and SURF on features computed with the BRISK detector as a fraction of the theoretically possible matches. Corresponding features are considered a correct match if they are not further away from the reference position than 1.5 pixels. Wrong matches, according to the reference disparity maps may theoretically exceed the ordinate of the graph but will generally lie below the percentage of correct matches. As was to be expected the BRISK descriptor generates the best results for its own detector, followed by the FREAK descriptor which follows a similar processing pattern. SIFT and SURF generate comparable results with much lower percentages of mismatches than the binary descriptors. The BRISK detector appears to have problems to produce distinctive features in the data-set containing well defined cracks, potentially due to their position on edges rather than corners. The amount of mismatches is high but was to be expected due to the weak performance of the feature detector in the detector evaluation process (Section 7.2.2.3).

Figure 94 shows the same evaluation on BRISK features computed on images pre-processed with adaptive histogram equalization leading to a better feature distribution within the image but to less stable points with a lower repeatability. The scores of correctly matched features stay almost constant but the proportion of wrong matches is strongly amplified.

Figure 95 and Figure 96 illustrate the deviation of the matched feature correspondences to the reference. Due to the difference in performance of the individual descriptors and the presence of gross outliers the data had to be visualized on a logarithmic scale.

The logarithmic scale is necessary to properly illustrate for instance the performance of the BRISK descriptor with an average accuracy of around 2 pixels and an average standard deviation of around 8 pixels on the same chart as SIFT with an average standard deviation of 218 pixels due to the overwhelmingly large amount of mismatches. The evaluation of features computed with AHE shows values in a similar range as without pre-processing.

However, these results are clearly not usable without an additional filtering of the detected matches as described in Section 6.1. Figure 44 shows the same data-set with the application of outlier filtering. A precise fundamental matrix describing the epipolar geometry of the images was calculated from the detected correspondences with the known image orientations as an initial estimate (Section 6.1.1). Feature correspondences further away from their epipolar lines than 1 pixel were eliminated. In order to also eliminate mismatches fitting to the calculated model correspondences were restricted based on the mean position on the epipolar line of the K nearest neighbors compared to the position of the current match. These measures on the one hand reduced the amount of correct matches as well as the amount of mismatches due to the strict policy leading to a certain amount of false positives.

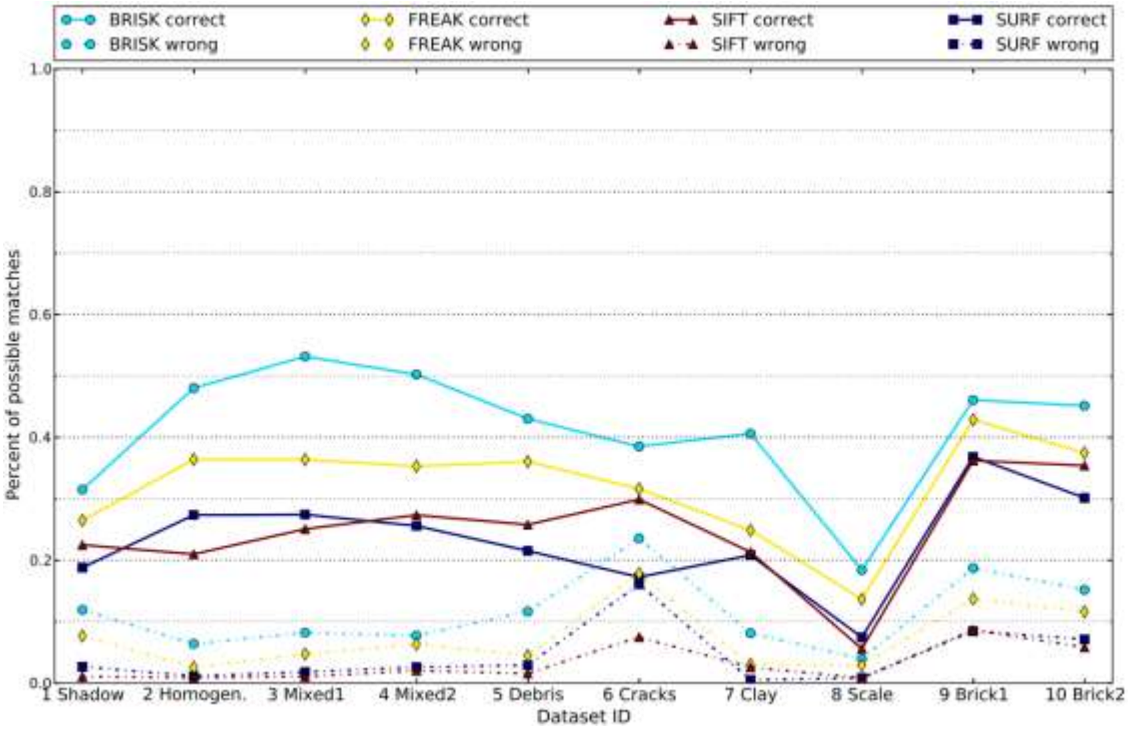

Figure 44: Matches from descriptors on BRISK features with additional outlier filtering

The immediate positive effect of the filtering is visible in Figure 45. Due to the lack of gross outliers a logarithmic scale is not necessary anymore. The improved accuracy can be shown with the same example used previously. The mean accuracy of BRISK over all data-sets was reduced from around 1.9 ± 7.9 to 1.1 ± 1.1 pixel whereas results from SIFT changed from around 56.1 ± 217.7 to 0.8 ± 0.7 pixel. All other descriptors also lie in a similar range of accuracy.

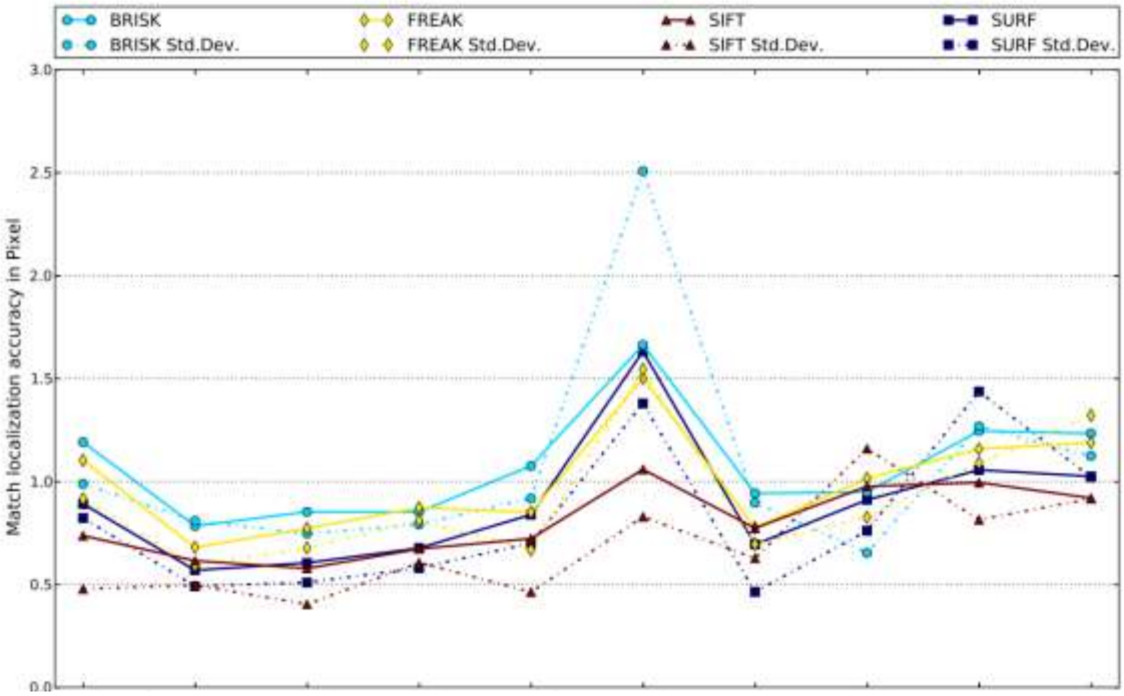

Figure 45: Matching accuracy on BRISK features with additional outlier filtering

Descriptors applied to DoH features

Features computed with the fast Hessian implementation of the DoH detector (Section 5.2.3.1) as introduced together with the SURF algorithm outperform the BRISK features. Interestingly other than in the previous case SURF descriptors on average receive the lowest scores compared to the other tested descriptors on its own detector. As also stated in Miksik & Mikolajczyk (2012) BRISK performs well combined with DoH. Further the DoH features allow a robust matching in the challenging scale change data-set for all descriptors again with BRISK outperforming the other methods. Similarly to the case of the BRISK detector the results based on features derived with AHE pre-processing do not deviate significantly from the results without pre-processing, the corresponding graphs are therefore omitted. Figure 97 and Figure 98 show unfiltered matches and match accuracies whereas Figure 99 and Figure 100 show filtered results.

Descriptors applied to DoG features

Features retrieved with the DoG detector are best evaluated with the SIFT descriptor the detection method was developed for. BRISK competes well but the challenging scale change data-set is handled far better by SIFT. The amount of gross outliers in the unfiltered data-set is already low in the case of SIFT and the matching accuracies in the filtered data-sets are among the highest of all detectors. Evaluations of features computed with AHE are again omitted.

Descriptors applied to FAST features

Using FAST features the SIFT descriptor is able to retrieve the highest amount of correct matches and even results in high accuracy values without additional outlier filtering with around 0.9 ± 0.9 pixel average match accuracy. Due to the property of FAST being a single scale detector no descriptor was able to achieve satisfactory results on the multi-scale data-set which reduces the overall matching scores.

FREAK did not pass the initially mentioned threshold of a minimum of ten matches in the multi-scale data-set and is therefore omitted in the corresponding sections of the graphs. Whereas BRISK and FREAK show the highest overall matching accuracy with around 0.5 ± 0.3 and 0.4 ± 0.3 pixel respectively, SURF performs significantly worse.

Descriptors applied to HARRIS features

HARRIS features have high recognition rates especially for SIFT and BRISK. Due to the fact that not enough features could be detected in the homogenous clay wall data-set there are also no matches visible in Figure 106. Therefore the evaluation of features with AHE pre-processing is shown in Figure 110 to Figure 113. The detector leads to high matching accuracy scores particularly in combination with SIFT. Due to the lack of multi-scale feature detection the scale change data-set cannot be matched satisfactorily.

Descriptors applied to Pyramid HARRIS features

The Pyramid HARRIS detector was developed by the author to make use of the well-known qualities of the HARRIS detector while retrieving scale invariant features. However, the granularity of the used reduction factor for image pyramids calculated prior to the application of the detector proved to be too coarse to successfully deal with the scale change data-set (Figure 114). Therefore only the filtered data-set and accuracy evaluation graphs are shown for this detector.

As for the single scale HARRIS detector not enough features to produce a sufficient amount of matches in the homogenous clay wall data-set were detected. Therefore no information about correspondences is contained in the respective sections of the graphs. The accuracy of the matches lies in a similar range as HARRIS for BRISK and FREAK, whereas SIFT and SURF perform significantly worse than in the single scale case.

Descriptors applied to MSER features

The MSER feature detector leads to the best results in combination with BRISK. Due to the low amount of features found in the homogenous clay wall data-set the evaluation of features computed after AHE pre-processing is shown below. Due to the excellent scale invariance of the detector the scale change data-set could be evaluated well with all descriptors. The resulting matching accuracy (Figure 118) is similar for all descriptors but does not reach a top ranking among the detectors.

Descriptors applied to ORB features

The ORB feature detector is based on a multi-scale implementation of FAST. However, it is not able to successfully deal with the scale change data-set. Due to otherwise lower matching scores it is inferior compared to the original FAST implementation. The combination with SIFT or BRISK work best and lead to acceptable matching accuracies even without outlier filtering. The matching accuracy after outlier filtering over all descriptors is among the lowest of all detectors.

Due to the low amount of features found in the homogenous clay wall data-set again the evaluation of features computed after AHE pre-processing is shown below.

Descriptors applied to STAR features

The STAR detector works best in combination with the SIFT descriptor. This combination leads to the highest overall rate of correct matches over all detector-descriptor combinations evaluated. The scale information given by the detector can be used successfully to also match the challenging scale change data-set, however, SIFT handles this data-set far better than its competitors. SIFT again also leads to useful results without applied outlier filtering. In contrast to the satisfying rate of detected matches the matching accuracy on STAR features is together with ORB and BRISK among the lowest of all detectors.

Comparison of all detector-descriptor combinations

The following charts contain the above shown information condensed into single charts by averaging the results of all detector-descriptor combinations including the described filters over all evaluated data-sets.

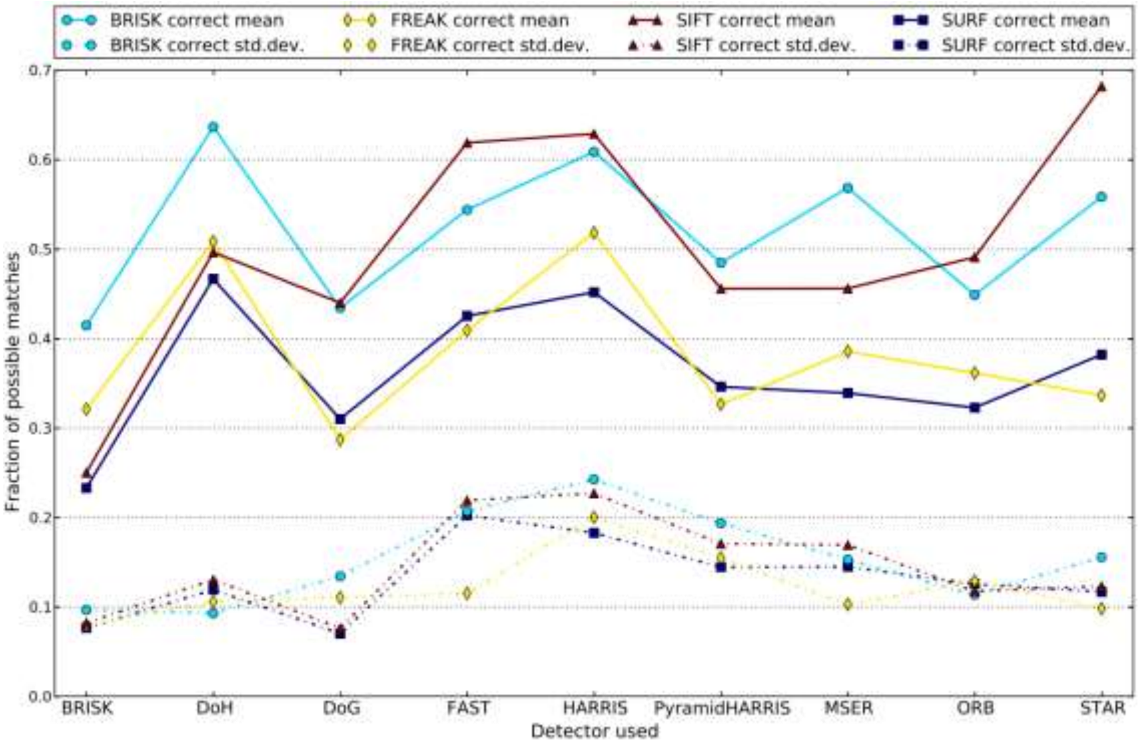

Figure 46: Average fraction of filtered correct matches of possible matches for all detectors and descriptors over all data-sets including standard deviation

Figure 46 shows that the STAR detector in combination with SIFT and the DoH detector used for BRISK lead to the highest rates of correct matches, however, Figure 47 shows that those two are also the combinations leading to the highest amount of mismatches. Therefore the performances of FAST and HARRIS primarily combined with SIFT but also with BRISK have to be valued higher as they lead to much fewer incorrect matches. Further it has to be considered that these two detectors only work in a single scale and cannot be applied in applications where larger scale changes are to be expected. On the other hand this means that the scores of correct matches only calculated over data-sets without scale change puts the combinations HARRIS/SIFT, HARRIS/BRISK and FAST/SIFT at an even level with STAR/SIFT while retaining the mentioned advantages.

As can be seen in Figure 48 DoG and HARRIS lead to the best matching accuracies for SIFT and SURF with the lowest standard deviation (Figure 49). FAST leads to the best matching accuracy for BRISK and FREAK also with the lowest standard deviation.

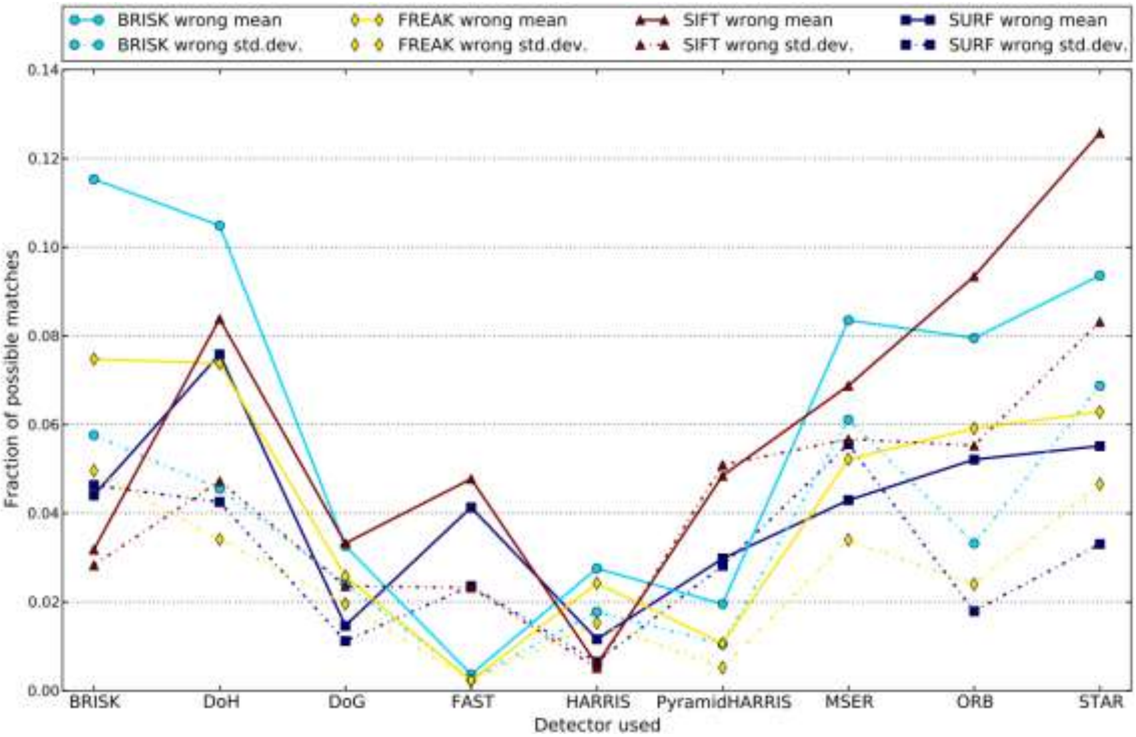

Figure 47: Average fraction of incorrect matches of possible matches (after filter)

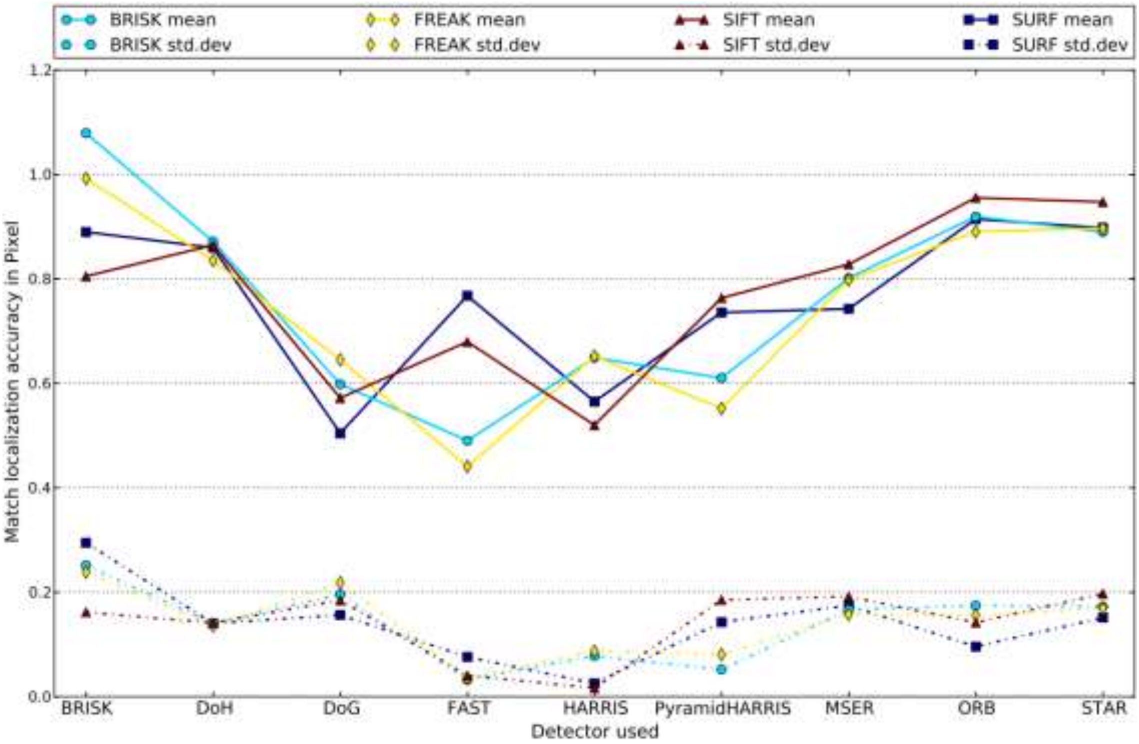

Figure 48: Average matching accuracy for all detectors and descriptors over averages of all data-sets including standard deviation

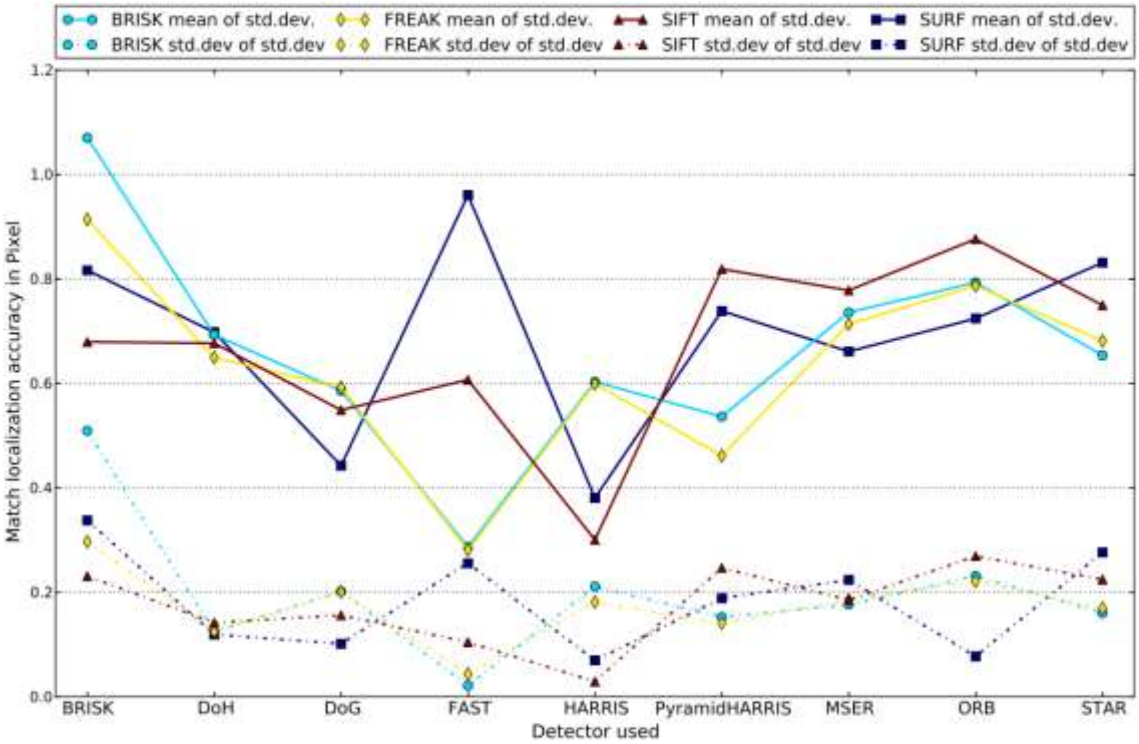

Figure 49: Average standard deviations of matching accuracy for all detectors and descriptors over all data-sets including standard deviation of standard deviations

7.3.3 Descriptor matching conclusion

The conducted evaluations lead to the conclusion that if high processing speed is a requirement the FAST/BRISK combination should be used for a maximum of matching accuracy and an excellent matching rate at a minimal processing time. If another requirement is scale invariance then DoG/BRISK is an adequate alternative, reducing the overall matching rate but able to deal with scale changes very well. If processing speed and scale invariance are not important the combination HARRIS/SIFT can be used.

For an application in the US the software patents covering DoG, DoH, SURF and SIFT have to be considered. In this case FAST/BRISK will be the best option. If scale invariance is absolutely necessary MSER/BRISK or STAR/BRISK are the best options.

7.4 Least squares sub-pixel refinement performance evaluation

7.4.1 Implementation details

The above described method was implemented by the author using the programming language C++ and enhanced to allow a calculation of refinements for different homologous points in parallel. This creates a linear speedup of the matching process by the amount of available cores. Modern Intel CPUs additionally offer virtual cores using a concept called Hyper-Threading, this doubles the amount of available CPUs for processing. On the author's computer with an Intel i7-3520M CPU the speedup factor that can be reached for LSM due to the implemented parallelization is four. Standard office PCs nowadays can have four to eight physical cores bringing the speedup factor to a value of eight or even 16 compared to linear processing. It has to be noted however that despite these optimizations allowing speedups of up to an order of magnitude, due to the iterative process of the algorithm LSM still is the computationally most expensive part of the developed processing chain and should only be applied in case of high accuracy requirements and low runtime restrictions.

7.4.2 Accuracy and runtime evaluation

The full potential of the LSM method is illustrated in a first accuracy evaluation wherein the search image is artificially created from the reference using an affine transformation reflecting a similar stereo set-up as used in data-set number 6 (Table 1). A reduced set of features from the DoH detector was matched with the SURF descriptor, filtered by the geometry based outlier detection method and restricted to 100 correspondences. These matches were refined with LSM. This does not represent a realistic trial as the two images are related by the same transformation that is estimated by the matching method however the convergence process and LSM's ability to approximate this transformation in highest accuracy can be shown. Using a matching window based on the scale of the DoH features and a maximum of 30 iterations the six parameters of an affine transformation and two radiometric parameters consisting of a gray value multiplication factor and a gray value offset where estimated. All 100 matches converged and also showed a back-matching distance of less than 0.1 pixel, resulting in a mean deviation from the calculated reference position of 0.0109 ± 0.0087 pixel. The maximum deviation from the reference was 0.0406 pixel. These impressive values show the maximum matching accuracies that can be achieved with LSM on locally well-defined features within homogenous neighborhoods where projective image distortions can be approximated using an affine model.

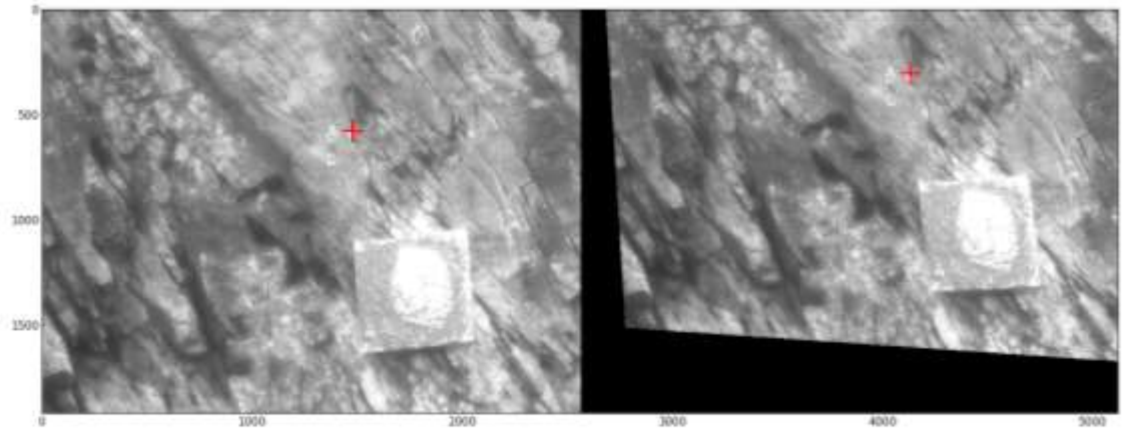

Figure 50: Single stereo-correspondence evaluated with LSM

Figure 51 shows an example of the LSM convergence process of one of the initial correspondences retrieved from feature matching (Figure 50). On the left side of the panels the reference window is shown which stays constant, on the right side the continuously transformed search windows over multiple iterations are displayed. To be able to calculate the necessary image gradients in the search window at every position of the reference window the former has to be one pixel larger.

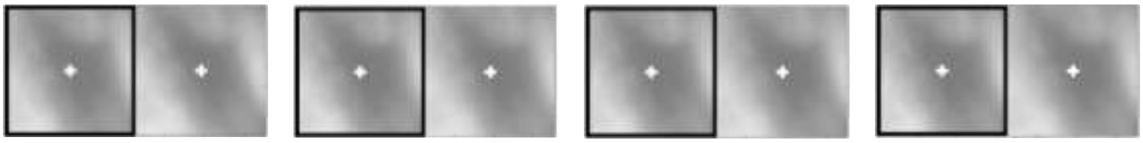

Figure 51: The convergence and refinement process of LSM

In the following analysis the LSM method was used as a post processing and accuracy refinement method for the full set of evaluations shown in Section 7.3.2. Features already filtered according to their averaged runtimes are given and the accuracy results are displayed.

The average runtimes of the parallel calculation of LSM on four separate cores on an Intel i7-3520M CPU with 8 GB of RAM are shown in Table 5 for multiple different parameter settings. Again the shown values are averages over the application of an 8 parameter LSM on all detector-descriptor combinations on all test data-sets. A fixed maximum amount of 30 iterations and a convergence threshold of 0.0001 (see Equation (6.4)). The runtime difference between LSM with and without back-matching is not as large as could be expected as the first matching run already produces accurate predictions that can be used for back-matching including the inverse of the retrieved reference to search window affine transformation. This generally leads to a back-matching convergence within the first iterations and therefore does not significantly increase the runtime.

Table 5: LSM runtimes comparison

Runtime per correspondence [milliseconds]		Window size	Back-Matching
3.631	± 0.047	31	Yes
2.916	± 0.039	31	No
1.016	± 0.038	15	Yes
1.094	± 0.031	15	No
0.842	± 0.025	9	Yes
0.703	± 0.024	9	No

To increase readability a majority of evaluation charts were again placed in Appendix E Least squares refinement performance evaluation charts.

Smaller LSM matching windows lead to a greater rate of rejected matches but higher matching accuracy as can be seen in Figures 126 - 134. The size of the matching window is the main influence on LSM runtime. The LSM evaluations are based on the filtered feature matching results as shown in the previous chapters. These matches are used as initial values for LSM and ensure a fast convergence process. Matching windows of size 9, 15 and 31 pixel were evaluated.

Figure 126 shows the fraction of matches after LSM of theoretically possible matches. The amount of matches was visibly reduced compared to the initial values shown in Figure 46. Additional filtering with the LSM back-matching distance further decreases this ratio but also strongly reduces the amount of incorrect matches as can be seen in comparison of Figure 128 and Figure 129. As a consequence also the resulting accuracy values are higher after an applied back-matching process as illustrated in Figure 130 and Figure 131. Results including filtering based on back-matching are only shown for LSM with a matching window of 9 pixels, deviations to LSM only based on the convergence however are in a similar range also for larger matching windows. As expected LSM results in significantly better accuracies as feature matching alone (Figure 48). With an increase of the matching window from 9 to 15 pixels the amount of correct matches is increased (compare Figure 126 and Figure 52; Figure 128 and Figure 53) whereas the matching accuracy is reduced (compare Figure 130 and Figure 54). While the second effect, the reduced matching accuracy is further increased with an additional increase of search window size from 15 to 31, the amount of correct matches does not change (compare correct matches in Figure 52 and Figure 132 and accuracies in Figure 54 and Figure 134). Based on this evaluation regarding the involved parameters of correct matches, matching accuracy and runtime a matching window of 15 pixels therefore seems to be the best overall tradeoff.

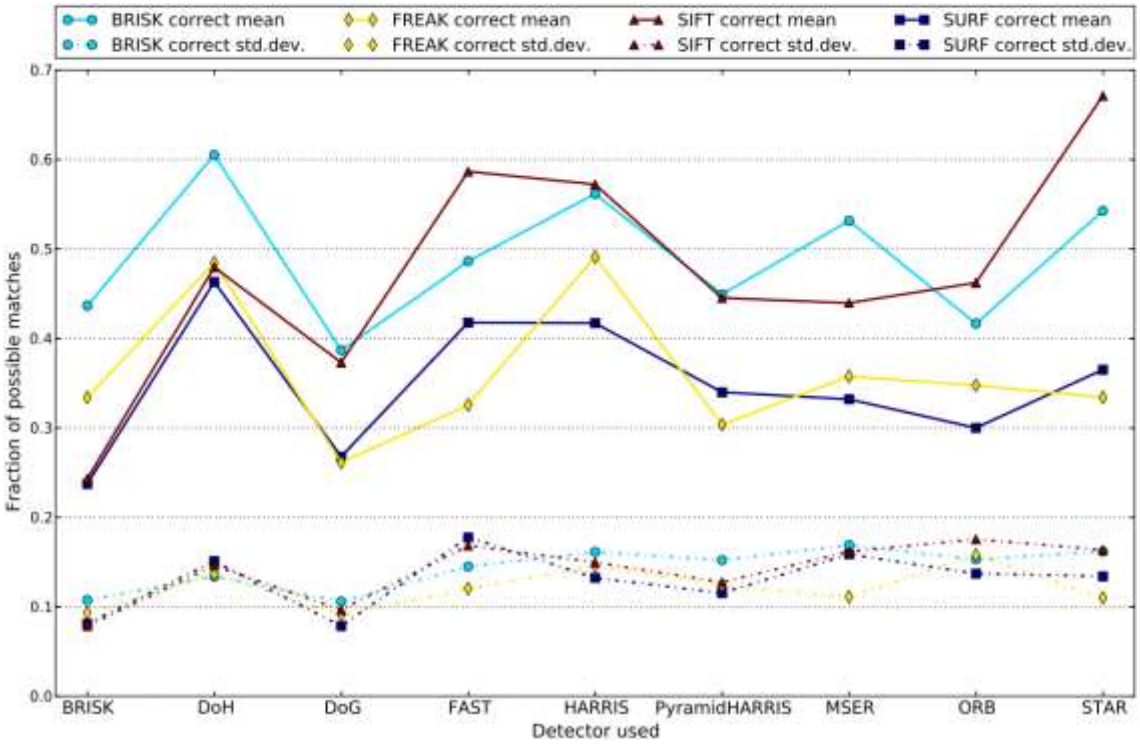

Figure 52: Average fraction of filtered correct matches of possible matches from LSM with a window size of 15 pixels

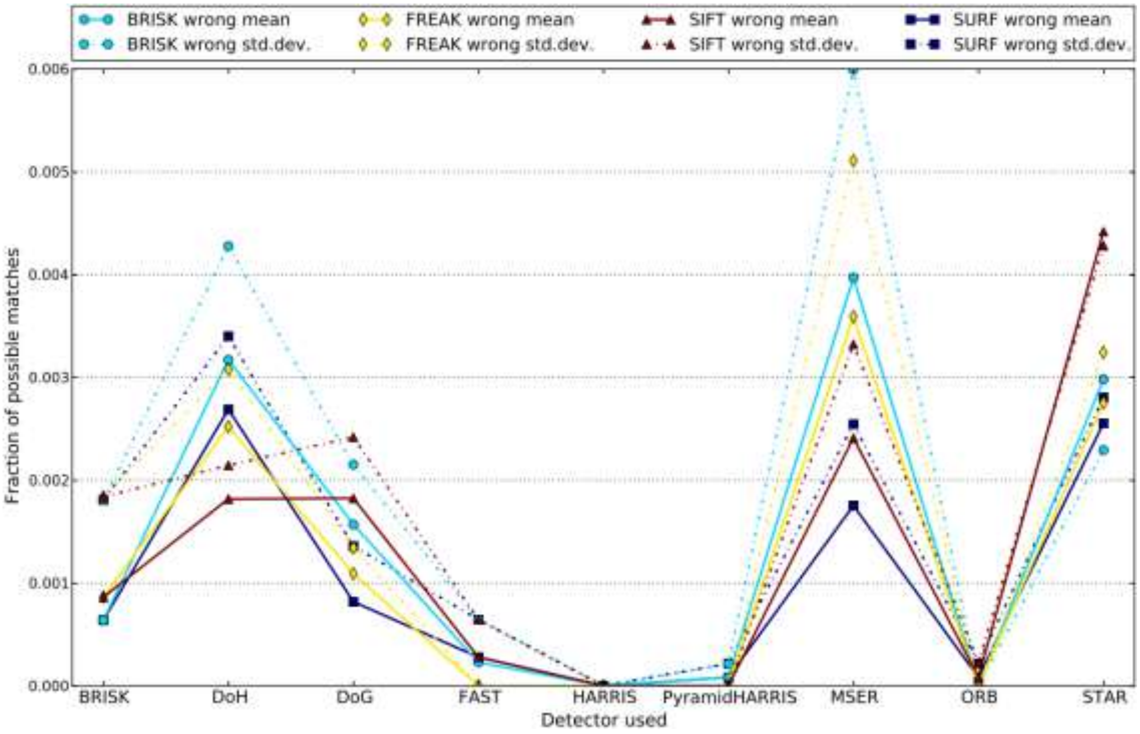

Figure 53: Average fraction of filtered incorrect matches of possible matches after LSM with window size of 15 pixels

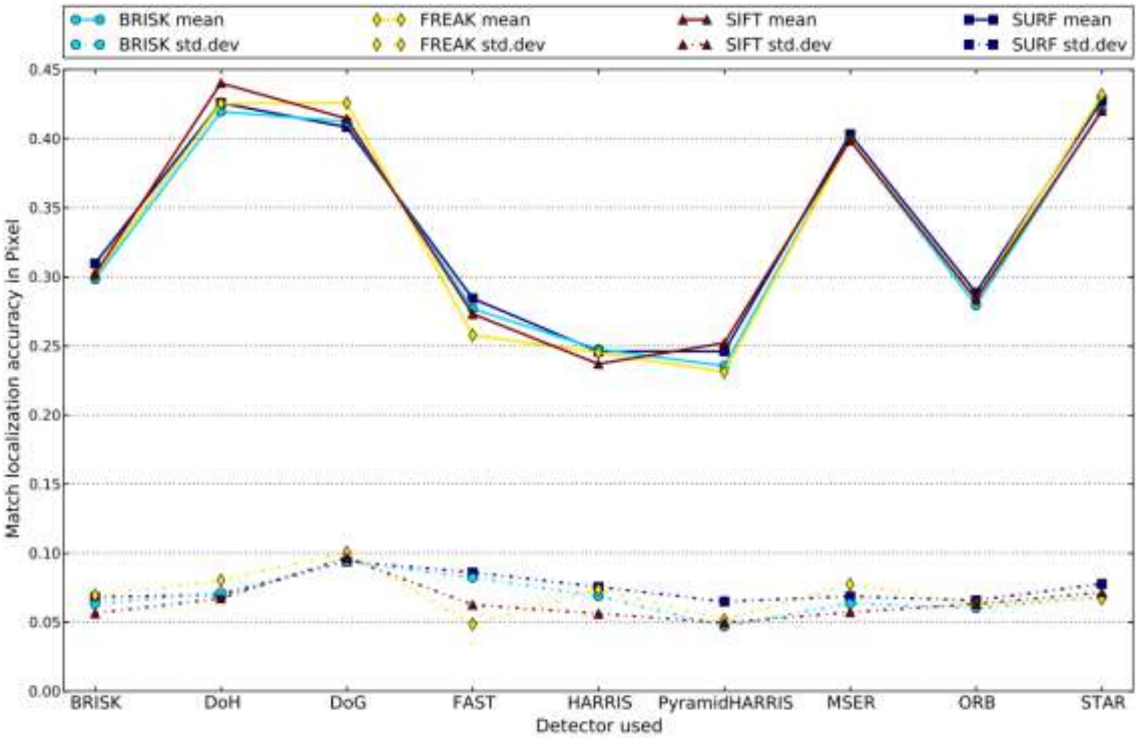

Figure 54: Average LSM accuracy with an LSM window size of 15 pixels

7.4.3 LSM evaluation conclusion

The applied feature descriptor does not influence the matching accuracy that can be reached with LSM. However, the used feature detector significantly influences the results. HARRIS and FAST lead to the highest LSM accuracies. Due to their close relation to FAST the ORB and BRISK detectors produce similar results when used as initial values for LSM.

Due to a satisfying trade-off of achievable accuracy, the amount of correct features that are preserved and runtime a LSM matching window of 15 pixels is preferable. If sufficient input correspondences are available and highest accuracy is required a matching process with a matching window of only 9 pixels and an additional validation of the results using a back-matching process can be applied. Larger matching windows are not reasonable.

7.5 Conclusion and interpretation

To bring the performed evaluations back into the context of geo-monitoring it has to be kept in mind that there are different requirement to a feature detection and matching algorithm to be applied in an actual monitoring system as mentioned earlier. On the one hand the matching accuracy is important as it significantly influences the final 3D accuracy of the measurements. On the other hand the repeatability and stability of derived results regarding outliers or mismatches is at least as important. Further, depending on the current measurement configuration the processing time constraints pose a limiting factor on the choice of algorithm.

An interpretation of the conducted evaluation based on these criteria shows that a proper detection of mismatches using global or local geometric constraints is crucial to achieve the required accuracies. This is clearly visible by comparing for instance the charts in Figure 45 and Figure 95 illustrating matching accuracy with and without applied outlier filtering. Both global and local outlier filtering methods are used in the following section of real world measurement data evaluation. Within stereo setups the epipolar geometry (Section 6.1.1) is used to filter outliers, over multiple time epochs the local RSOC filter (Section 6.1.2) is applied.

The combined analysis of feature detector performance and its influence on feature description and matching allows the identification of best performing detector/descriptor pairs to be used for certain measurement configurations. The FAST feature detector should be used in combinations with the BRISK feature descriptor for highest processing speed. It leads to a maximum of matching accuracy combined with a low amount of mismatches computed with the lowest processing time of all evaluated methods for the detector as well as the descriptor. The combination of the HARRIS detector and SIFT descriptor leads to similar accuracy values and slightly higher matching rates, however, at a much slower processing speed. Furthermore, the descriptor is covered by US software patents which may prohibit its application outside of academic use.

If the measurement configuration covers larger scale differences, as may be encountered if the stereo-base line lies in the direction of to the measured object (i.e. Section 7.1 data-set ID8) a scale invariant feature detector is required. This leads to the combination of DoG as detector and BRISK as descriptor, producing satisfying results at the cost of longer feature detection processing time. The application of the DoG detector is covered by the same US software patents as SIFT. To avoid this issue while retaining a high level of scale invariance the detector descriptor combination MSER/BRISK can be used for higher accuracy or STAR/BRISK for better matching rates. Both options are however inferior to the recommended, scale invariant DoG/BRISK combination.

Using these feature matching methods mean accuracies of around half a pixel can be reached leading to an angular resolution in the range of the accuracy of the base-instrument (Section 2.1.). For better matching results a sub-pixel refinement method such as LSM is necessary. A previously applied feature matching process delivers excellent predictions for this method. The resulting accuracy however varies with the applied feature detector. The use of a HARRIS detector leads to the best LSM results closely followed by FAST. Both resulting in a mean matching accuracy of less than 0.3 pixel and a greatly reduced standard deviation lying between 0.05 and 0.1 pixel. It remains to be noted however that the application of LSM may significantly increase the required processing time. Therefore the method will most likely be used in applications without strong time constraints. The positive results of the application of LSM on the accuracy and variation of derived 3D deformation vectors can be seen in the evaluation of field trial data in the following chapter (Figure 66 and Figure 67).

8 Evaluation of system design using measurement data

To evaluate the developed system under field conditions multiple tests were conducted to validate the IATS calibration, set-up and control software on the one hand and the accuracy of the applied processing on the other hand. Among smaller measurement campaigns targeted at system set-up and orientation two larger field experiments were carried out to capture data including measurements from theodolites and terrestrial laser scanners in order to provide reference data for the IATS measurements. The following chapter contains the evaluation of the developed processing chain using the captured reference data.

8.1 Field trial Graz Weinzödl

This field trial was carried out at a rock wall in the north of Graz on the 25th of February 2013 by the DE-MONTES project consortium (Bauer, et al., 2014). The main target of this field trial was on the one hand a technical end to end testing of the applied device control software and on the other hand an accuracy comparison of the applied methods of stereo photogrammetry, IATS and TLS. Important for the scope of this work a test with an artificial deformation was conducted using a rock model made of structured and painted foam to target the field of deformation monitoring. Targets on this artificial rock were measured with a total station. These artificial deformation data-sets were evaluated using the developed processing chain.

Due to the date of the measurement campaign in the end of winter the whole measurement system proved to be able to withstand non-ideal weather conditions as it was snowing during half of the data capturing.

8.1.1 Object and measured region of interest

The target object was the midsection of a rock climbing wall of around 20 by 40 meters (Figure 56), measured from a distance of around 150 meters. Multiple data-sets where acquired using a stereo set-up of IATS (Leica IATS2 see Section 2.2), a Riegl VZ 400 terrestrial laser scanner and a conventional photogrammetric stereo set-up. Fourteen reflecting control points in the form of circular targets where placed on and around the object and measured using a Leica TCRA1201 Total Station to establish a control network. The positions of the IATS, TLS and Total Station can be seen in Figure 55.

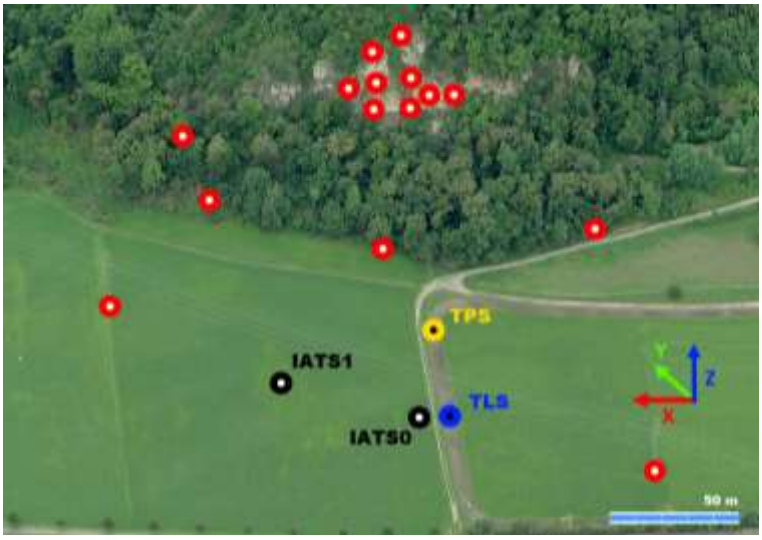

Figure 55: Field Trial Test Site at Climbing Wall Weinzödl north of Graz. Aerial Image © Microsoft

Figure 56 gives an overview of the dimension of the measured Region of Interest and the distribution of the placed control points. A local left handed Cartesian coordinate system was defined with the measured control points using polar point measurements from a single position. The used disc shaped targets have a diameter of around 20 centimetres and where sheeted with a retro-reflecting lamination to be clearly detectable in the visible spectrum as well as in the reflectance image of the Laser Scan. The well distributed targets are not necessary for the IATS measurements but were used to register the measurement results from all used sensors into one common coordinate frame to compare their relative accuracy. The comparison and evaluation of TLS and conventional photogrammetry however lie outside the scope of this work. Only the artificial deformation measured with a total station and the IATS system was used in this evaluation. The artificial deformation was carried out close to target 14 visible in Figure 56.

Table 6: Target accuracies as a result of TPS network adjustment

	Accuracy [m]	
Target	**Position**	**Height**
8	0.0040	0.0062
11	0.0066	0.0104
12	0.0045	0.0071
13	0.0035	0.0054
14	0.0045	0.0071

Table 6 gives accuracy values of the control points that are later visible in both IATS bundles (Figure 57). For more information refer to Appendix G Adjustment report (TPS network Weinzödl).

Figure 56: Measured Region of Interest, including numbered control points and close-up of mounting of target 14

8.1.2 IATS measurements

As depicted in Figure 55 the IATS Stations where set up in an average distance of around 147 m and 163 m to the object with a baseline of around 47 meters. Due to their different angles of observation the two devices captured 150 (left station) and 220 (right station) images to cover the area of interest with image bundles of 10x15 and 11x20 rows and columns respectively. This resulted in an observation time of ten minutes for the left station and fifteen minutes for the right station. With an angle resolution of 0.61 mgon per pixel the field of view of one image is ~1.56 x 1.71 gon or ~3.68 x 2.76 m in a distance of 150 m. The ground sampling distance (GSD) of one pixel at that distance is therefore ~1.4 mm.

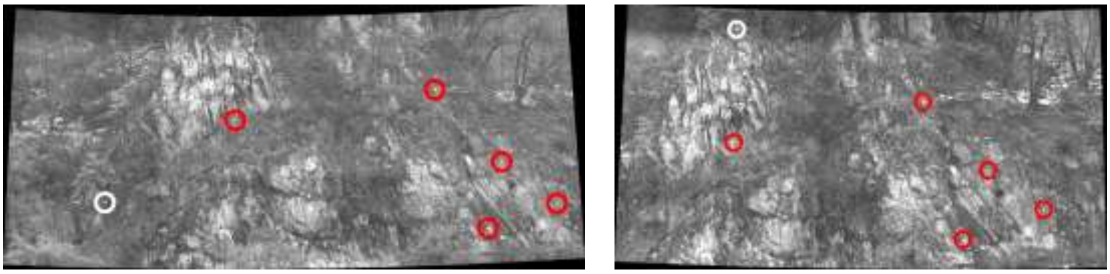

Figure 57: Spherical Panoramas composed of the captured image Mosaics; red targets are visible from both stations – white targets only from one station

The orientations measured by the IATS together with the known IATS calibration where used to project the captured images into a spherical panorama as shown in Figure 57.

Table 7: Residuals of IATS measured targets to TPS Measurements

Target	Δ X [m]	Δ Y [m]	Δ Z [m]
8	0.000	-0.001	0.003
11	-0.008	-0.004	-0.008
12	0.000	0.006	-0.008
13	-0.004	-0.003	0.003
14	-0.002	-0.001	-0.001

The centres of the individual targets where extracted within the images in a high sub-pixel accuracy using intensity thresholding operations and the calculation of the center of gravity. Taking into account the intrinsic calibration of the two cameras a forward intersection was calculated for each target resulting in the 3D residuals shown in Table 7.

8.1.3 Artificial deformation

In order to capture a large enough controlled deformation an artificial rock was used consisting of a square wooden plate with a side length of about one meter. In order to simulate well-textured and structured rock polyurethane foam, gravel and paint were applied. Four reflective targets on the corners of the artificial rock allow precise positioning measurements with a total station. The model was placed in a flat section of the rock wall below target number 14 about three meters above the ground (Figure 56 and Figure 58).

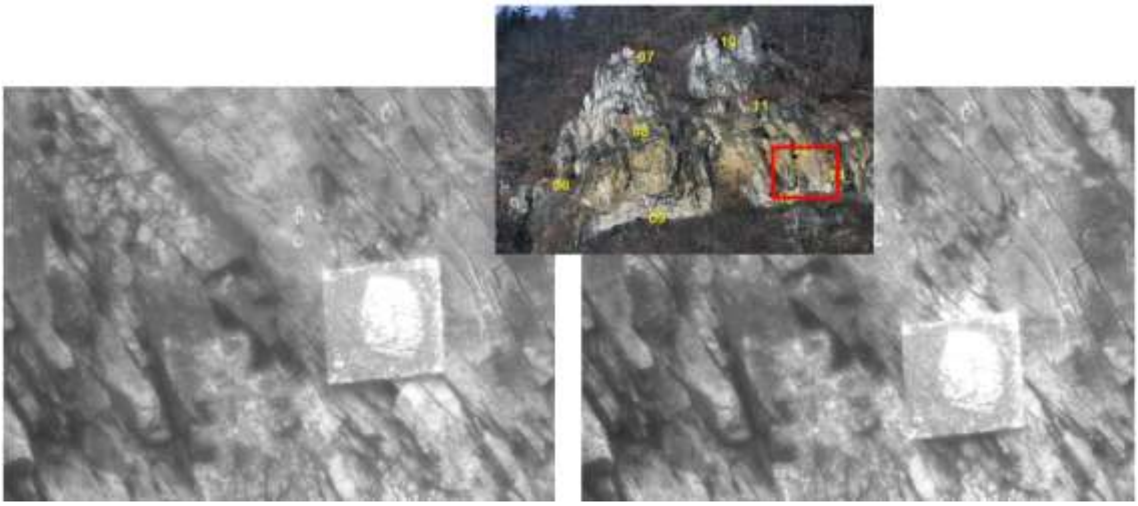

Figure 58: Artificial rock model as seen from the left IATS station before and after an artificial deformation

The artificial rock was measured in single IATS image tiles from each station (Figure 58) as a part of a full image bundle while the four signalized corner targets were observed with a Leica TCRA1201 tacheometer. Due to weather conditions and time restrictions the theodolite measurements were carried out from only one location. Therefore no quality measures for the observed points are available. The theoretical accuracy of the polar distance measurements of the device is stated by the manufacturer with an RMS of ±2 mm by the manufacturer for each target (Bauer, et al., 2014).

8.2 Field trial Pellheim

The second major field trial was conducted in a clay-pit in Pellheim near Munich. The pit is owned and operated by Hörl & Hartmann Ziegeltechnik GmbH & Co. KG. The field trial was carried out to obtain control points of higher accuracy and higher resolution TLS measurements compared to the measurement campaign in Graz (Bauer, et al., 2014).

8.2.1 Object and measured region of interest

A vertical wall at the west end of the clay-pit was chosen as measurement area due to the interesting structure and texture and its accessibility from above as well as from below. The IATS devices where set up in a mean distance of around 261 and 292 meters from the measured object (Figure 61). The control and reference points of the ground control point network (Figure 60) were measured by a Leica TPS1201 and a Leica MS50 total station which have angle measurement accuracies of 0.3 mgon each. Distance measurements can be carried out with an accuracy of 1 mm ±1.5 ppm for prisms and 2 mm ±2ppm for surfaces. Additionally for every measurement epoch the region of interest was also scanned using a Leica HDS7000 laser scanner which was set up at a distance of only around 80 meters from the ROI. Scans were taken with a resolution of 10 mm point distance in the target area.

Figure 59: Field trial clay-pit Pellheim overview. Satellite image © Google

Figure 60 shows the region of interest including the control points used to register the measured data-sets from the different sensors. Further the artificial rock and its targets are visible on the measured wall.

Figure 60: Overview of region of interest including control points and artificial rock with targets

To measure the control point, network observations from four different stations were carried out using forced centering for targets and instruments. Resulting accuracies and confidence levels after a network adjustment conducted at the Technical University of Munich are visible in Figure 61 (refer to Appendix F for the adjustment report). The accuracy of the control points and the artificial rock was determined as 3 mm for position and height respectively. For the TPS, TLS and IATS positions 2 mm were derived for position and height.

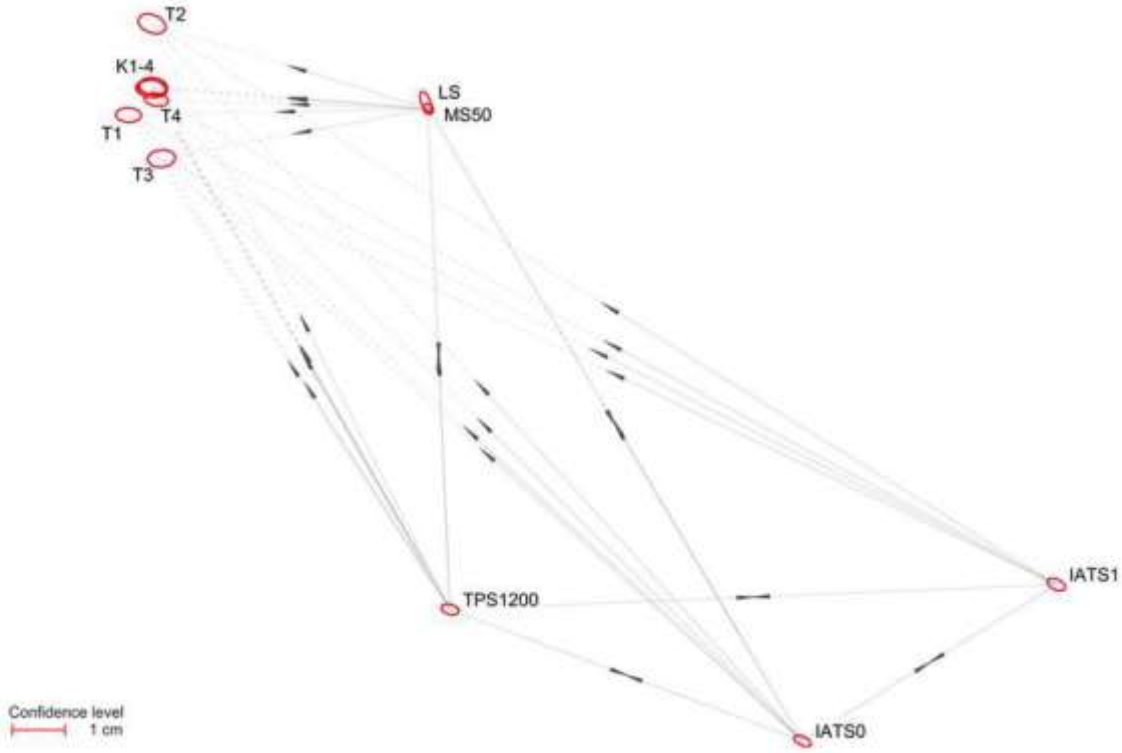

Figure 61: TPS network and point accuracies

To orient the laser scans in each epoch the four control points (T1 to T4) consisting of black and white targets were scanned separately with the highest possible resolution of ~4.5 mm point distance. The TLS measurements and the TPS measured target positions were used to calculate the three rotations and translations of the transformation orienting the laser scans resulting in maximal residuals of ±3 mm. From a comparison of two scans of the target region an average range noise of the laser scanner of around ±3 mm was calculated corresponding well with given manufacturer specifications (Bauer, et al., 2014).

8.2.2 IATS measurements

The two IATS stations were set-up in an average distance to the region of interest of 261 m for IATS0 and 292 m for IATS1. The baseline between the instruments was 84.47 m. This leads to a mean intersection angle in the measured object points of around 18 gon. This value represents an adequate compromise between the image matching condition of as little viewpoint change as possible to retrieve highest accuracy correspondences and the geometrical condition whereupon the highest accuracies in all three directions can be reached at intersection angles of 100 gon. The locations of measured object points are not on the perpendicular bisector of the baseline as would be ideal for the triangulation. However, the set-up is representative for real world applications wherein the conditions of the site often prohibit a theoretically optimal setting leading to compromises such as the above. It could be shown that highly accurate results can be retrieved also from non-optimal measurement configurations.

Within each measurement epoch 32 images where taken by IATS1 and 40 images by IATS0 to cover the measured region. With an acquisition time of around five to six seconds per image this process takes around three and four minutes respectively.

Table 8: Residuals of IATS measured targets to TPS Measurements

Target	Δ X [m]	Δ Y [m]	Δ Z [m]
T1	0.005	-0.003	0.001
T2	0.001	0.001	-0.001
T4	0.005	0.004	0.001

To verify the orientation of the IATS a template matching was performed to retrieve the sub-pixel image positions of the targets visible in the images. Table 8 shows the residuals of the target coordinates calculated from IATS to the adjusted values from the TPS network (Wagner, et al., 2014).

8.2.3 TLS measurements

A reference model of the target area was captured using a Leica HDS7000 TLS from a distance of only around 80 m which is less than a third of the distance of the IATS to the object. A point density of 10 mm was used and the scan was oriented by scanning the black and white targets T1-T4 (Figure 60). The scanner is subject to range noise of ±2 – 8 mm at this measurement distance according to the manufacturer. Based on the comparison of two successive scans of the same region a range noise of 2.7 mm was calculated. Other errors such as wrong absolute distance measurements are eliminated by the registration of the scan.

Table 9: Residuals of transformation to register TLS

Target	Δ X [m]	Δ Y [m]	Δ Z [m]
T1	-0.003	-0.003	0.001
T2	-0.001	0.001	0.001
T3	0.003	-0.001	-0.001
T4	-0.003	-0.001	0.001

Table 9 shows the residuals of the transformation (three rotations, three translations) used to orient the laser scan (Wagner, et al., 2014).

8.2.4 Artificial deformation

Similarly to the field trial in Graz an artificial deformation was measured using the same artificial rock equipped with signalized targets at its corners. Due to the fact that the test site was accessible from above as well as from below the artificial rock could be moved comfortably (Figure 60). In contrast to the previous measurement campaign the signalized points on the moved target where measured from two positions to increase the localization accuracy and to provide quality estimates for the observations.

8.3 Processing chain

This work deals with the task of geo hazard monitoring based on a vision based geodetic system using a stereo set-up of IATS. Simplified the necessary steps are the detection of non-signalized stereo-correspondences and the tracking of those features into consecutive measurement epochs wherein stereo-correspondences have to be established again. This leads to highly accurate 3D deformation vectors describing the movement of a measured object.

The processing chain was developed in the high-level programming language python and integrates the necessary individual pieces of developed software for

- image pre-processing
- feature detection, description and matching,
- mismatch detection and
- match refinement.

These components are mostly C++ applications except the RSOC mismatch detection filter which was only implemented in pure python based on numpy the optimized python library for numeric calculations. Other functionality such as the least squares matching were first implemented and tested in python and then ported to a C++ application in order to speed up processing. The python workflow mainly handles interfacing issues. This means on the one hand the interfacing to the user based on configuration files or command line parameters and on the other hand the preparation and assembly of parameters needed by the individual sub-functions. Another important task is data format conversion and preparation. As an example the feature matching may write a correspondence file containing information not only about the sub-pixel positions of homologous features but also their scale space position and extraction filter response value, whereas the tool executing the forward intersection needs a file containing only an index and the left and write column row pixel positions. In this case the workflow handles the conversion and keeps track of the files to be used.

Most functionality developed by the author is contained in a single C++ library accessibly from a command line application containing histogram-equalization or adaptive-histogram-equalization pre-processing methods, all evaluated feature detectors and descriptors, brute-force, adaptive nearest neighbors and epipolar-restricted matching methods, mismatch detection based on a homography or fundamental matrix estimation or refinement and other filter methods and a match refinement using massively parallelized LSM as described in Section 6.2. Further the calculated features and feature vectors can be saved as well as diverse output images illustrating features and correspondences (refer to Appendix H for the full list of available command line options of the mentioned tool). Other used C/C++ command line tools are DispTo3d.exe to calculate the spatial forward intersection using two camera calibration files and a feature correspondence file, and the tool 3dToDisp.exe which projects a list of 3D coordinates into an image of which the calibration was specified. These two tools were developed at the Joanneum Research as part of the Impro image processing library.

The matching parameters can be configured either as fast processing which will use FAST feature detection and BRISK feature description without LSM refinement or as high accuracy processing which applies HARRIS feature detection and SIFT feature description including LSM refinement. If this granularity of parameter definition is not fine enough the individual parameters can be exchanged as well in a parameter file. The structure of the processing chain is as follows.

The left image of epoch one is L1, the right image is R1. L2 and R2 are the images of the consecutive epoch.

- L1 match to R1
    - Pre-processing
    - Feature detection
    - Feature description
    - Feature matching
    - Mismatch detection
    - (LSM match refinement)
- Forward intersection for correspondences from epoch 1
- L1 match to L2
    - Pre-processing
    - Loading corresponding matches from L1 R1 as features for L1
    - Feature detection in L2
    - Feature description
    - Feature matching
    - Mismatch detection
    - (LSM match refinement)
- RSOC mismatch filtering
- L2 match to R2
    - Pre-processing
    - Loading corresponding matches from L1 L2 as features for L2
    - Feature detection in R2
    - Feature description
    - Feature matching
    - Mismatch detection
    - (LSM match refinement)
- Forward intersection for correspondences from epoch 2
- Creation of output files
    - Creation of 3D deformation vectors file
    - Back-projection of epoch 2 coordinates into L1
    - Creation of 2D projections of deformation vectors in L1
    - Draw deformation vectors in L1
    - Save processing log
    - Save processing time

8.4 Results and evaluation

The processing chain outlined in the previous section was applied to the acquired test data described in 8.1 and 8.2 and the processing results are compared to the reference data.

8.4.1 Field trial Graz Weinzödl evaluation

As described in Section 8.2.1 the artificial deformation carried out in Graz was applied manually and measured with polar distance measurements of four signalized targets in both measurement epochs. The used stereo data-sets, consisting of four single IATS tiles from two measurement epochs are shown in Figure 62. The images from the left station cover a field of view of ~3.6 x 2.7 m in a distance of 147 m with a pixel resolution of ~1.4 mm. The right station is placed at a distance of 163 m and the images cover an area of ~4 x 3 m with a pixel resolution of ~1.,6 mm.

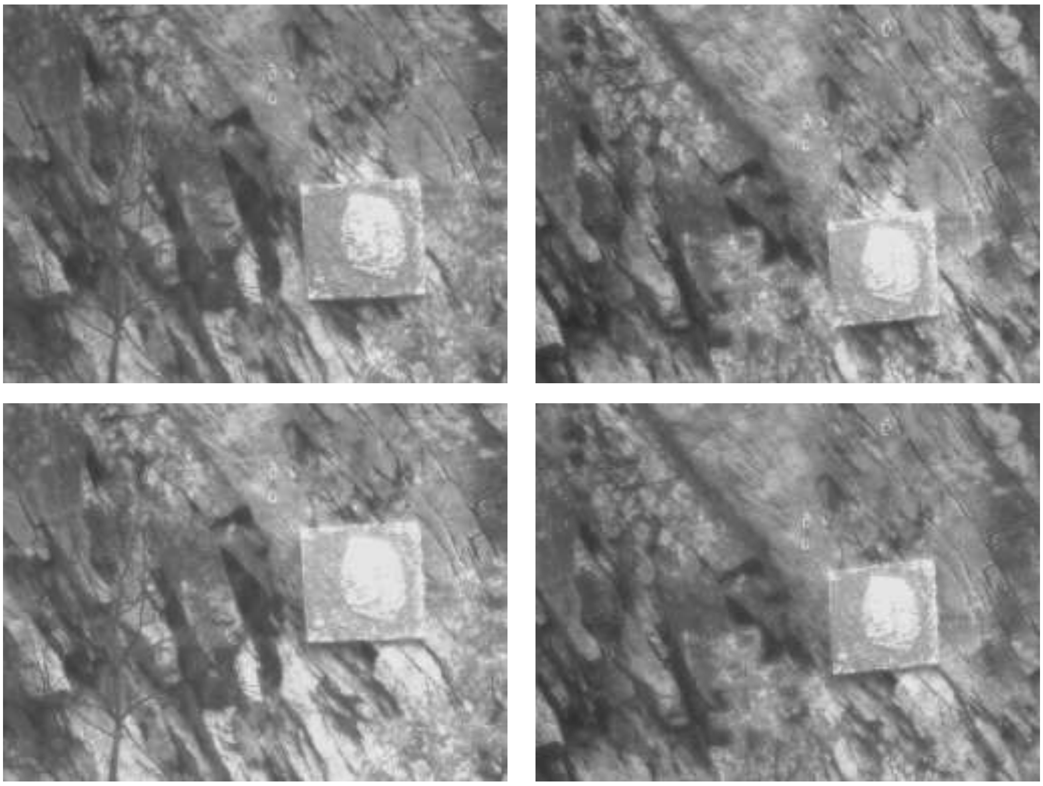

Figure 62: Artificial deformation in consecutive measurement epochs

The applied deformation was a vertical shift by around 0.424 m according to the differences of measured targets as shown in Table 10.

Table 10: Coordinate differences between epochs

Target	Δ X [m]	Δ Y [m]	Δ Z [m]	Displacement [m]
TP01	0.001	0.000	0.427	0.427
TP02	0.002	-0.003	0.419	0.419
TP03	0.009	-0.005	0.425	0.425
TP04	0.008	0.004	0.423	0.423
Ø	**-0.001**	**0.005**	**0.423**	**0.424**

Following the findings from Chapter 7 a fast and coarse processing run using the processing chain described in Section 8.3 with the FAST detector and the BRISK descriptor and an epipolar restricted matching was applied. The resulting projected deformation vectors are shown in Figure 63.

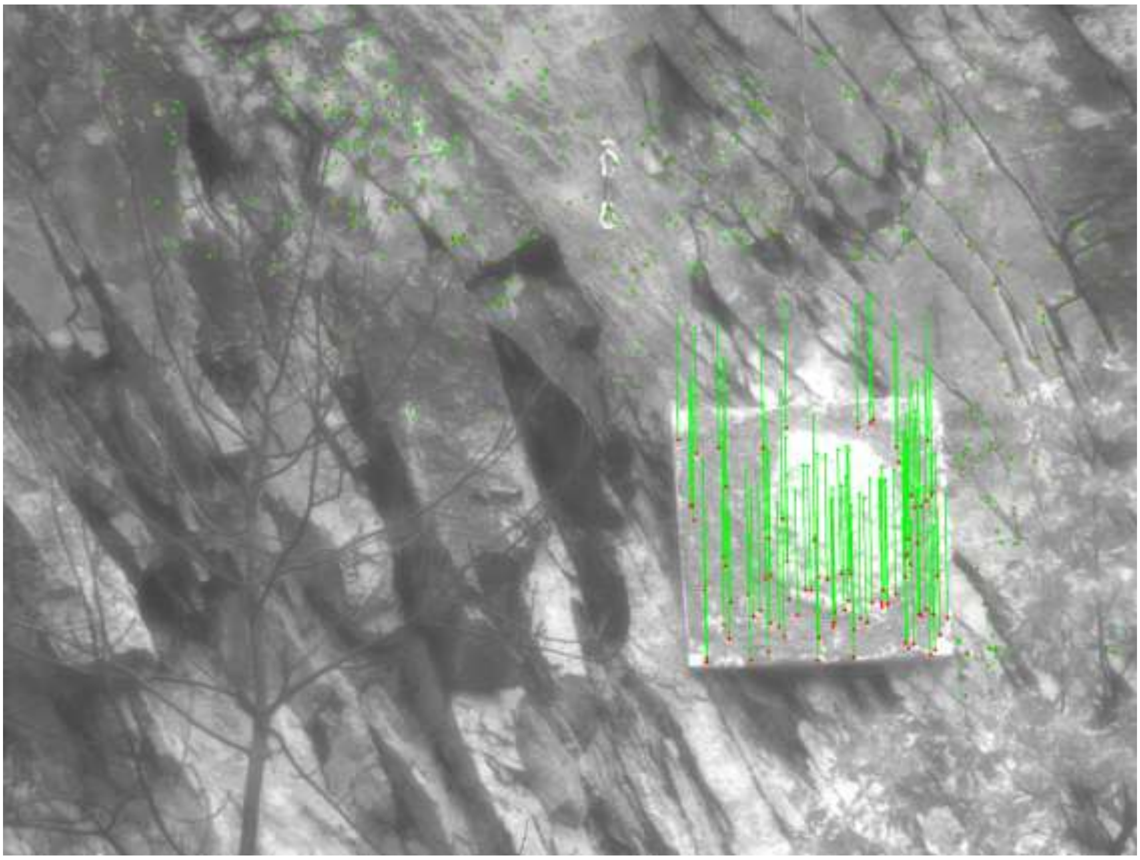

Figure 63: 3D vectors from FAST/BRISK matching projected into left image of first epoch

A comparison of the length of the automatically derived deformation vectors of the artificial target with the TPS measurements leads to the results shown in Figure 64.

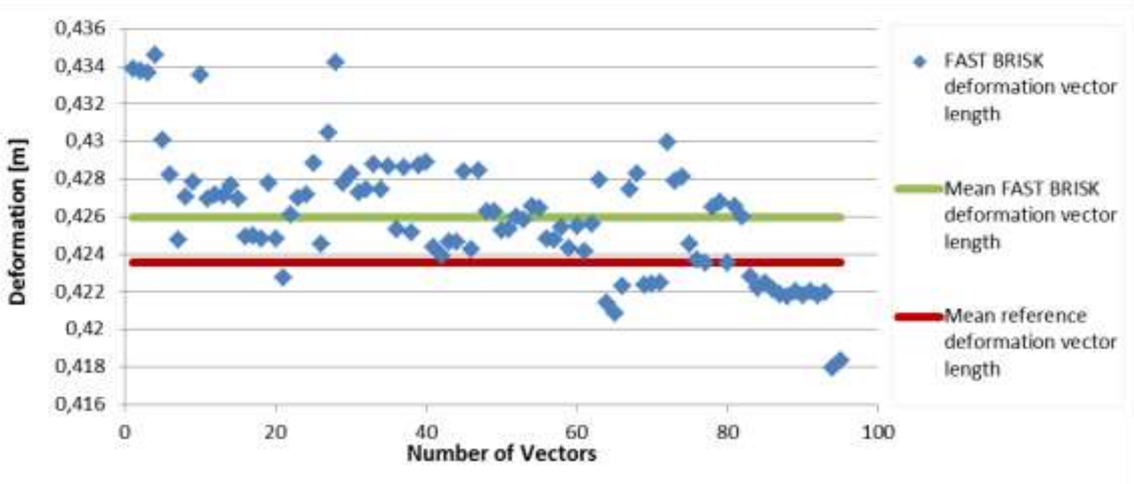

Figure 64: FAST BRISK deformation vector length compared to mean length from TPS

The processing time of the FAST BRISK based application of the automated deformation measurement on a Laptop with an Intel i7-3520M CPU and 8 GB of RAM is given in Table 11. The values were averaged over 10 processing runs leading to an overall standard deviation of around 0.35 seconds for the whole processing chain.

Table 11: Processing time using FAST and BRISK

Processing Step	Execution Time [seconds]
Pointmatching L1R1	3.25
Format conversion and forward intersection	0.39
Pointmatching L1L2	1.87
RSOC filter (python implementation)	1.75
RSOC filter (estimation for C++ implementation)	*0.18*
Pointmatching L2R2	1.59
Format conversion and forward intersection	0.22
Output generation and visualization (python implementation)	2.20
Output generation and visualization (estimation for C++ implementation)	*0.22*
Total (with current python implementations)	**11.27**
Estimated total with non-existent C++ implementations	**7.71**

Although a constant amount of initial features is extracted the timing values depend on the amount of retrieved matches in the first epoch and may therefore vary over different data-sets. Values of the shown data-set therefore just serve as a representative example.

Two steps of the processing, the RSOC filtering and the output generation and visualization are still implemented in python only. Based on experiences with the implementation of LSM both in python as well as in C++ a speedup by a factor of 10 to 20 is possible when implementing processing steps in C++ and additionally making use of parallelization. These changes could bring the amount of required processing time from more than 11 seconds down to only around 8 seconds. Considering these figures were derived on a two year old laptop, values in the range of under 5 to 6 seconds needed for a fully synchronous data capturing and evaluation are without question possible using an up-to date high-level laptop or mid-level desktop workstation.

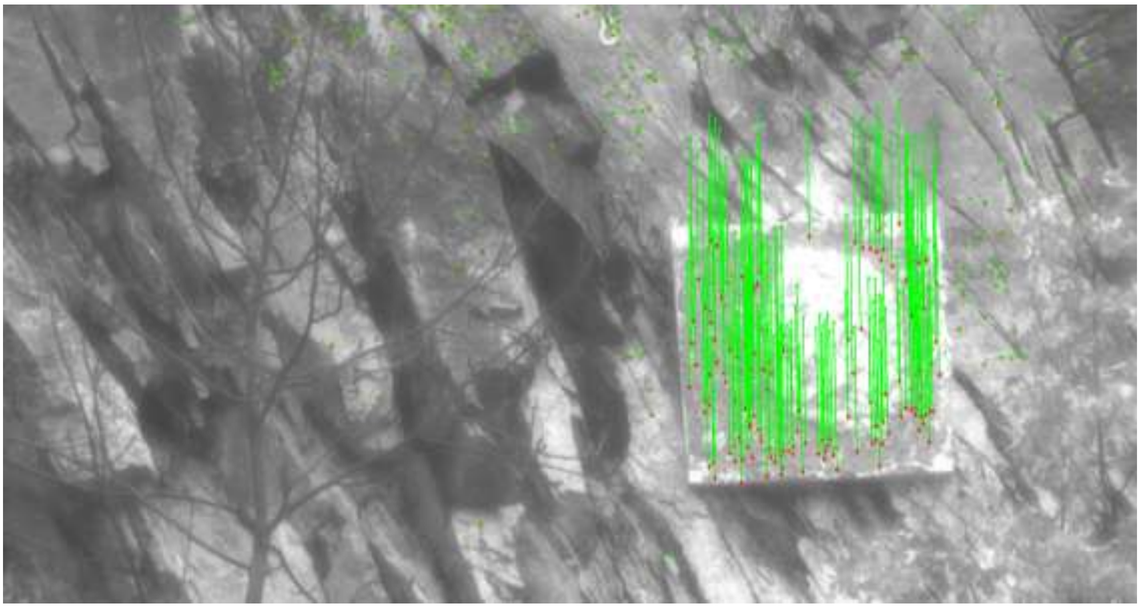

Figure 65: 3D vectors from FAST/SIFT matching projected into left image of first epoch

The accuracy of this processing run is in agreement with processing results from other feature detectors and descriptors as will be shown in the following figures. The consistency of the calculated IATS deformation vectors and the fact that the TPS measurements were carried out as polar measurements which are not capable to sufficiently capture the vertical direction of deformation leads to the conclusion that the calculated IATS measurements may be more accurate than the acquired “reference” data. For this reason another field trial had to be carried out to capture reference data in higher accuracy.

Figure 65 shows the resulting deformation vectors from a processing run with the detector, descriptor combination FAST/SIFT. A higher amount of deformation vectors could be detected but the matching accuracy seems to be below the FAST, BRISK combination in this case. This results in a higher variance of the length of deformation vectors as visible in Figure 66. The mean deformation vector length however stays the same as for the FAST, BRISK based processing.

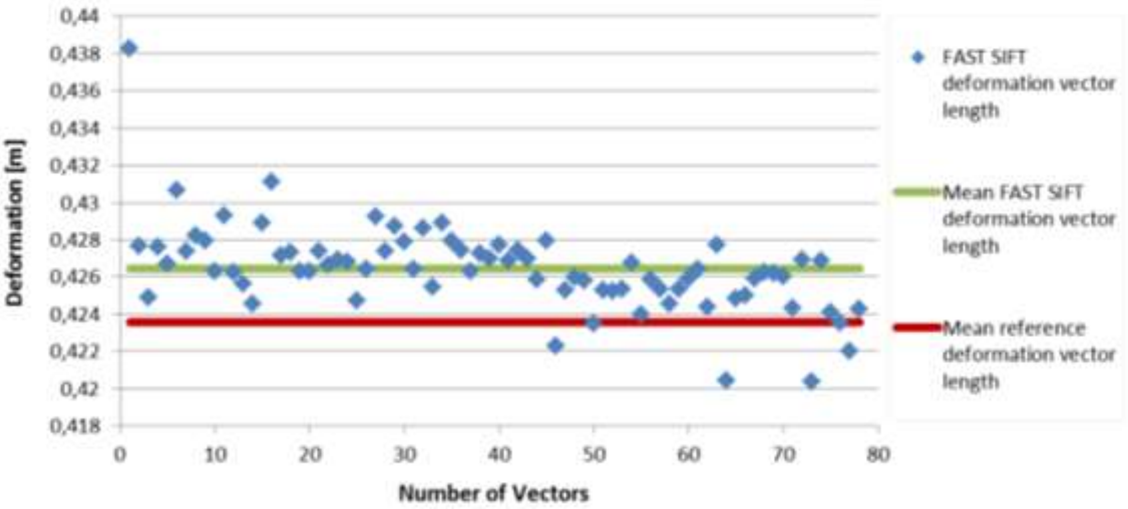

Figure 66: FAST SIFT deformation vector length compared to mean length from TPS

As expected the execution time of the SIFT based processing chain is significantly higher (Table 12).

Table 12: Processing time using FAST and SIFT

Processing Step	**Execution Time [seconds]**
Pointmatching L1R1	2.92
Pointmatching L1R1 (with LSM)	15.82
Format conversion and forward intersection	0.26
Pointmatching L1L2	2.20
Pointmatching L1L2 (with LSM)	5.91
RSOC filter (python implementation)	2.49
RSOC filter (estimation for C++ implementation)	*0.25*
Pointmatching L2R2	2.49
Pointmatching L2R2 (with LSM)	3.77
Format conversion and forward intersection	0.20
Output generation and visualization (python implementation)	2.64
Output generation and visualization (estimation for C++ implementation)	*0.26*
Total (python)	**13.66**
Total with LSM refinement (python)	**31.44**
Estimated total (C++)	**9.66**
Estimated total with LSM refinement (C++)	**27.63**

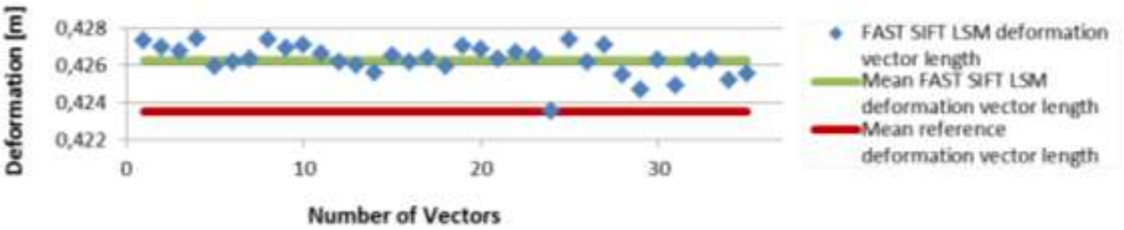

Figure 67: FAST SIFT LSM deformation vector length compared to mean length from TPS

Additional LSM match refinement reduces the amount of detected deformation vectors as visible in Figure 67. The increased matching accuracy reduces the variance in the length of the resulting deformation vectors. However, it increases the already long processing times even further as visible in Table 12. The timing values were again retrieved as an average of 10 processing runs with a standard deviation of 0.62 and 0.87 seconds with and without LSM.

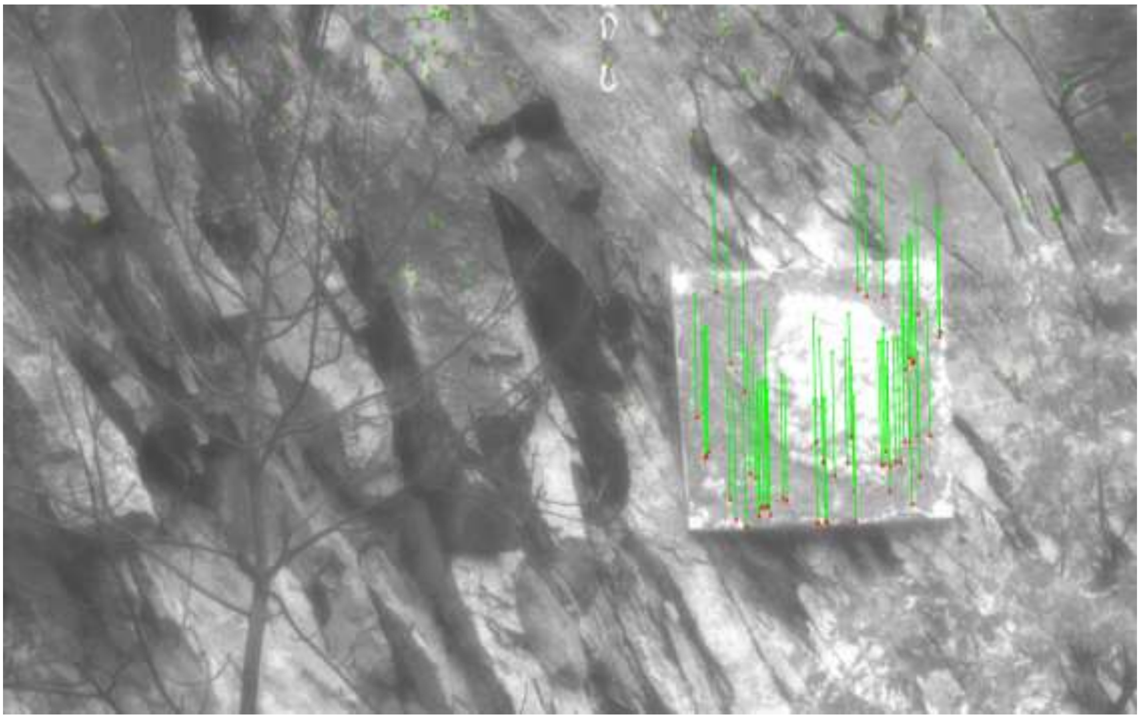

Figure 68: 3D vectors from HARRIS/SIFT matching projected into left image of first epoch

As found in the evaluation of synthetic test data in Section 7.2 the detector, descriptor combination HARRIS/SIFT (Figure 68) leads to the best results in terms of matching accuracy and repeatability. The results in this evaluation lead to less deformation vectors that could be tracked than the FAST/SIFT combination. However, the standard deviation of the matching accuracy could be lowered to a value of ±1.2 mm as can be seen in Figure 69. An additional application of LSM refinement further decreases this value to only ±0.7 mm (Figure 70).

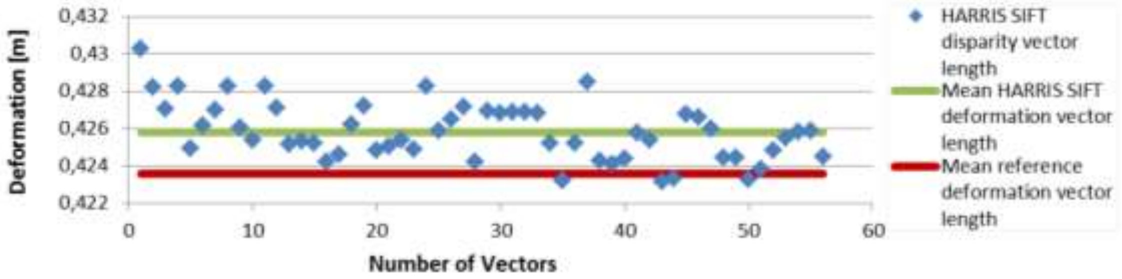

Figure 69: HARRIS SIFT deformation vector length compared to mean length from TPS

The runtimes for the HARRIS/SIFT combination with and without LSM refinement shown in Table 13 lie above the values for FAST/SIFT due to the longer processing times needed for the detection of HARRIS features. The values are again averaged over 10 processing runs and show standard deviations of around 0.71 and 0.83 seconds for normal and LSM refined matches.

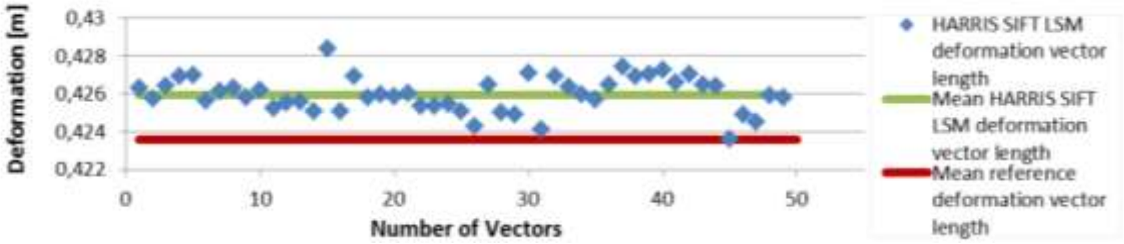

Figure 70: HARRIS SIFT LSM deformation vector length compared to mean length from TPS

The different matching methods lead to similar results in terms of accuracy as illustrated in Table 14. The main difference is the processing time and the amount of detected deformation vectors.

Table 13: Processing time using HARRIS and SIFT

Processing Step	Execution Time [seconds]
Pointmatching L1R1	4.83
Pointmatching L1R1 (with LSM)	13.59
Format conversion and forward intersection	0.28
Pointmatching L1L2	4.52
Pointmatching L1L2 (with LSM)	8.44
RSOC filter (python implementation)	6.51
RSOC filter (estimation for C++ implementation)	*0.65*
Pointmatching L2R2	3.44
Pointmatching L2R2 (with LSM)	4.78
Format conversion and forward intersection	0.22
Output generation and visualization (python implementation)	4.31
Output generation and visualization (estimation for C++ implementation)	*0.43*
Total (python)	**24.11**
Total with LSM refinement (python)	**31.39**
Estimated total (C++)	**15.41**
Estimated total with LSM refinement (C++)	**29.19**

Table 14 shows a systematic offset between TPS and IATS measurements that is also visible in Figure 64, 66, 67, 69 and 70. This was most likely caused by an outdated calibration of intrinsic camera parameters of the used IATS devices at time of the experiment. The faulty values for the principle point and focal length cause deviations from the TPS based measurements in the range of around 2.5 mm.

Table 14: Summary of calculated deformation vector deviation from TPS measurements

Detector	Descriptor	Refine	Mean Displacement [m]	Std. Dev. Displacement [m]	Deviation from TPS [m]
FAST	BRISK	No	0.4260	0.0025	0.0024
FAST	SIFT	No	0.4264	0.0016	0.0029
FAST	SIFT	LSM	0.4263	0.0006	0.0027
HARRIS	SIFT	No	0.4258	0.0012	0.0023
HARRIS	SIFT	LSM	0.4260	0.0007	0.0024

8.4.2 Field trial Pellheim evaluation

In the Pellheim experiment TPS and TLS based reference data was captured and compared to IATS measurements. This section is split into the comparison of IATS based results with TPS measurements, a comparison with TLS results and the illustration of the evaluation of a time series consisting of four measurement epochs.

8.4.2.1 IATS compared to TPS reference

In contrast to the measurement campaign in Graz the artificial deformation applied in the field trial in Pellheim was observed from a stereo set-up of total stations as shown in Figure 61. The network adjustment resulted in an accuracy of ±3 mm in height and position for the signalized corner points on the moved artificial rock (Table 15). In the following a comparison of the results of the developed processing chain applied to two consecutive measurement epochs with the mentioned reference measurements will be shown.

Table 15: Coordinate differences between epochs

Target	Δ X [m]	Δ Y [m]	Δ Z [m]	Displacement [m]
Kx.1	-0.031	-0.055	-0.001	0.063
Kx.2	-0.035	-0.054	0.005	0.065
Kx.3	-0.020	-0.057	-0.002	0.060
Kx.4	-0.023	-0.054	0.003	0.059
Ø	**-0.027**	**-0.055**	**0.001**	**0.062**

Figure 71 shows the resulting deformation vectors of the FAST/BRISK based processing. The increased distance compared to the previous example is visible in the size of the artificial stone in the image. Due to the weakly textured homogenous clay wall fewer features could be detected and tracked. The smaller deformation of only around 6.2 cm is illustrated as projected 3D deformation vectors into the left image of the first epoch. The starting points of the vectors are shown in red and the end points in green.

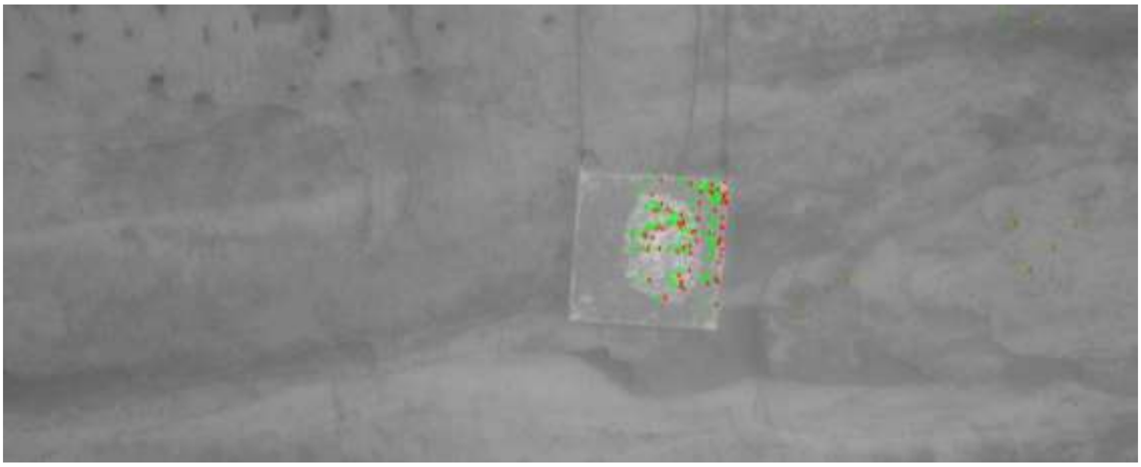

Figure 71: 3D vectors from FAST/BRISK matching projected into left image of first epoch

Since the reference measurements are to be expected far more accurate than the polar measurements in the campaign in Graz the comparison of IATS based results and the mean deformation vector length of the reference shows deviations in the range of less than 1 mm for the detector, descriptor pair FAST/BRISK (Figure 72). The standard deviation of the retrieved deformation vector length is ±3.3 mm. The used IATS stereopairs where taken from a distance of 261 m and 292 m respectively. The left images cover a field of view of around 6.4 x 4.8 m on the object with a pixel resolution of around 2.5 mm. The right images show an area of around 7.2 x 5.4 m with a resolution of around 2.8 mm.

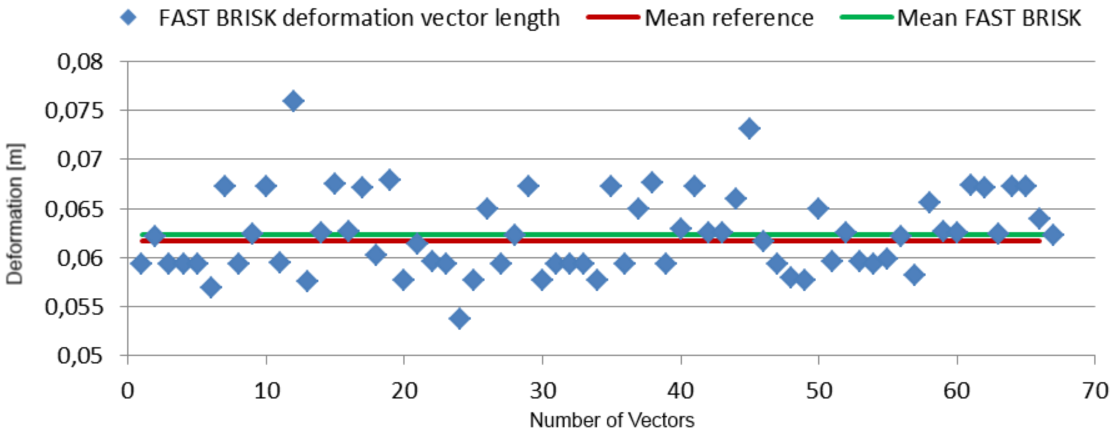

Figure 72: FAST BRISK deformation vector length compared to mean length from reference

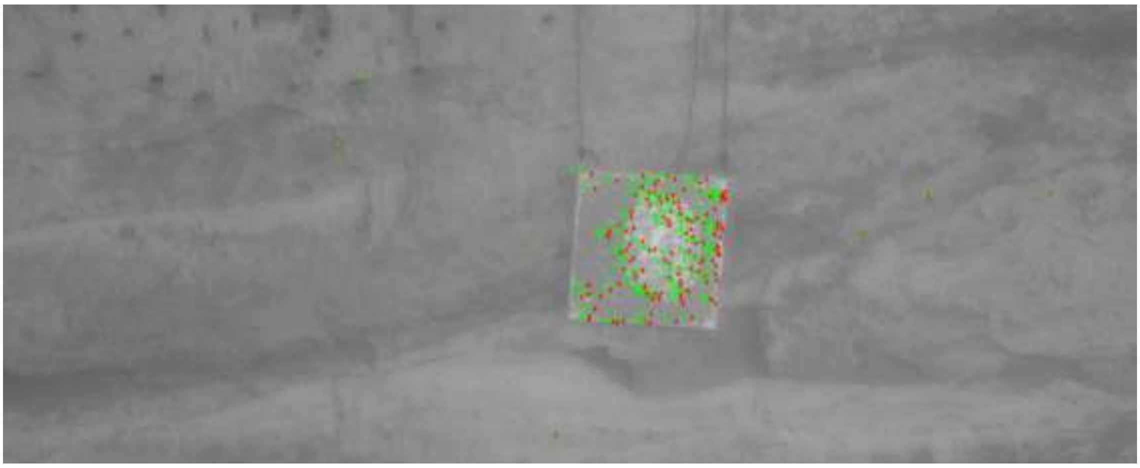

Figure 73: 3D vectors from FAST/SIFT matching projected into left image of first epoch

The resulting deformation vectors using the same detector with the SIFT descriptor lead to higher amounts of matches as shown in Figure 73. The evaluation however also results in a higher variance in the length of the vectors (Figure 74). The mean deviation from the reference is still below 2 mm. The standard deviation is almost twice as high as in the FAST/BRISK case and reaches 6.2 mm. As already seen in the evaluation of the data from Graz the additional application of LSM to refine the sub-pixel matching accuracy leads to a strong reduction of the variance of results (Figure 75). In this case a mean deviation to the reference of 1.2 mm with a standard deviation of only ±2 mm could be reached.

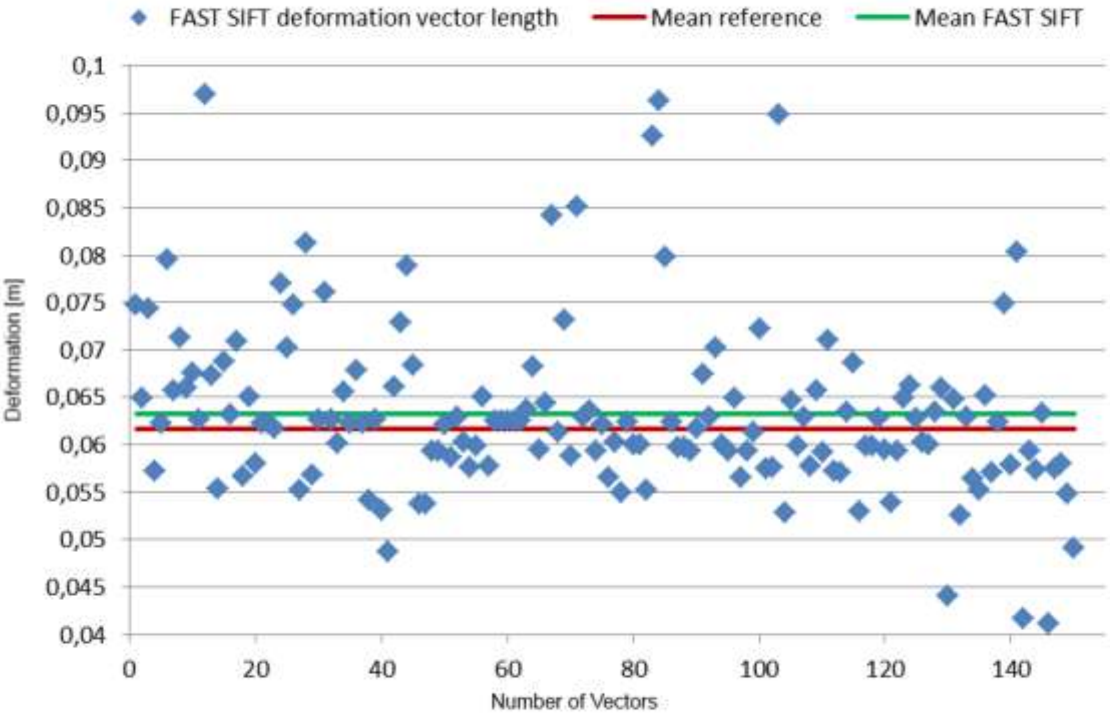

Figure 74: FAST SIFT deformation vector length compared to mean length from reference

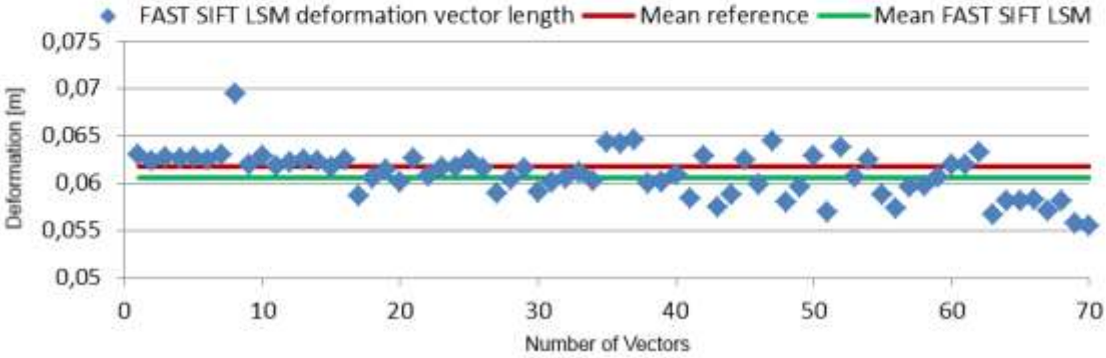

Figure 75: FAST SIFT LSM deformation vector length compared to mean length from reference

Due to the results from the synthetic test data evaluation also the detector descriptor pair HARRIS/SIFT was applied to the Pellheim artificial deformation. Figure 76 shows the according deformation vectors.

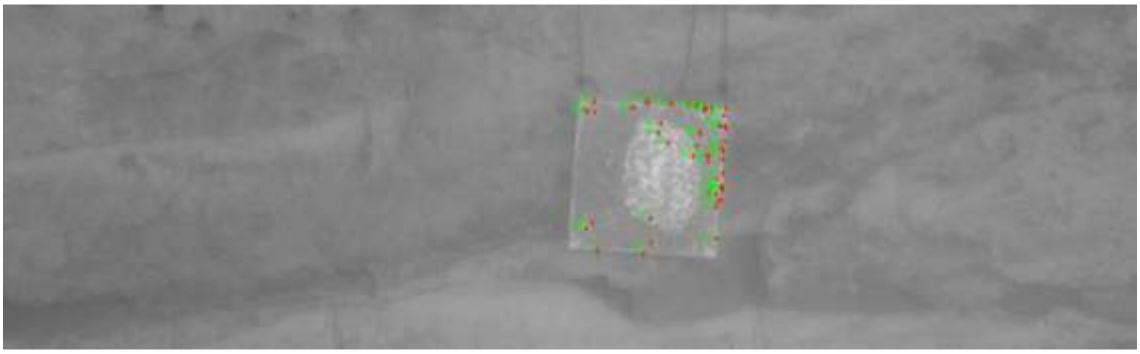

Figure 76: 3D vectors from HARRIS/SIFT matching projected into left image of first epoch

The application of HARRIS/SIFT leads to a small standard deviation of deformation vector lengths of ±2.8 mm with a low deviation of the mean processing derived vectors from the mean reference vectors of only 0.4 mm (Figure 77). An additional match refinement using LSM leads to an even lower standard deviation of ±1.6 mm as illustrated in Figure 78.

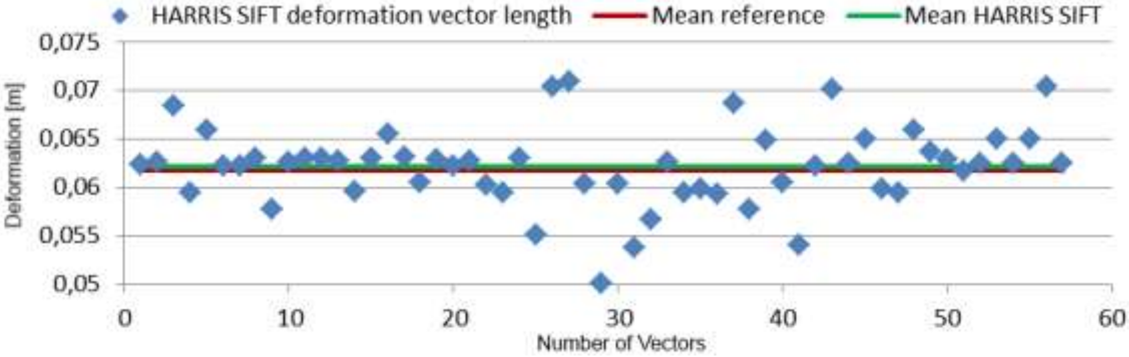

Figure 77: HARRIS SIFT deformation vector length compared to mean length from reference

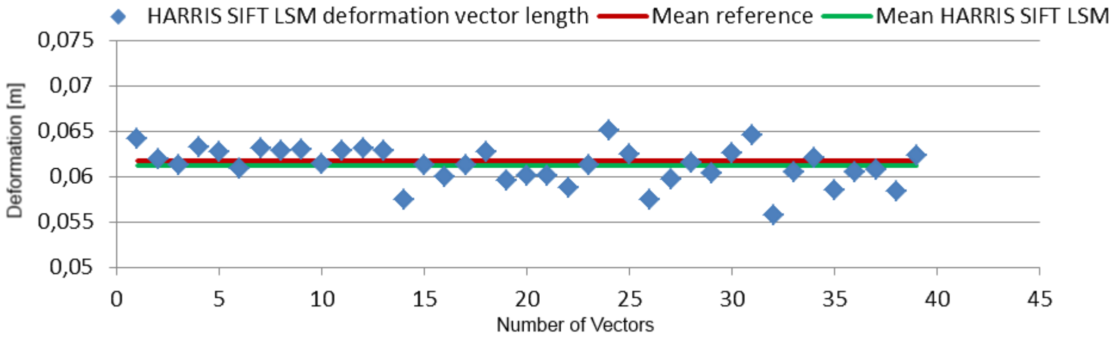

Figure 78: HARRIS SIFT LSM deformation vectors compared to mean length from reference

Table 16 shows a summary of the evaluation. The IATS based results coincide with reference measurements well within the calculated accuracy of the reference data. As was to be expected the application of LSM leads to a reduction of the variance of the retrieved results, however, processing solely based on feature descriptor matching also leads to similarly low average deviations from the reference.

Table 16: Summary of calculated deformation vector deviation from TPS measurements

Detector	Descriptor	Refined	Mean Displacement [m]	Std. Dev. Displacement [m]	Deviation from TPS [m]
FAST	BRISK	No	0.0623	0.0033	0.0006
FAST	SIFT	No	0.0633	0.0062	0.0016
FAST	SIFT	LSM	0.0605	0.0020	-0.0012
HARRIS	SIFT	No	0.0622	0.0028	0.0004
HARRIS	SIFT	LSM	0.0613	0.0016	-0.0004

8.4.2.2 IATS compared to TLS measurement

Based on the formulas given in Section 3.1 a range of expected errors was calculated for the IATS measurement configuration in the Pellheim experiment. The eigenvalues and eigenvectors of the covariance matrix lead to an error ellipse with semi-axes of 14.4 mm and 2.1 mm at an orientation of 159 gon leading to a Helmert point error of 14.6 mm. In the direction of the baseline which is also the approximate direction of the ROI this leads to a mean standard deviation of 13.6 mm in longitudinal and 5.6 mm in transversal direction (Wagner, et al., 2014). To compare the TLS scan data with the IATS a triangulated irregular network (TIN) was created from the raw scan data. The IATS the image bundles where stitched together to two spherical panoramas based on their known orientation and a sub-pixel matching process at the tile borders to ensure a smooth transition. These stereo panoramas were mutually matched using the correlation based dense matching algorithm HFVM (Kolesnik, et al., 1998) and reconstructed in 3D based on the known orientation of the panoramas. The resulting point cloud was also triangulated and can therefore be compared to the TLS results. The shortest distance of each point of the two datasets is illustrated in Figure 85.

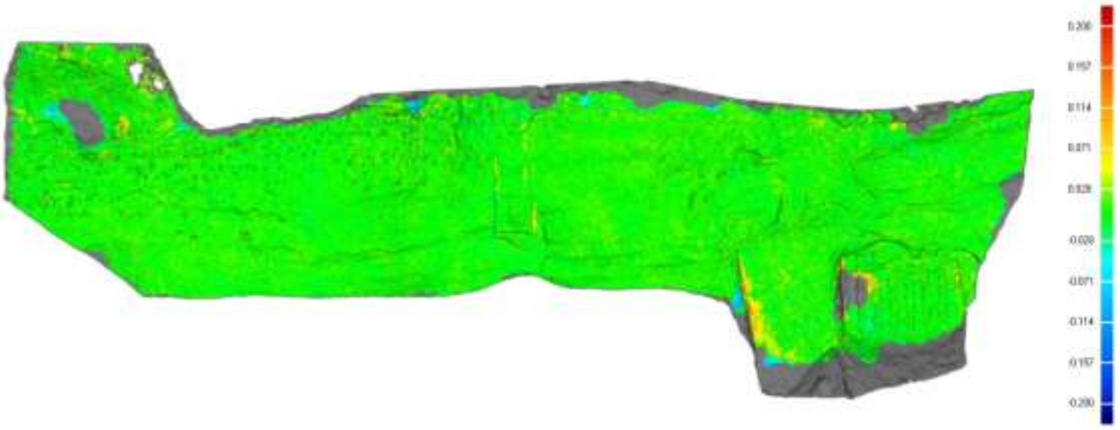

Figure 79: IATS to TLS results comparison, differences in m (Wagner, et al., 2014)

The TLS model is illustrated as background whereas the distance to the IATS model is color-coded from negative values in blue to positive values in red. The range illustrated in green is chosen based on the calculated Helmert point error of 14.6 mm and a hypothesis test with a 95% level of significance. This evaluation shows a good accordance of the IATS point cloud with the TLS data within the calculated and expected range of accuracy (Wagner, et al., 2014).

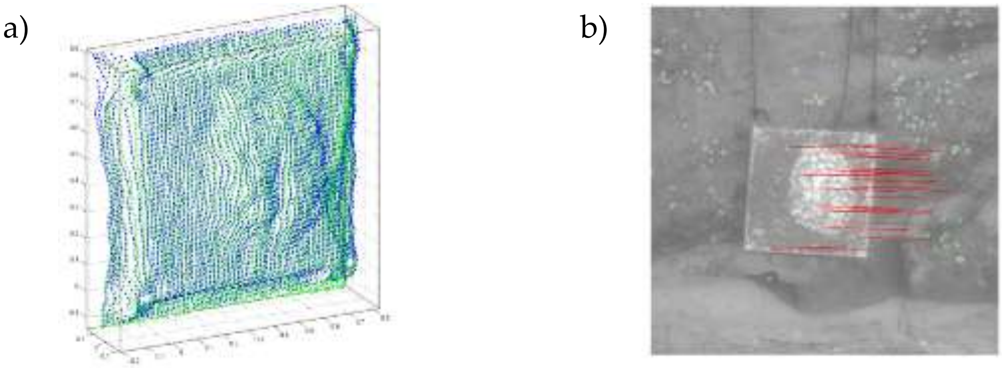

Figure 80: a) ICP registration of TLS scans b) IATS based deformation vectors exaggerated by a factor of 10 (Wagner, et al., 2014)

The artificial deformation was also captured by the TLS. Using an ICP algorithm cut-outs of the target region containing the artificial deformation where registered to each other to retrieve transformation parameters representing the rigid body motion of the artificial rock. The developed IATS based processing chain using the feature detector FAST and the descriptor SIFT was applied to retrieve 3D deformation vectors in the same way as described in the previous Sections. The parameters of the transformation could be directly calculated from the 3D deformation vectors. Results of both TLS and IATS were compared to transformation values based on the TPS reference measurements.

Table 17: TLS and IATS rigid body motion residuals to reference from TPS

	TLS	IATS
Translation [m]		
X	0.005	-0.005
Y	-0.003	0.007
Z	0.005	0.000
Rotation [gon]		
X	-0.085	0.173
Y	-0.088	0.236
Z	-0.068	0.494

Table 17 shows the residuals of the TLS and IATS based transformation parameters to the TPS based reference. Within their measurement accuracy both, the TLS and IATS lead to similar results, however with the TLS at a distance of ~80 meters and the IATS at 261 m and 292 m from the object.

It remains to be mentioned that the evaluated case of moving a well-defined artificial rock is especially beneficial for the ICP algorithm used to register the two laser scan point clouds. The clearly separated and temporally unchanged structure of the artificial rock facilitates the registration in all three directions. In the theoretical case where only the horizontal edge of such a target is measured, a lateral deformation might not be quantifiable by the laser scanner. This aperture problem is however mitigated by the high point density of the scanner and will only become problematic at further distances. Admittedly the textural information needed by the IATS system is generally accompanied by structural variations in natural objects such as rock slopes, which means that both systems will show a similar capacity to register multi temporal datasets.

8.4.2.3 IATS time series

In the Pellheim experiment five measurement epochs were captured from the same stations wherein only the artificial target was moved and finally removed in the last epoch. The processing chain described in Section 8.3 was applied to retrieve deformation vectors between the individual epochs.

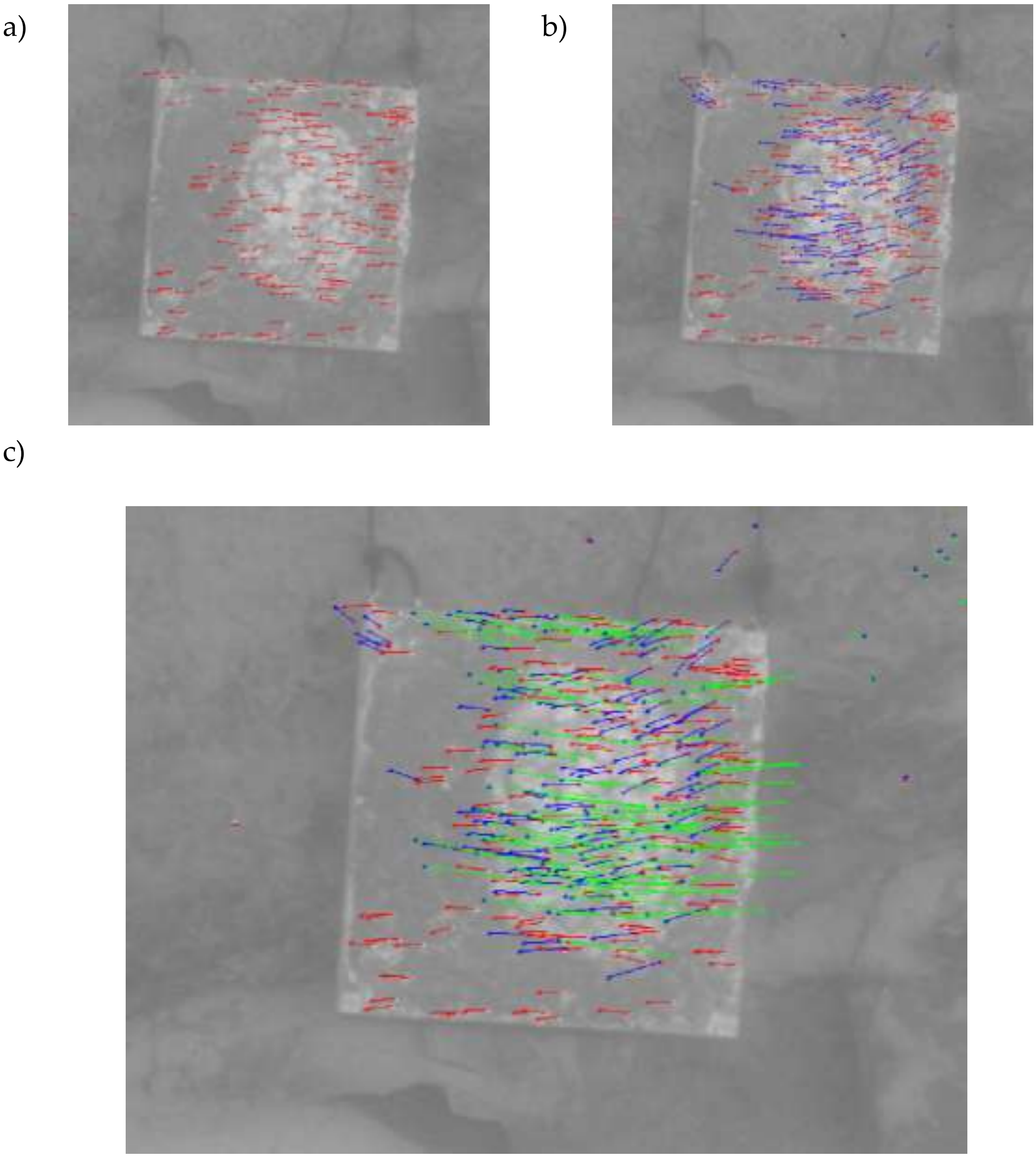

Figure 81: Illustration of time series, deformation vectors are projected in the left image of the first epoch a) deformation vectors between epoch 0 and 1 in red, start points gray b) added vectors between epoch 1 and 2 in blue, start points red c) added vectors between epoch 2 and 3 in green, start points blue

The processing was based on the detector descriptor pair FAST/SIFT. The horizontal movement between epoch 0 and 1 can be seen as red deformation vectors in Figure 81a. The endpoints of these vectors are illustrated in the form of small red crosses. Figure 81b shows the consecutive deformation consisting of a rotation, illustrated in blue. The beginnings of the deformation vectors are marked in red whereas the endpoints have the same color as the lines. Figure 81c shows a slight rotation and larger horizontal displacement in the opposite direction as the previous movements. The green deformation vectors start at positions marked with a blue cross.

Most deformation vectors only connect two measurement epochs. A continuous tracking of the occurred deformation over all epochs is however possible by interpolation due to the sufficiently high density of deformation vectors.

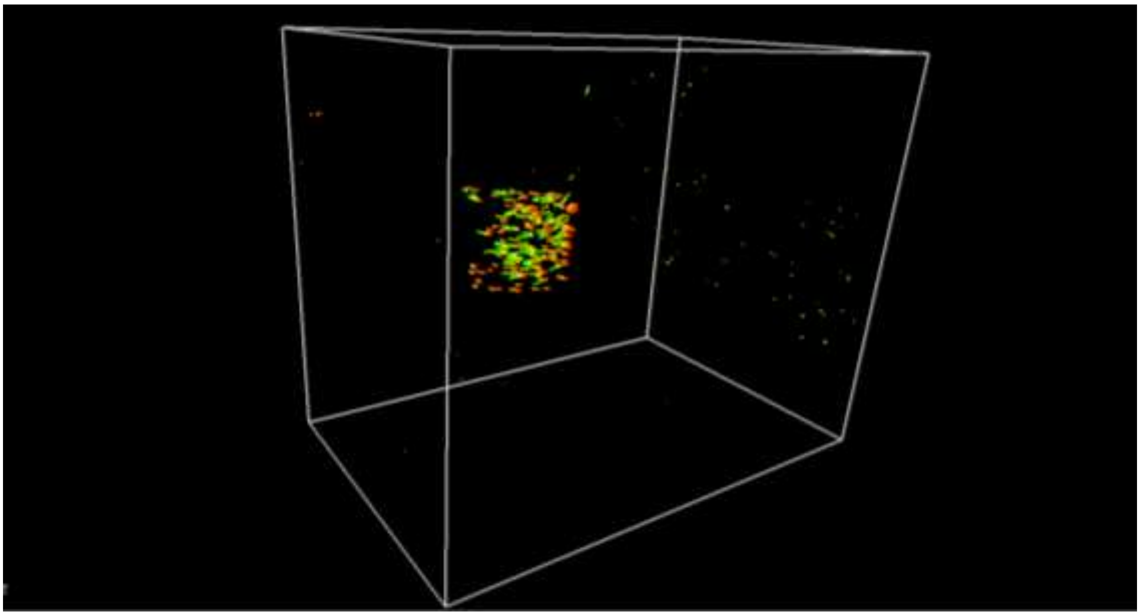

Figure 82: 3D VRML model of deformation vectors between epoch 0, 1 and 2

The deformation vectors can also be directly analyzed in 3D using the VRML format and a suitable 3D viewer. Figure 82 shows the deformation vectors between epoch 0, 1 and 2 in 3D. The vectors are displayed via a gradual color change from red to yellow between epoch 0 and 1, and a change from yellow to green between epoch 1 and 2. This method allows a simple yet efficient analysis of the calculated results directly in 3D.

8.5 Conclusion and interpretation

The evaluation of field trial measurement data using the developed processing chain allowed an end to end testing of the system design from data acquisition to the final processing. The results were assessed and validated using ground truth theodolite measurements. The full used processing chain was described and executed with different detector, descriptor combinations. Furthermore, processing runs with and without LSM sub-pixel refinement were executed. To reduce possible mismatches to a minimum a global geometrical epipolar outlier filtering was applied in between stereo measurements and a local RSOC filter was used for different temporal measurement epochs preserving the matched deformation which would violate a global outlier detection model.

The timing of the fastest available detector/descriptor combination without LSM refinement shows values around 11 seconds for a full processing and multi-temporal connection of two stereopairs from different measurement epochs resulting in 3D deformation vectors. Considering the fact that some components are not implemented as efficiently as they could be and the used processing hardware is not up to date, an evaluation within 5 to 6 seconds required for a fully continuous monitoring is reachable. The mean accuracy of this fast processing run was compared with other detector, descriptor combinations including LSM refinement and was found to be in good agreement. However, the refinement process reduces the standard deviation of measurements.

Because of unfavorable weather conditions in the field trial in Graz Weinzödl the reference measurements were carried out only as polar distance measurements. Caused by IATS calibration issues a systematic deviation between IATS and TPS measurements occurred in the range of around 2.5 mm. However, measurement errors could not sufficiently be evaluated due to an insufficient coverage with accurately measured ground control points. For these reasons another field trial had to be carried out to capture reference data in higher accuracy.

The reference data of the second field trial in Pellheim, Germany, was captured from two TPS positions. A network adjustment lead to an accuracy of ±3 mm for the signalized corner points on the moved artificial rock. The mean length of the resulting automatically derived 3D deformation vectors showed deviations of only between 0.4 mm and 1.2 mm from the reference measurements at an IATS distance of 261 m and 292 m from the object. LSM refined matches lead to a standard deviation of better than ±2.0 mm whereas unrefined results where only below ±6.5 mm varying for different detector, descriptor pairs.

The comparison of IATS with TLS illustrate that the IATS based system is able to compete with TLS, in terms of accuracy. When it comes to measurement set-up and data acquisition time the IATS system is more complex and time consuming. If dense point clouds are calculated from IATS data longer processing times in the range of multiple hours, depending on the size of the image bundle and the set resolution of the panorama, for the mosaicking and dense matching process have to be accepted. In this case also a parallel data acquisition and processing is not possible as the full panoramas have to be available before they can be evaluated. These points are the main reasons the processing chain developed in this thesis is focused on continuous pairwise evaluations of stereo-images and their connection to the temporally adjacent measurement epochs and not the calculation of dense 3D point clouds. This way data acquisition and evaluation can be executed in parallel in about 5-6 seconds per image (on suitable processing hardware) automatically leading to already temporally connected 3D deformation vectors. The whole method however, is limited by the need of illumination and sufficient texture on the measured object.

9 Conclusions

The main objective of this thesis was to develop a photogrammetric processing chain for oriented image bundles, acquired with a stereo set-up of "Image Assisted Total Stations". The work was targeted at the task of continuous rock slope deformation assessment from measurements taken in consecutive intervals, so called measurement epochs, in order to determine and monitor objects which may be at risk of undergoing deformation.

Therefore a processing chain for the automated calculation of 3D deformation vectors based on IATS image data was developed and evaluated. After appropriate system calibration and set-up the accuracies lie in a similar range as manually executed point measurements on signalized targets. However, these results can be retrieved automatically, solely based on image texture and the detection of dozens or even hundreds of deformation vectors is only a matter of seconds.

This aspect represents the main advantage of an IATS based measurement system in combination with appropriate image analysis: the resulting 3D point density can reach similar values as a laser scan combined with the accuracy of conventional theodolites. In comparison to TLS the IATS based angle measurements, like conventional theodolite measurements do not have the issue of large laser beam footprints. On the other hand small patches of image texture around points are still necessary to automatically calculate homologues points needed to retrieve the 3D coordinates. Further statistical error information such as error ellipses for each calculated 3D point can be derived. This requires the calculation of the forward intersection as a constrained adjustment as described in Section 6.4 and was not performed in the conducted field experiments. However, for the Pellheim experiment the accuracy range to be expected was calculated based on the measurement configuration (Section 8.4.2.2) leading to an error ellipse with semi-axes of 14.4 mm and 2.1 mm and a Helmert point error of 14.6 mm for a set-up with a baseline of around 84 m and a mean distance to the object of 261 m and 292 m for the respective stations.

For IATS direct point to point connections over multiple measurement epochs based on image texture are possible, directly leading to 3D deformation vectors. Laser scans can only be registered using the structure of the object without the possibility to retrieve direct point to point correspondences without signalized targets. Therefore the method is especially well suited to detect deformations in lateral direction which may not be recognizable in laser scans in case of insufficient structural variation in the scanned region. In the Pellheim experiment a well-defined target was moved to create an artificial deformation.

In this case it is unproblematic for an automated algorithm to register laser scans. The high point density of the scans generally mitigates the aperture problem. In contrast, the necessary aperture to calculate homologous points in images only consists of small patches of texture around the points, depending on the scale of the point and the used algorithm and is therefore far smaller than the larger patch needed for a laser scan registration.

Using suitable instrument control software a continuous data acquisition and parallel evaluation using the developed processing chain is possible leading to an automated deformation monitoring system, thereby fulfilling the main objective of this thesis.

In order to reach these results many distinctive underlying processing steps are necessary. This thesis focused on developing, evaluating and combining these components to a full workflow. In order to automatically retrieve direct 3D deformation vectors the concept of feature detection and matching was applied. Various state of the art methods were developed and evaluated based on synthetic reference data. As mismatches can lead to wrong deformation alerts they cannot be tolerated in the evaluation. For that reason they were reduced by a set of specialized filters based on local or global geometric constraints. For further refinement of the acquired feature correspondences the LSM method was successfully implemented and evaluated.

The in depth evaluation of all methods lead to a set of algorithms which can be processed in the same amount of time necessary to capture a new stereopair of images. This allows a possible fully continuous monitoring. The second set of algorithms offers a higher accuracy but is more time consuming leading to longer necessary measurement intervals.

The developed processing chain was evaluated using test data from two executed measurement campaigns and compared to captured ground truth data from TPS and TLS measurements. The first experiment (Section 8.4.1) lead to a systematic offset of IATS measurements from the TPS reference in the range of 2.5 mm, due to IATS calibration issues. The evaluation of the second experiment (Section 8.4.2) showed a good accordance of IATS and TPS as well as of IATS and TLS measurements.

Based on these results it can be stated that the IATS system combines the strength of a theodolite based measurement system in terms of highest accuracy and the high point density and automated processing of a terrestrial laser scanner. Furthermore, it is possible to calculate a statistical evaluation of each measured point. Due to the fact that errors in an IATS system, set up with a shorter baseline to increase matching accuracy, mainly occur perpendicular to the base the system also shows a high complementarity with TLS or polar distance measurements.

Leica Geosystems equipped their new total station MS50 with a five megapixel camera coaxial to the optical path of the telescope opening up the described field of IATS based measurements for commercial operators and other manufacturers are likely to follow. The MS50 also includes a limited laser scanning functionality therefore perfectly integrating the described complementarity of passive and active vision based sensors.

Another exciting field of future application of this technology is based on cheaper, less accurate angle measurement devices: so called pan-tilt units. In combination with more advanced processing algorithms including for example bundle adjustment methods may pave the way to a wide spread distribution and use of this class of devices. This indicates an overall trend from laser scanners to image-based sensors for many applications.

The possibilities arising from industrial versions of IATS may have a significant impact on the field of surveying leading to a fusion of ideas and methods from conventional geodesy and photogrammetry.

10 References

Abellán, A., Oppikofer, T., Jaboyedoff, M., Rosser, N. J., Lim, M. and Lato, M. J., 2014. Terrestrial laser scanning of rock slope instabilities.
Earth Surf. Process. Landforms, 39: 80–97.

Abellán, A., Vilaplana, J., Calvet, J., Blanchard, J., 2010. Rock face monitoring by Terrestrial Laser Scanning - change detection and spatial prediction of rockfall. *Geomorphology*, 119: 162-171.

Agrawal, M., Konolige, K., Blas, M. R., 2008. Censure: Center surround extremas for realtime feature detection and matching.
Computer Vision – ECCV 2008, Lecture Notes in Computer Science, 5305: 102-115.

Alahi, A., Ortiz, R., Vandergheynst, P., 2012. Freak: Fast retina keypoint.
IEEE Conference on Computer Vision and Pattern Recognition, New York.

Albertz, J., 2001. Albrecht Meydenbauer – Pioneer of photogrammetric documentation of the cultural heritage.
Proceedings 18th International Symposium CIPA, Potsdam

Aryal, A., Brooks, B. A., Reid, M. E., Bawden, G. W., Pawlak, G., 2012. Displacement fields from point cloud data: application of particle imaging velocimetry to landslide geodesy. *Journal of Geophysical Research*, 117.

Bauer, A., Huber, N. B., Wagner, A., Nauschnegg, B., 2014. D6.1 Report on Accptance Test Campaigns and Results. *DE-MONTES Consortium, Graz.*

Bauer, A., Paar, G., Kaltenböck, A., 2005. Mass Movement Monitoring Using Terrestrial Laser Scanner for Rock Fall Management.
Geo-information for Disaster Management, Delft: 393-406.

Bayer, B. E., 1976. Color imaging array. *USA, Patent No. US3971065 (A).*

Bay, H., Ess, A., T., T., Van Gool, L., 2008. SURF: Speeded Up Robust Features.
Computer Vision and Image Understanding (CVIU), 110 (3): 346-359.

Besl, P. & McKay, N., 1992. A Method for Registration of 3-D Shapes.
IEEE Transactions on Pattern Analysis and Machine Intelligence, 14 (2): 239-256.

Bethmann, F. & Luhmann, T., 2011. Least-squares matching with advanced geometric transformation models.
Photogrammetrie-Fernerkundung-Geoinformation, 2: 57-69.

BIPM, 2008. International vocabulary of metrology - Basic and general concepts and associated terms (VIM). *Joint Committee for Guides in Metrology, Paris.*

Bradski, G., 2000. The OpenCV Library. *Dr. Dobb's Journal of Software Tools. [online][accessed 2014-28-09]* http://opencv.org/documentation.html

Briechle, S., 2014. Evaluierung berührungsloser 3D-Messmethoden mit der MS50. *Master's Thesis, Technische Universität München*.

Brückl, E., Brunner, F., Kraus, K., 2006. Kinematics of a deep-seated landslide derived from photogrammetric, GPS and geophysical data. *Engineering Geology*, 88: 149-159.

Brückl, E., Brunner, F. K., Lang, E., Mertl, S., Müller, M., Stary, U., 2013. The Gradenbach Observatory - monitoring deep-seated gravitational slope deformation by geodetic, hydrological, and seismological methods. *Landslides*, 6: 815-829.

Bürki, B., Guillaume, S., Sorber, P., Oesch, H., 2010. DAEDALUS: A versatile usable digital clip-on measuring system for Total Stations. *2010 International Conference on Indoor Positioning and Indoor Navigation (IPIN), Zürich*: 1-10.

Calonder, M., Lepetit, V., Strecha, C., Fua, P., 2010. BRIEF: binary robust independent elementary features. *Computer Vision – ECCV 2010, Lecture Notes in Computer Science*, 6314: 778-792.

Chmelina, K., Eichhorn, A., Haberler-Weber, M., Kahmen, H., 2006. Deformation analysis with knwoledge and fuzzy based systems. *Proceedings of the 3rd IAG Symposium on Geodesy for Geotechnical and Structural Engineering / 12th FIG Symposium on Deformation Measurements, Baden*.

Debella-Gilo, M., Kääb, A., 2011. Sub-pixel precision image matching for measuring surface displacements on mass movements using normalized cross-correlation. *Remote Sensing of Environment*, 115: 130-142.

Debella-Gilo, M., Kääb, A., 2012a. Measurement of surface displacement and deformation of mass movements using least squares matching of repeat high resolution satellite and aerial images. *Remote sensing*, 4: 43-67.

Debella-Gilo, M., Kääb, A., 2012b. Locally adaptive template sizes for matching repeat images of Earth surface mass movements. *ISPRS Journal of Photogrammetry and Remote Sensing*, 69: 10-28.

Devoli, G., Taurisano A., Eikenaes, O., Hermanns, R.,L., Oppikofer, T., Fischer, L., Bunkholt, H., 2011. Plan for skredfarekartlegging (Plan for landslide hazard mapping). *Norwegian Water Resources and Energy Directorate, Oslo*.

Duchon, C., 1979. Lanczos filtering in one and two dimensions. *Journal of Applied Meteorology*, 18: 1016-1022.

Engler, C., Hollenberg, A., Kast, H., 1897. Dinglers Polytechnisches Journal. Volume 306.

Evans, C., 2009. Notes on the OpenSURF Library. *[online][accessed 2014-28-09]* http://opensurf1.googlecode.com/files/OpenSURF.pdf

Fan, B., Wu, F., Hu, Z., 2011. Aggregating gradient distributions into intensity orders: A novel local image descriptor. *2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Providence*: 2377-2384.

Fan, B., Wu, F., Hu, Z., 2012. Rotationally invariant descriptors using intensity order pooling.
IEEE Transactions on Pattern Analysis and Machine Intelligence, 34 (10): 2031-2045.

Faulhaber GmbH & Co.KG., 2007. Bedienungsanleitung MCDC 3003/06. *[online][accessed 2014-28-09]* http://www.faulhaber.com/manuals/pdf/geraetehandbuch/mcxx300x_v2-5/DE_7000_00038.pdf

Fischler, M. A. & Bolles, R. C., 1981. Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. *Comm. Assoc. Comp. Mach.*, 624: 381–395.

Freidman, J. H., Bentley, J. L., Finkel, R. A., 1977. An algorithm for finding best matches in logarithmic expected time. *ACM Trans. Math. Softw., 3 (3)*: 209-226.

Funayama, R., Yanagihara, H., Van Gool, L., Tuytelaars, T., Bay, H., 2007. Robust interest point detector and descriptor.
USA, Patent No. US 20090238460 A1.

Furuya, K., Nauschnegg, B., Kathrin, S., 2013. D5.3 Monitoring Overlay.
DE-MONTES Consortium, Graz.

Gauglitz, S., Höllerer, T., Turk, M., 2011. Evaluation of interest point detectors and feature descriptors for visual tracking.
International journal of computer vision, 394: 335-360.

Gottwald, R., 1987. Kern E2-SE — Ein neues Instrument nicht nur für die Industrievermessung? *AVN Allgemeine Vermessungs- Nachrichten*, 4: 75.

Grün, A., 1985. Adaptive Least Squares Correlation: A Powerful Image Matching Technique. *South African Journal of Phntogrammetry, Remote Sensing and Cartography*, 314: 175-187.

Guillaume, S., Buerki, B., Griffet, S., Mainaud Durand, H., 2012. QDaedalus: Augmentation of Total Stations by CCD Sensor for Automated Contactless High-Precision Metrology. *Knowing to manage the territory, protect the environment, evaluate the cultural heritage, FIG Working Week, Rome.*

Haag, R., Bayer, G., Zimmermann, M., 1997. Vermessen mit der automatischen Feinzielung des TCA1800 von Leica.
VPK Vermessung, Photogrammetrie und Kulturtechnik, 7: 466-468.

Harris, C. & Stephens, M., 1988. A Combined Corner and Edge Detector.
In Proc. of Fourth Alvey Vision Conference: 147-151.

Hartley, R. & Zisserman, A., 2000. Multiple view geometry in computer vision. Ed.2. *Cambridge University Press, New York*.

Hauth, S. & Schlüter, M., 2010. Digitalkameratachymeter-Einsatzmöglichkeiten für modulare und integrierte Systeme. *30. Wissenschaftlich-Technische Jahrestagung der DGPF 2010, 3-Ländertagung, Publikationen der DGPF Band 19, Vienna*.

Heid, T. & Kääb, A., 2012. Evaluation of existing image matching methods for deriving glacier surface displacements globally from optical satellite imagery. *Remote Sensing of Environment*, 118, pp. 339-355.

Hirschmuller, H., 2006. Stereo vision in structured environments by consistent semi-global matching. *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, New York*, 2: 2386-2393.

IDS GmbH, 2009. User Manual uEye Cameras, Driver Version 3.32. *Obersulm, [online][accessed 2014-28-09]* http://www.micropticsl.com/wp-content/uploads/2013/09/ueye_usb_manual.pdf

Jazayeri, I. & Fraser, C., 2008. Interest Operators in Close-Range Object Reconstruction. *Proceedings of the 21st ISPRS Congress. Beijing*: 69–74.

Juretzko, M., 2005. Reflektorlose Video-Tachymetrie – ein integrales Verfahren zur Erfassung geometrischer und visueller Informationen.
Deutsche Geodätische Kommission, Reihe C, 588: 1-122.

Kahmen, H. & Steudel, J., 1988. Das automatisch zielende Messsystem GEO-ROBOT II. *Ingenieurvermessung Dümmler Verlag*, Bonn.

Kälber, S., Jäger, R., Schwäble, R., 1999. GPS-basierte Online Kontroll - und Alarmsysteme (GOCA). *X. Internationale Geodätische Woche in Obergurgl/Österreich, Februar 1999. Institutsmitteilungen, Institut für Geodäsie der Universität Innsbruck*. 18: 231-238.

Kolesnik, M., Paar, G., Bauer, A., Ulm, M., 1998. Algorithmic Solution for Autonomous Vision-Based Off-Road Navigation.
Orlando, International Society for Optics and Photonics: 230-247.

Kraus, K., 1996. Photogrammetrie Band 2: Verfeinerte Methoden und Anwendungen. *3. Auflage, Walter de Gruyter*.

Kraus, K., 2007. Photogrammetry: Geometry from Images and Laser Scans, Band 1. *2 ed. Walter de Gruyter*.

Kristensen, L., Blikra, L.H., Hole, J., 2010. Åknes: State of instrumentation and data analysis. *Åknes/Tafjord Early Warning Centre*: 1-43. *[online][accessed 2014-28-09]* http://aknes.no/sites/default/files/Aaknesrapport02_2010.pdf

Ladstädter, R., 2004. Digital-photogrammetrische Bewegungsmessung von Blockgletschern - Entwurf eines Monitoringsystems.
Dissertation, Institut für Fernerkundung und Photogrammetrie, TU Graz.

Leica Geosystems AG, 2003. TPS Info Kalibrierung 01/2003. *Heerbrugg*.

Leica Geosystems AG, 2007a. Leica TPS1200+ Series Technical Data. *Heerbrugg*.

Leica Geosystems AG, 2007b. Leica TPS1200+ White Paper. *Heerbrugg*.

Leica Geosystems AG, 2010. Leica Viva TS15 Brochure. *Heerbrug*.

Leica Geosystems AG, 2013a. Leica Nova MS50 Datasheet. *Heerbrugg*.

Leica Geosystems AG, 2013b. Leica Nova MS50 Whitepaper. *Heerbrugg*.

Leutenegger, S., Chli, M., Siegwart, R., 2011. BRISK: Binary robust invariant scalable keypoints. *2011 IEEE International Conference on Computer Vision (ICCV), Barcelona:* 2548–2555.

Liu, Z., An, J., Jing, Y., 2012. A Simple and robust feature point matching algorithm based on restricted spatial order constraints for aerial image registration. *IEEE Transactions on Geoscience and Remote Sensing*, 50 (2): 514–527.

Lowe, D., 2004a. Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision*, 60: 91-110.

Lowe, D., 2004b. Method and apparatus for identifying scale invariant features in an image and use of same for locating an object in an image.
USA, Patent No. US 6711293 B1.

Marr, D. & Hildreth, E., 1980. Theory of Edge Detection.
Proceedings of the Royal Society of London, 207: 187-217.

Matas, J., Chum, O., Urban, M., Pajdla, T., 2002. Robust wide-baseline stereo from maximally stable extremal regions.
Proceedings of the British Machine Vision Conference, 22 (7): 384–393.

Matsuoka, R., Sone, M., Sudo, N., Yokotsuka, H., 2008. Comparison of Image Interpolation Methods Applied to Least Squares Matching. *International Conference on Computational Intelligence for Modelling Control & Automation, Vienna, IEEE:* 1017–1022.

Mikolajczyk, K. & Schmid, C., 2001. Indexing Based on Scale Invariant Interest Points. *Eighth IEEE International Conference on Computer Vision, ICCV*: 525-531.

Mikolajczyk, K. & Schmid, C., 2002. An affine invariant interest point detector. *Computer Vision — ECCV 2002, Lecture Notes in Computer Science*, 2350: 128-142.

Mikolajczyk, K. & Schmid, C., 2004. Scale & affine invariant interest point detectors. *International Journal of Computer Vision*, 60: 63-86.

Mikolajczyk, K. & Schmid, C., 2005. A Performance Evaluation of Local Descriptors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27 (10): 1615-1630.

Miksik, O. & Mikolajczyk, K., 2012. Evaluation of local detectors and descriptors for fast feature matching.
21st International Conference on Pattern Recognition (ICPR): 2681–2684.

Mischke, A. & Kahmen, H., 1997. A New Kind of Measurement Robot System for Surveying of non Signalized Targets.
Optical 3D measurement techniques Karlsruhe, 4: 311-318.

Muja, M. & Lowe, D. G., 2009. Fast Approximate Nearest Neighbors with Automatic Algorithm Configuration. *VISAPP International Conference on Computer Vision Theory and Applications*: 331–340.

Nauschnegg, B., Sander, K., Huber, N. B., 2013. D5.3 Monitoring Overlay. *DE-MONTES Consortium, Graz.*

Niemeier, W., 2002. Ausgleichungsrechnung: eine Einführung für Studierende und Praktiker des Vermessungs- und Geoinformationswesens.
Walter de Gruyter, Berlin.

Niemeier, W., 2008. Ausgleichungsrechnung: statistische Auswertemethoden. *Walter de Gruyter, Berlin.*

Pentax, 2009. R-400VDN Visio Datasheet. *[online][accessed 2014-28-09]* http://www.pentaxsurveying.com/en/pdfs/R400VDN-FOLDER-DE.pdf

Petley, D., 2012. Global patterns of loss of life from landslides.
Geology, Geological Society of America, 1040: 927-930.

Poisel, R. & Preh, A., 2004. Rock slope initial failure mechanisms and their mechanical models. *Felsbau*, 222: 40-45.

Redpath, T., Sirguey, P., Fitzsimons, S. & Kääb, A., 2013. Accuracy assessment for mapping glacier flow velocity and detecting flow dynamics from ASTER satellite imagery: Tasman Glacier, New Zealand.
Remote Sensing of Environment, 133: 90-101.

Reiterer, A., 2004. A Knowledge-Based Decision System for an On-Line Videotheodolite-Based Multisensor System. *Dissertation, Institute of Geodesy and Geophysics, Engineering Geodesy, Vienna University of Technology*.

Reiterer, A., 2012. Modeling Atmospheric Refraction Influences by Optical Turbulences Using an Image-Assisted Total Station.
zfv - Zeitschrift für Geodäsie, Geoinformation und Landmanagement, 137 (3).

Reiterer, A., Huber, N. B., Bauer, A., 2010. Image-Based Point Detection and Matching in a Geo-Monitoring System.
Allgemeine Vermessungsnachrichten (AVN), 4.

Reiterer, A., Lehmann, M., Miljanovic, M., Ali, H., Paar, G., Egly, U., Eiter, T., Kahmen, H., 2009. A 3D Optical Deformation Measurement System Supported by Knowledge-Based & Learning Techniques. *Journal of Applied Geodesy*, 3: 1-13.

Roic, M., 1996. Erfassung von nicht signalisierten 3D-Strukturen mit Videotheodoliten, *Veröffentlichung des Inst. für Landesvermessung und Ingenieurgeodäsie, Abt. Ingenieurgeodäsie, Wien*.

Rosten, E., Drummond, T., 2005. Fusing Points and Lines for High Performance Tracking.
Tenth IEEE International Conference on Computer Vision, 2005. ICCV. 2: 1508-1511.

Rosten, E., Drummond, T., 2006. Machine Learning for High-Speed Corner Detection.
Computer Vision – ECCV 2006, Lecture Notes in Computer Science, 3951: 430-443.

Rublee, E., Rabaud, V., Konolige, K. & Bradski, G., 2011. ORB: an efficient alternative to SIFT or SURF.
2011 IEEE International Conference on Computer Vision (ICCV): 2564-2571

Scharstein, D. & Szeliski, R., 2002. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms.
International journal of computer vision, 47: 7-42.

Scherer, M., 2004. Intelligent Scanning with Robot-Tacheometer and Image Processing a Low Cost Alternative to 3D Laser Scanning?
FIG working week 2004, Athens.

Schlüter, M. & Hauth, S., Heß, H., 2009. Selbstkalibrierung motorisierter Digitalkameratheodolite für technische Präzisionsmessungen. *Zeitschrift für Geodäsie, Geoinformation und Landmanagement (zfv). Bd. Jg.* 134: 22 - 28.

Smith, S. & Brady, J., 1997. SUSAN – A New Approach to Low Level Image Processing. *International Journal of Computer Vision*, 23: 45-78.

Sonka, M., Hlavac, V., Boyle, R., 2008.
Image Processing, Analysis, and Machine Vision. 3 ed. Toronto, Thomson.

Tola, E., Lepetit, V., Fua, P., 2010. DAISY: An Efficient Dense Descriptor Applied to Wide-Baseline Stereo.
IEEE Transactions on Pattern Analysis and Machine Intelligence, 32 (5): 815-830.

Topcon, 2004. GPT-7000i Series Specifications. *[online][accessed 2014-28-09]*
http://www.topcon.hk/positioning/brochure/GPT7000i.pdf

Topcon, 2008. Imaging Station Specifications. *[online][accessed 2014-28-09]*
http://www.topconpositioning.com/sites/default/files/literature/IS_4PG_Brochure_RevA.pdf

Topcon, 2012. IS-3 Specifications. *[online][accessed 2014-28-09]*
http://www.topconpositioning.com/sites/default/files/IS-3_Broch_7010_2096_RevB_sm_0.pdf

Triggs, B., 2004. Detecting keypoints with stable position, orientation, and scale under illumination changes.
Computer Vision - ECCV 2004, Lecture Notes in Computer Science, 3024: 100-113.

Trimble, 2007. VX Spatial Station Datasheet. *[online][accessed 2014-28-09]*
http://www.trimble.com/3d-laser-scanning/vx.aspx?dtID=technical

Tuytelaars, T. & Mikolajczyk, K., 2008. Local invariant feature detectors: a survey. *Foundations and Trends in Computer Graphics and Vision*, 33: 177–280.

U.S. Geological Survey, 2013. *[online][accessed 2014-28-09]*
http://landslides.usgs.gov/aboutus/

Vicovac, T., 2008. Evaluation of the Leica IAT. *Master's Thesis, Institute of Geodesy and Geophysics, Engineering Geodesy, Vienna University of Technology*.

Vicovac, T., Reiterer, A., Egly, U., Eiter, T., Rieke-Zapp., D., 2010. Knowledge-based geo-risk assessment for an intelligent measurement system.
Artificial Intelligence in Theory and Practice III: 215-224.

Vogel, M., 2006. Vom Pixel zur Richtung – Die räumlichen Beziehungen zwischen Abbildungsstrahlen und Tachymeter-Richtungen.
Dissertation. Techn. Univ. Darmstadt, Geodätisches Inst: 1-118.

Vyner, N. A. & Hanold, J. M., 1982. A.I.M.S. Analytical Industrial Measuring Systems. *International Archives of Photogrammetry*, V/224: 524-532.

Wagner, A., Huber, N. B., Wiedmann, W., Paar, G., 2014. Long-Range Geo-Monitoring using Image Assisted Total Stations.
Journal of Applied Geodesy, Ahead of print. Online: DOI 10.1515/jag-2014-0014

Wagner, A. & Reith, C., 2013. D3.1 IATS Control.
Munich: DE-MONTES project consortium.

Wagner, A., Stylianidis, E., Smagas, K., Trdlicka, J., Paar, G., Huber, N.B., Reith Ch., 2013a. Geo-Monitoring By High-Resolution Optical Sensors.
Photogrammetrie - Laserscanning - Optische 3D-Messtechnik. Beiträge der Oldenburger 3D-Tage 2013, Wichmann H., Berlin: 166-177.

Wagner, A., Trdlicka, J., Huber, N.,B., Sander, K., 2013b. D3.5 DE-MONTES Hardware System.
Graz: DE-MONTES project consortium.

Wallis, R., 1976. An approach to the space variant restoration and enhancement of images. *Proceedings on Current Mathematical Problems in Image Science, Monterey.*

Walser, B., 2005. Development and Calibration of an Image Assisted Total Station. *Dissertation, Inst. f. Geodäsie u. Photogrammetrie, ETH Zürich.*

Wang, Z., Fan, B., Wu, F., 2011. Local intensity order pattern for feature description.
IEEE International Conference on Computer Vision (ICCV), Barcelona: 603-610.

Wasmeier, P., 2009. Grundlagen der Deformationsbestimmung mit Messdaten bildgebender Tachymeter. *Dissertation,Fakultät für Bauingenieur- und Vermessungswesen, Techn. Univ. München.*

Wasmeier, P. & Reith, C., 2013. D3.4 IATS Calibration report.
Munich: DE-MONTES project consortium.

Welsch, W. & Heunecke, O., 2001. Models and Terminology for the Analysis of Geodetic Monitoring Observations - Official Report of the Ad-Hoc Committee of FIG Working Group 6.1. *Orange, International Federation of Surveyors.*

Welsch, W., Heunecke, O. & Kuhlmann, H., 2000. Handbuch Ingenieurgeodäsie - Auswertung geodätischer Überwachungsmessungen.
1. ed. Heidelberg: Herbert Wichmann Verlag.

Wong-Riley, M., 1979. Changes in the visual system of monocularly sutured or enucleated cats demonstrable with cytochrome oxidase histochemistry.
Brain Research, 171: 11-28.

Xu, L., Jia, J., Kang, S. B., 2012. Improving sub-pixel correspondence through upsampling.
Computer Vision and Image Understanding, 116 (2): 250–261.

Zuiderveld, K., 1994. Contrast Limited Adaptive Histogram Equalization.
Graphics Gems IV, Academic Press Professional, San Diego: 474-485.

11 List of Abbreviations

Abbreviation	Meaning
IATS	Image Assisted Total Station
AIMS	Analytical Industrial Measuring System
API	Application Programming Interface
BETS	Brunson Electronic Triangulation System
BRIEF	Binary Robust Independent Elementary Features
BRISK	Binary Robust Invariant Scalable Keypoints
CCD	Charge-Coupled Device
CenSurE	Center Surround Extremas
DGPS	Differential GPS
CMOS	Complementary Metal-Oxide-Semiconductor
CPU	Central Processing Unit
DEM	Digital Elevation Model
DE-MONTES	Deformation Monitoring by High-resolution Terrestrial Long Range Sensing
DoG	Difference of Gaussians
DoH	Determinant of the Hessian
ECDS	Electronic Coordinate Determination System
FAST	Features from Accelerated Segment Test
FREAK	Fast Retina Keypoint
GB	Gigabyte
GOCA	GPS-based Online Control and Alarm System
GPS	Global Positioning System
GUI	Graphical User Interface
HD	High Definition
HFVM	Hierarchical Feature Vector Matching
HoG	Histograms of oriented Gradients
ICP	Iterative Closest Point
IDS	Imaging Development Systems
IEEE	Institute of Electrical and Electronics Engineers
i-MeaS	Intelligent Image-Based Measurement System for Geo-Hazard Monitoring
IO	Input Output

Abbreviation	Meaning
IR	Infrared
LIOP	Local Intensity Order Pattern
LSM	Least Squares Matching
MB	Megabyte
MROGH	Multisupport Region Order-Based Gradient Histogram
MRRID	Multisupport Region Rotation and Intensity Monotonic Invariant Descriptor
MSER	Maximally Stable Extremal Regions
NCC	Normalized Cross Correlation
ORB	Oriented Robust BRIEF
PC	Personal Computer
RAM	Random Access Memory
RANSAC	Random Sample Consensus
RGB	Red Green Blue
RMS	Root Mean Square
ROI	Region Of Interest
RSOC	Restricted Spatial Order Constraints
SD card	Secure Digital Memory card
SIFT	Scale Invariant Feature Transform
SURF	Speeded Up Robust Features
SUSAN	Smallest Uni-Value Segment Assimilating Nucleus
SVD	Singular Value Decomposition
TIN	Triangulated Irregular Network
TLS	Terrestrial Laser Scanning
TPS	Total Positioning System
VGA	Video Graphics Array
VRML	Virtual Reality Modelling Language
Wi-Fi	WLAN products that are based on the IEE's 802.11 standards
WLAN	Wireless local area network
XGA	Extended Graphics Array

12 Appendix

A. Evaluation of feature detection methods

Table 18: Evaluation of corner based feature detectors without image pre-processing

ID	**FAST**		**ORB**		**BRISK**		**HARRIS**		**PyrHARRIS**	
	Number of detected features	Processing time [sec]								
01	1559	0.0055	2075	0.0250	2587	0.1308	1677	0.6757	1131	0.9294
02	1896	0.0044	1636	0.0488	1692	0.0742	2090	0.7447	1258	1.0245
03	2882	0.0056	2261	0.0301	3654	0.1381	2209	0.6785	2754	0.9249
04	2657	0.0058	2387	0.0199	2791	0.1046	2406	0.6998	1241	0.9322
05	1850	0.0056	2299	0.0329	2533	0.0998	2258	0.6720	2872	0.9261
06	1277	0.0063	903	0.0182	1483	0.1066	387	0.6858	414	0.9164
07	1007	0.0051	3	0.0119	1275	0.0857	7	0.6706	12	0.9138
08	2859	0.0069	185	0.0110	2602	0.1031	2740	0.6753	1373	0.9238
09	1168	0.0051	1451	0.0527	3209	0.1465	2096	0.7029	1005	0.9816
10	1404	0.0052	1557	0.0543	3317	0.1485	756	0.6963	825	0.9654
Ø	**~1856**	**0,0056**	**~1527**	**0,0443**	**2463**	**0,1032**	**~1663**	**0,6902**	**~1289**	**0,9438**

Table 19: Evaluation of blob based feature detectors without image pre-processing

ID	**DoH**		**STAR**		**DoG**		**MSER**	
	Number of detected features	Processing time [sec]						
01	1151	0.3571	2636	0.2726	2412	1.9063	1972	0.9756
02	1224	0.3389	2026	0.2749	1055	2.4436	264	0.5658
03	1512	0.3377	1976	0.2687	2215	2.0419	2414	0.8165
04	1466	0.3617	2115	0.2672	2301	1.8760	1578	0.8652
05	2034	0.3419	2840	0.2679	2247	2.1071	3279	1.0647
06	1082	0.3484	1264	0.2706	1791	2.0739	1201	0.9263
07	1494	0.3871	2260	0.2718	847	2.3813	14	0.4226
08	1181	0.3416	2264	0.2677	2762	2.3274	288	0.5533
09	1254	0.3389	1995	0.2896	1945	1.9006	1172	1.1112
10	1354	0.3391	2206	0.2884	717	1.9159	726	0.9581
Ø	**1375,2**	**0,3492**	**2158,2**	**0,2739**	**1829,2**	**2,0974**	**1290,8**	**0,8259**

Table 20: Evaluation of corner based feature detectors with adaptive histogram equalization

ID	FAST		ORB		BRISK		HARRIS		PyrHARRIS	
	Number of detected features	Processing time [sec]								
01	1796	0.0068	1753	0.0763	3196	0.1554	785	0.6818	2315	0.9336
02	1538	0.0046	2296	0.1813	2322	0.0918	2277	0.9704	3008	1.0288
03	2707	0.0068	1967	0.1057	2497	0.1025	1704	0.6841	2549	0.9294
04	2649	0.0055	2577	0.0948	3957	0.1181	2490	0.6733	3287	0.9343
05	1904	0.0067	2131	0.1102	3082	0.1111	2038	0.6881	2904	0.9507
06	1677	0.0058	2075	0.0621	2655	0.1194	893	0.6954	2462	0.9231
07	2790	0.0074	2677	0.0970	1692	0.0763	2633	0.6821	3499	0.9348
08	1510	0.0048	2478	0.1029	3095	0.1200	2140	0.6817	3151	0.9377
09	2173	0.0063	2113	0.2120	2913	0.1190	2097	0.6933	2861	0.9568
10	2629	0.0064	2334	0.2077	3038	0.1209	1542	0.7037	3107	0.9649
Ø	**~2137**	**0.0061**	**~2240**	**0.1250**	**~2845**	**0.1135**	**~1860**	**0.7154**	**~2914**	**0.9494**

Table 21: Evaluation of blob based feature detectors with adaptive histogram equalization

ID	DoH		STAR		DoG		MSER	
	Number of detected features	Processing time [sec]						
01	875	0.3431	2438	0.2726	1768	1.8710	1595	1.3043
02	1481	0.3366	3096	0.2742	2698	1.9896	849	1.2230
03	1331	0.3413	1747	0.2682	1861	2.0519	3974	1.5321
04	2419	0.3606	2050	0.2786	2400	2.0114	2530	1.3777
05	1256	0.3399	2682	0.2655	2101	2.0903	2373	1.2644
06	1611	0.3937	1869	0.2739	1388	1.8841	1375	1.3532
07	2168	0.3645	1668	0.2733	1217	1.8724	1415	1.1577
08	1394	0.3463	2115	0.2734	2497	1.9487	1416	1.3393
09	1762	0.3378	2213	0.3023	2384	2.0211	1880	1.7387
10	1957	0.3427	2317	0.2926	663	1.8879	1975	1.7262
Ø	**1625.4**	**0.3507**	**2219.5**	**0.2775**	**1897.7**	**1.9628**	**1938.2**	**1.4017**

B. Visual evaluation of detected feature distribution

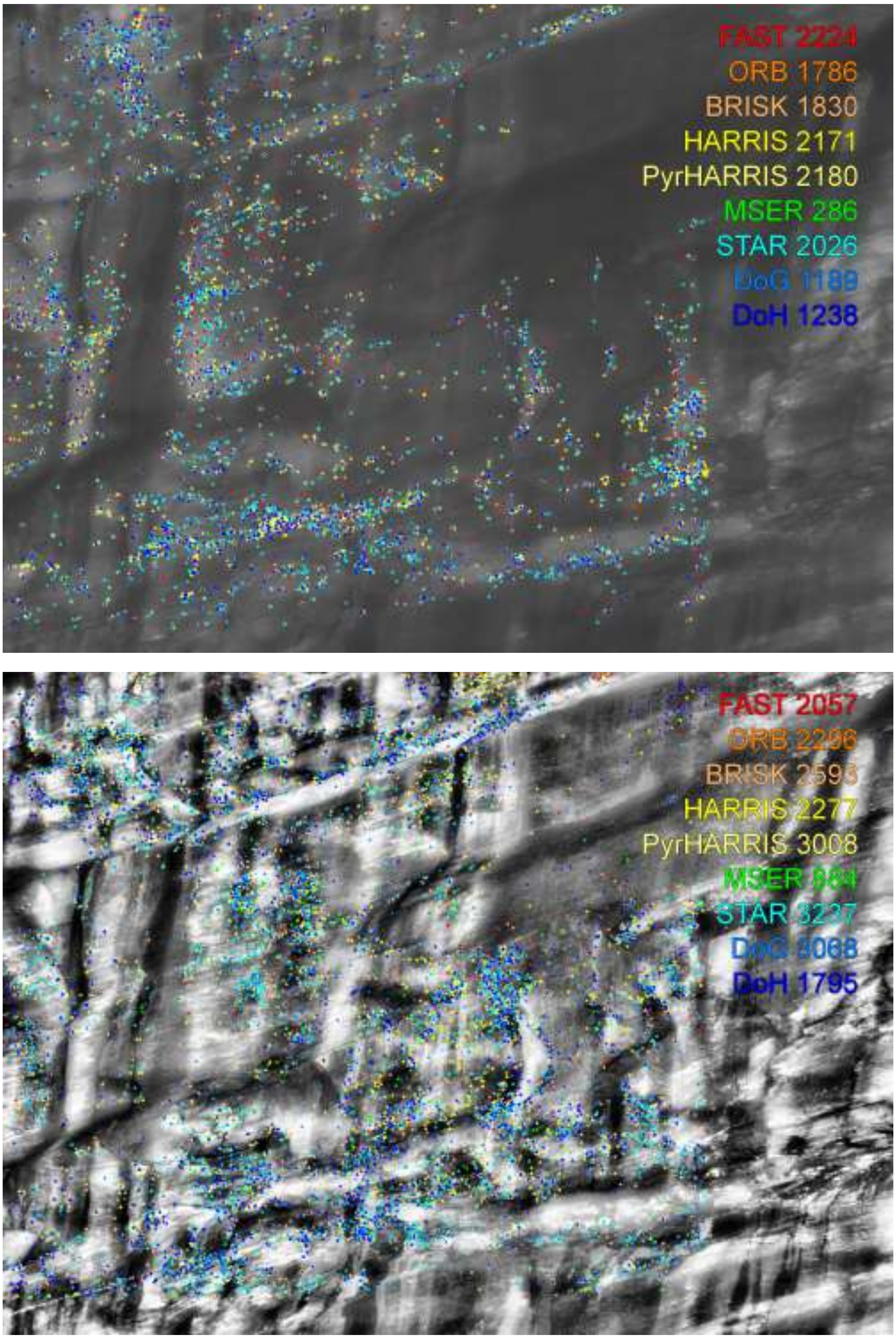

Figure 83: Feature distribution on homogeneous rock wall, before and after adaptive histogram equalization

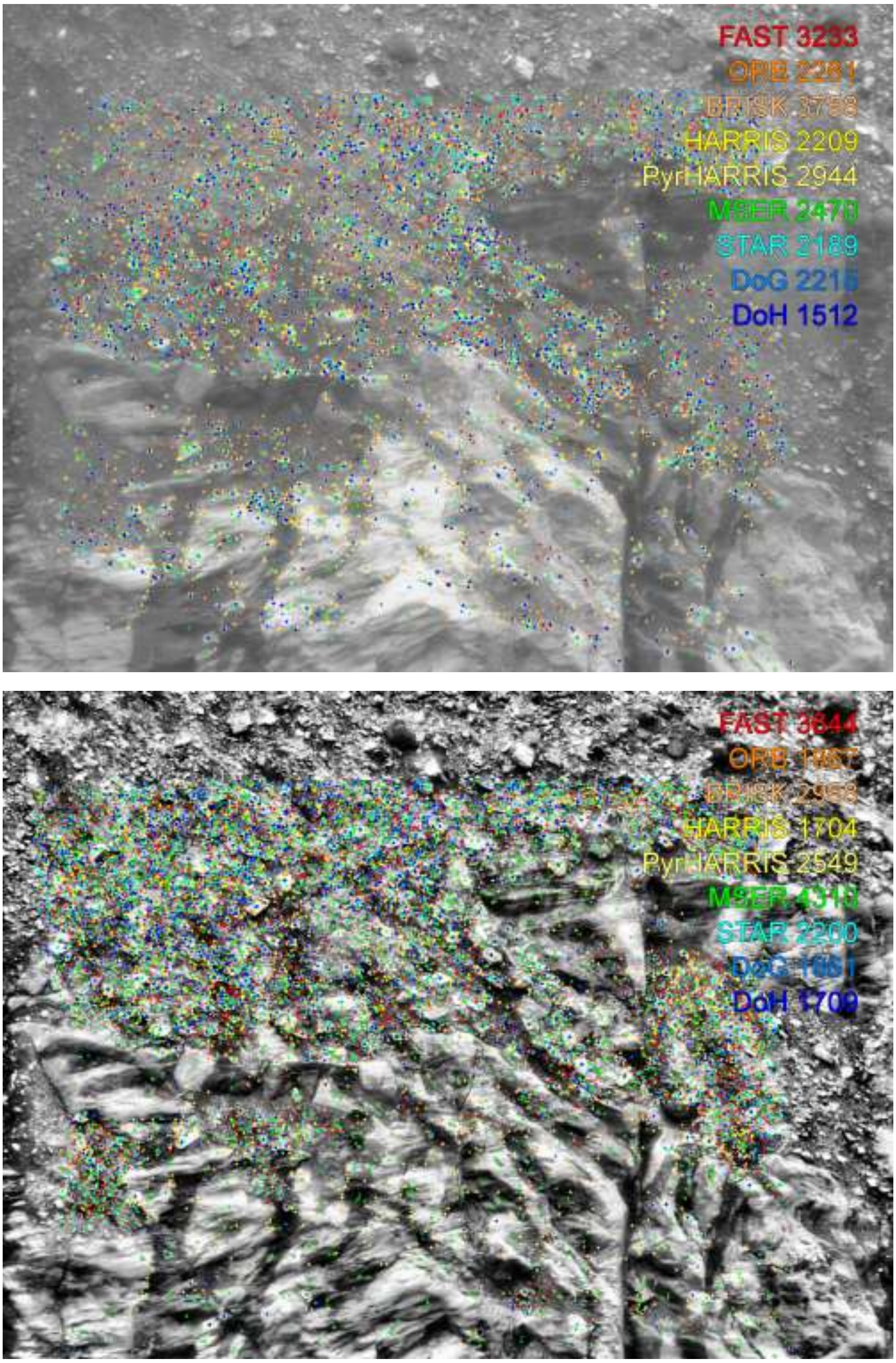

Figure 84: Feature distribution on debris above parts of a homogeneous rock wall

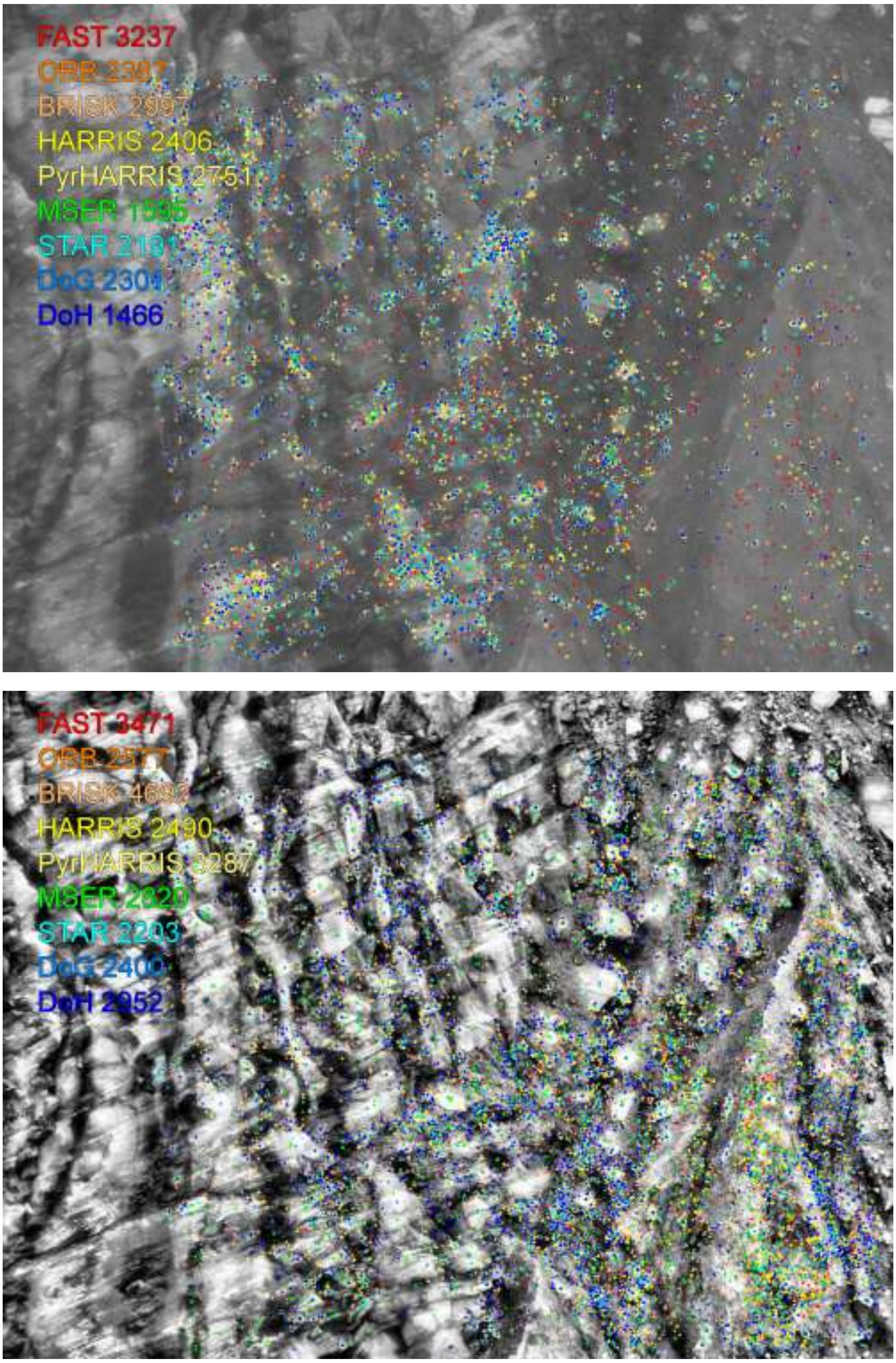

Figure 85: Feature distribution on homogeneous rock wall mixed with inhomogeneous debris

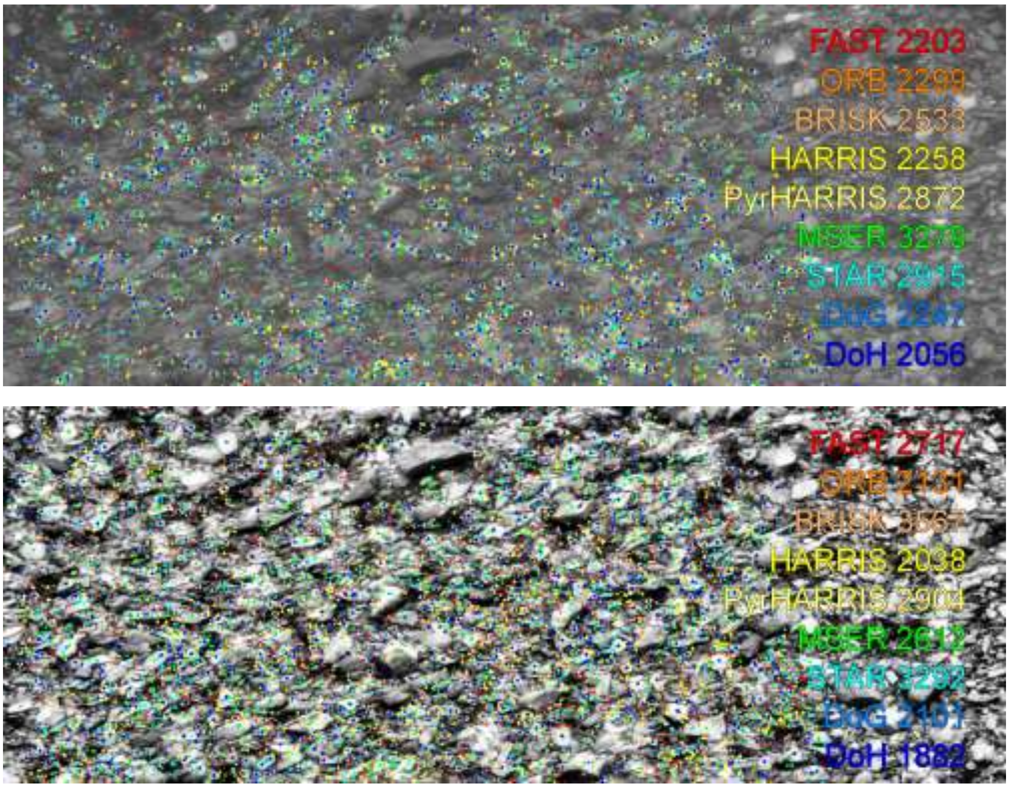

Figure 86: Feature distribution on in-homogenous debris (image cropped)

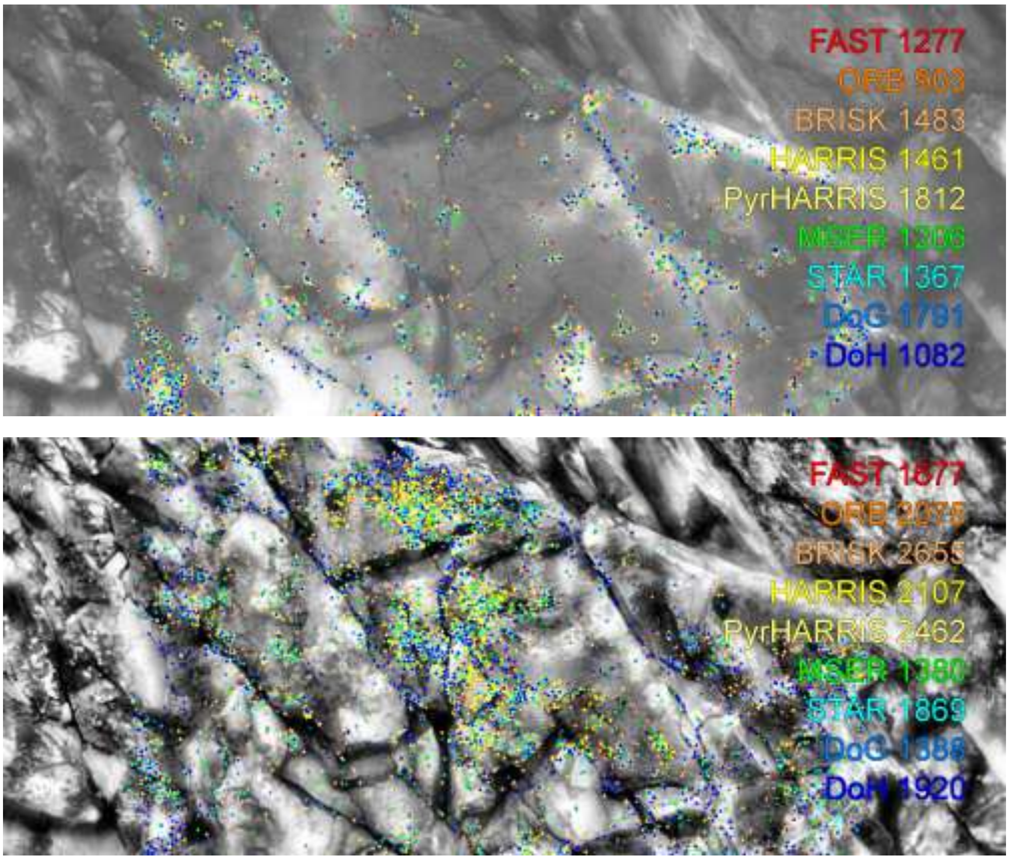

Figure 87: Feature distribution on cracked rock wall (image cropped)

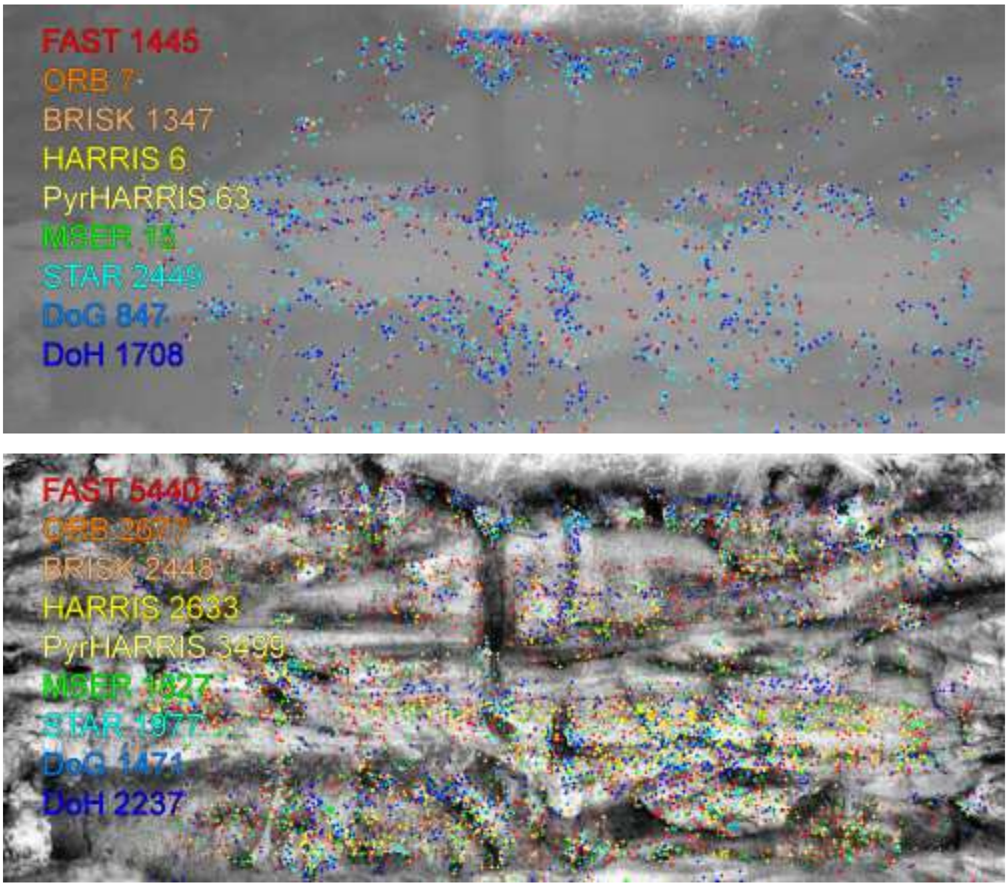

Figure 88: Feature distribution on weakly textured clay wall (image cropped)

C. Feature detector performance evaluation charts

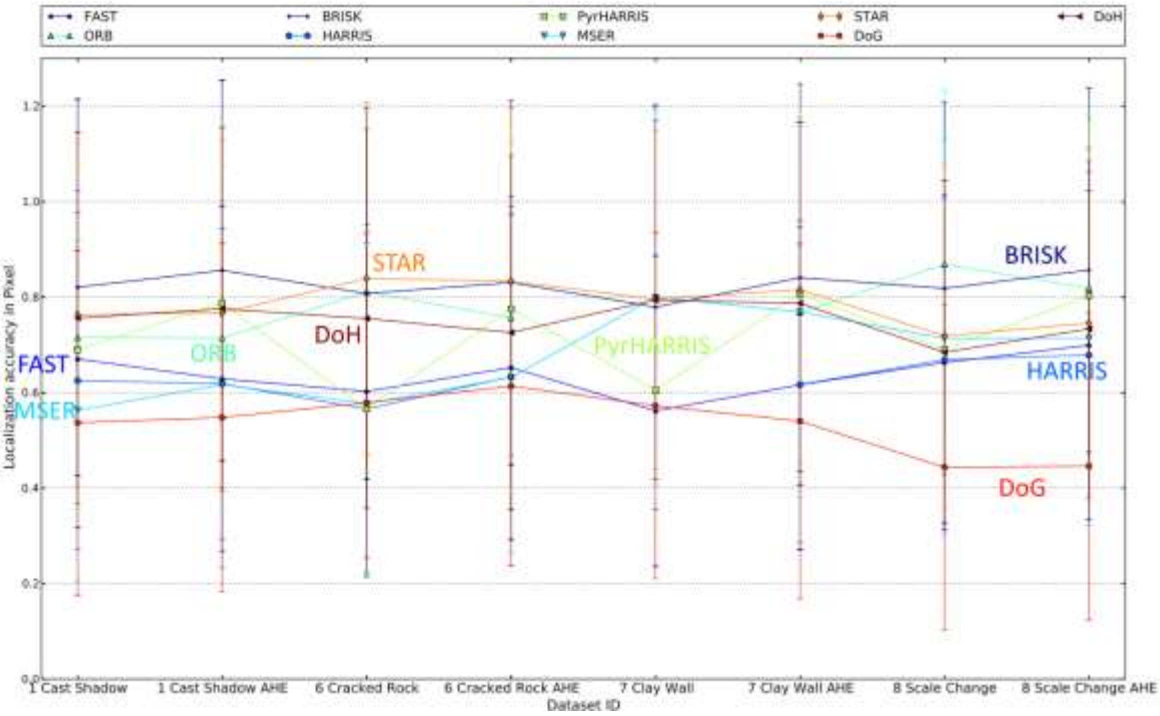

Figure 89: Localization accuracy on cast shadow, cracked rock, clay wall and scale change data-sets

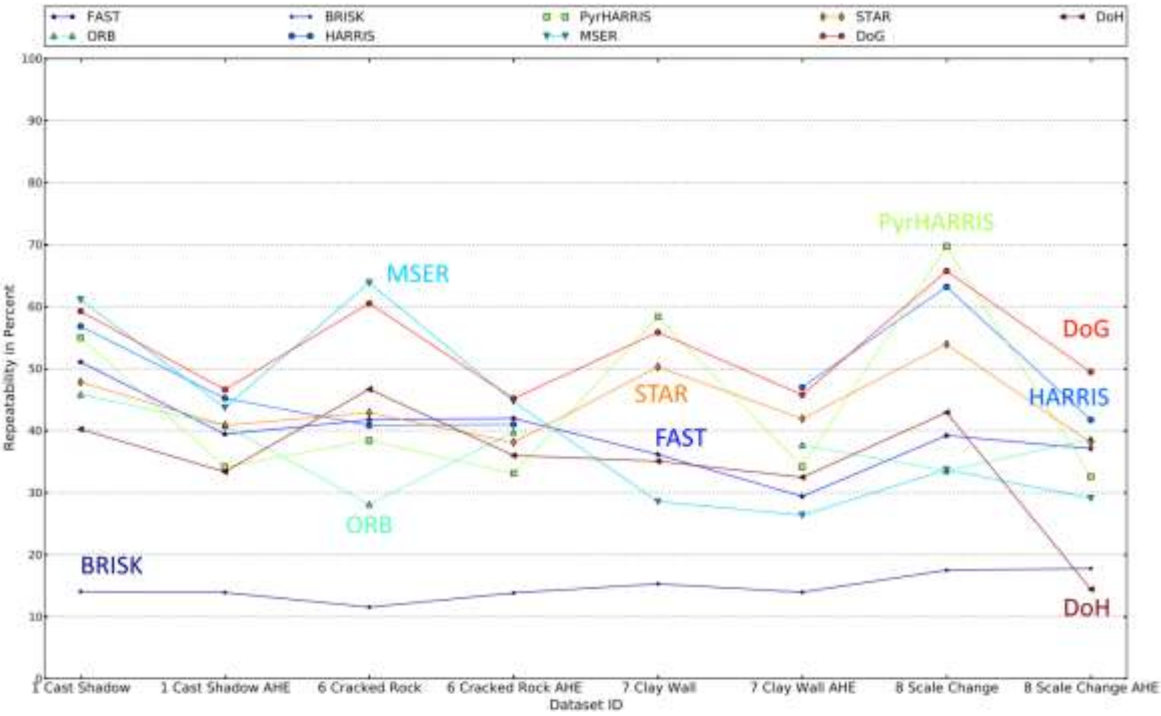

Figure 90: Detector repeatability on cast shadow, cracked rock, clay wall and scale change data-sets

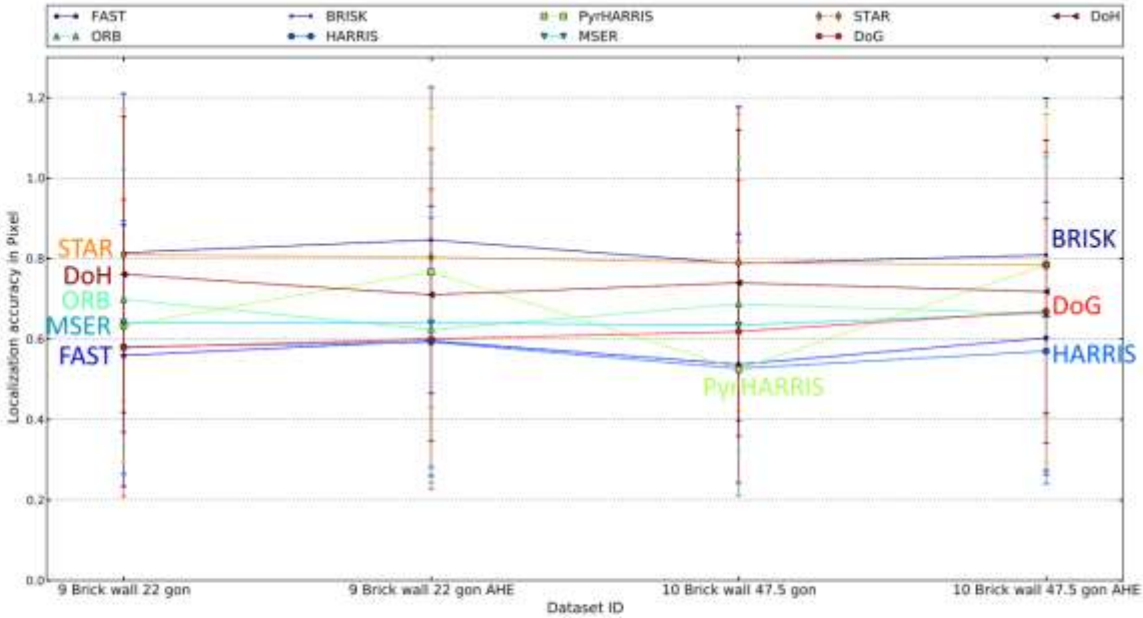

Figure 91: Brick Wall increasing viewpoint data-sets detector localization accuracy

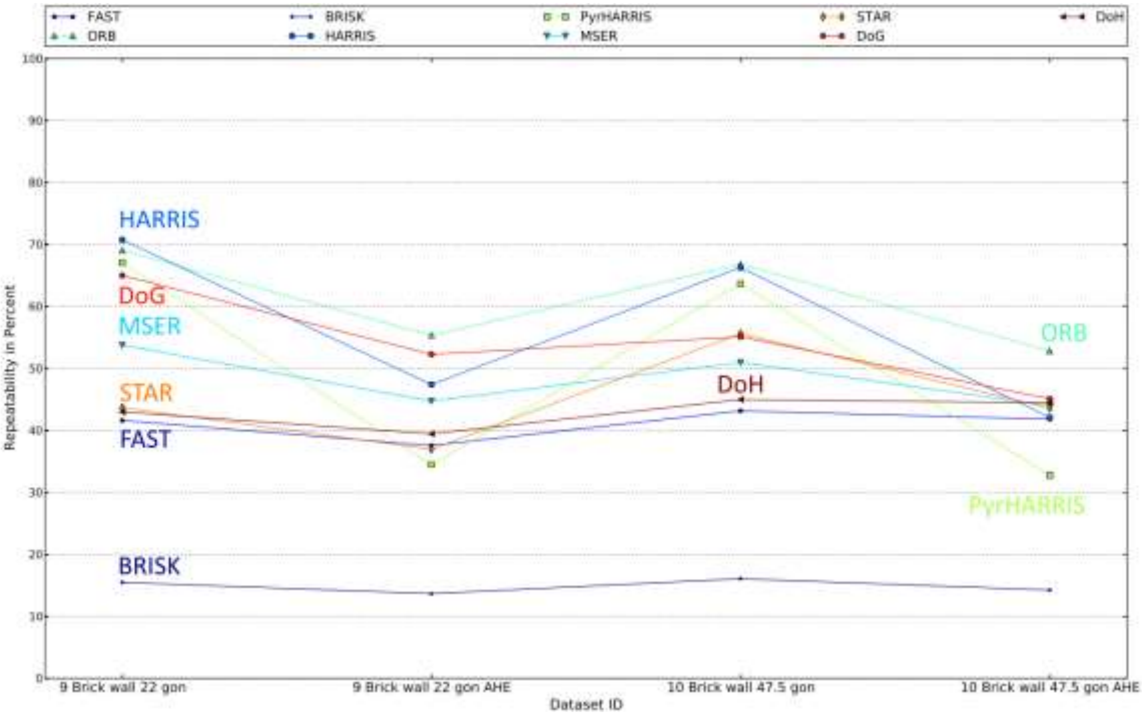

Figure 92: Brick Wall increasing viewpoint data-sets detector repeatability

D. Feature descriptor performance evaluation charts

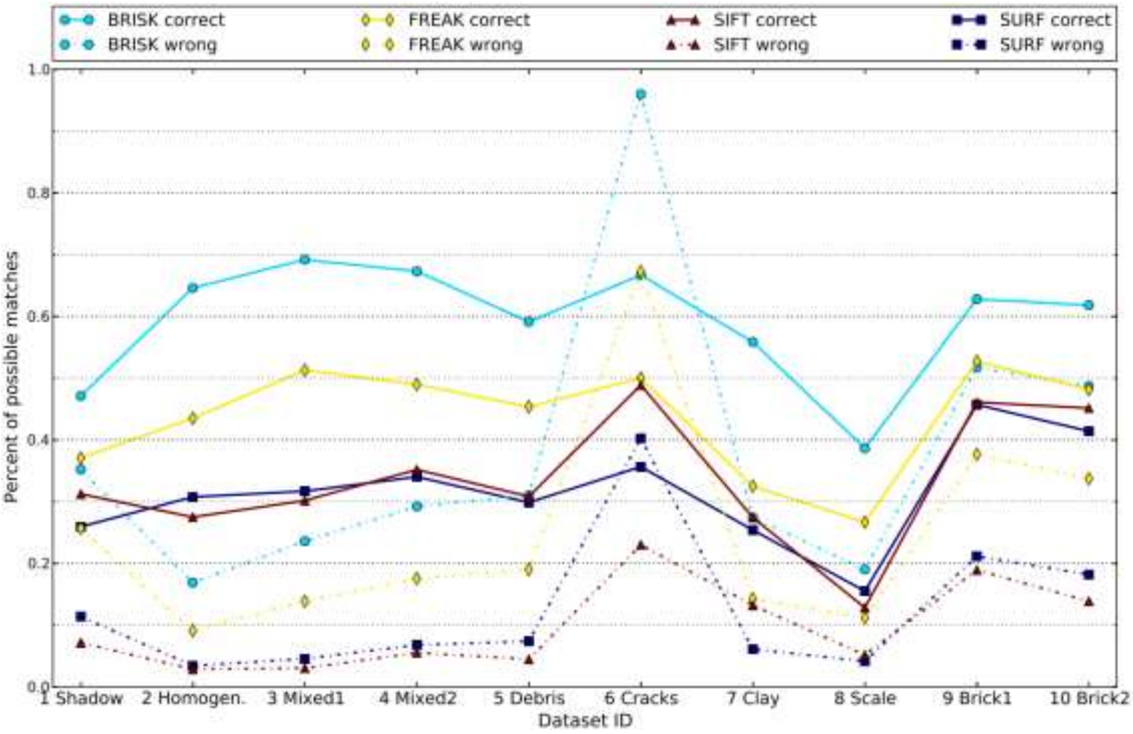

Figure 93: Matches from descriptors on BRISK features as fraction of maximally possible matches, divided in correct and incorrect matches

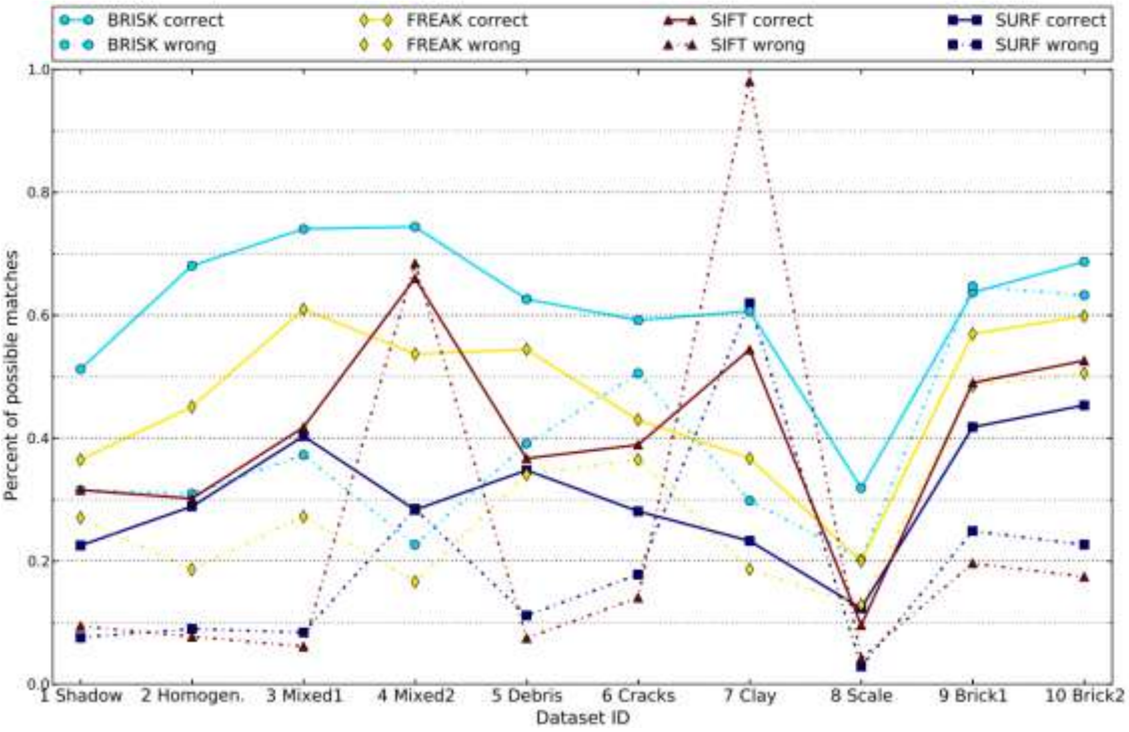

Figure 94: Matches from descriptors on BRISK features computed with AHE pre-processing

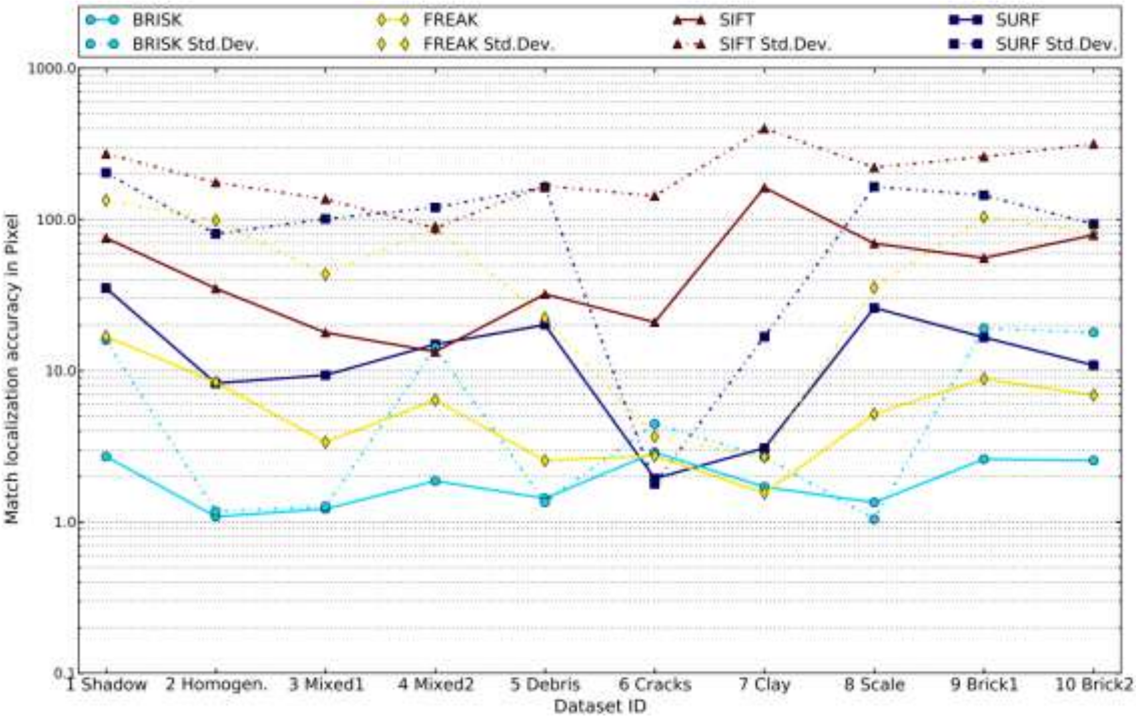

Figure 95: Matching accuracy on BRISK features including empirical standard deviation on logarithmic scale

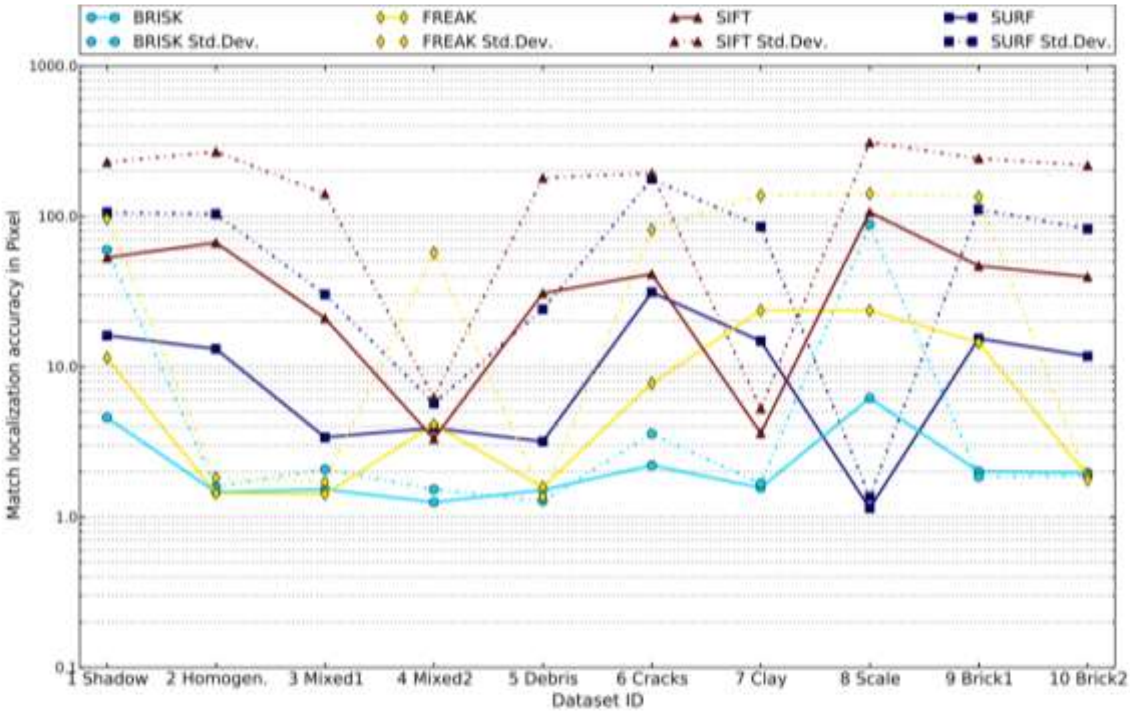

Figure 96: Matching accuracy on BRISK features computed with AHE pre-processing

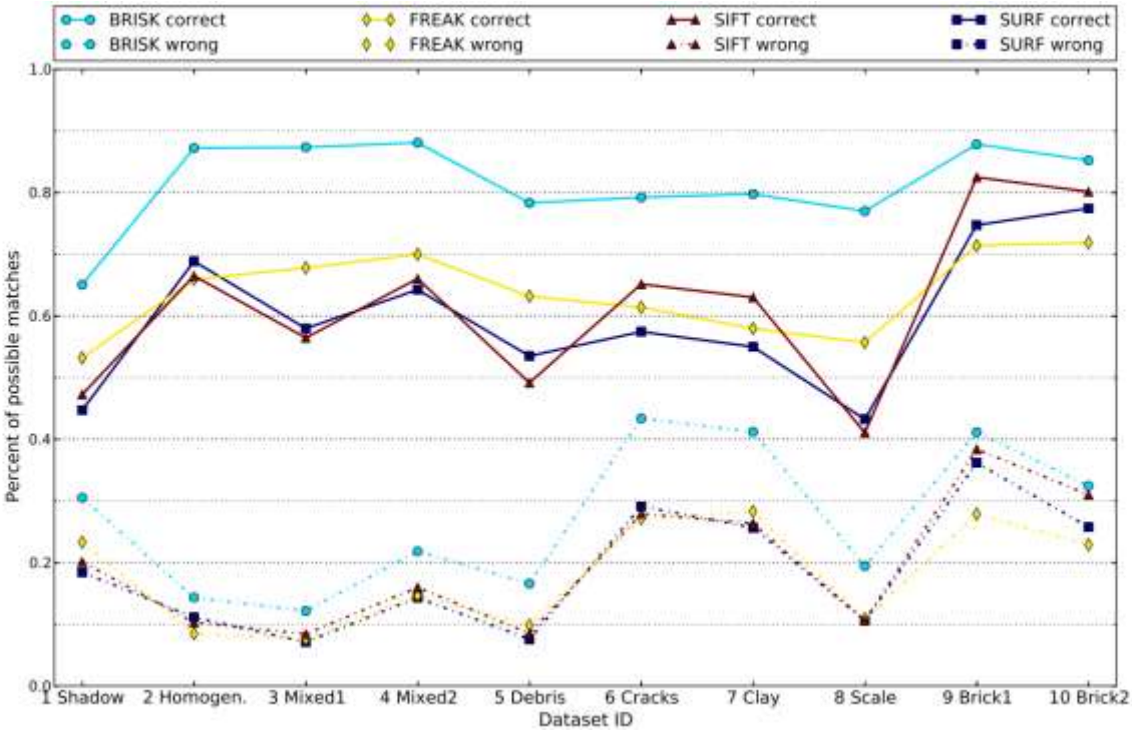

Figure 97: Matches from descriptors on DoH features

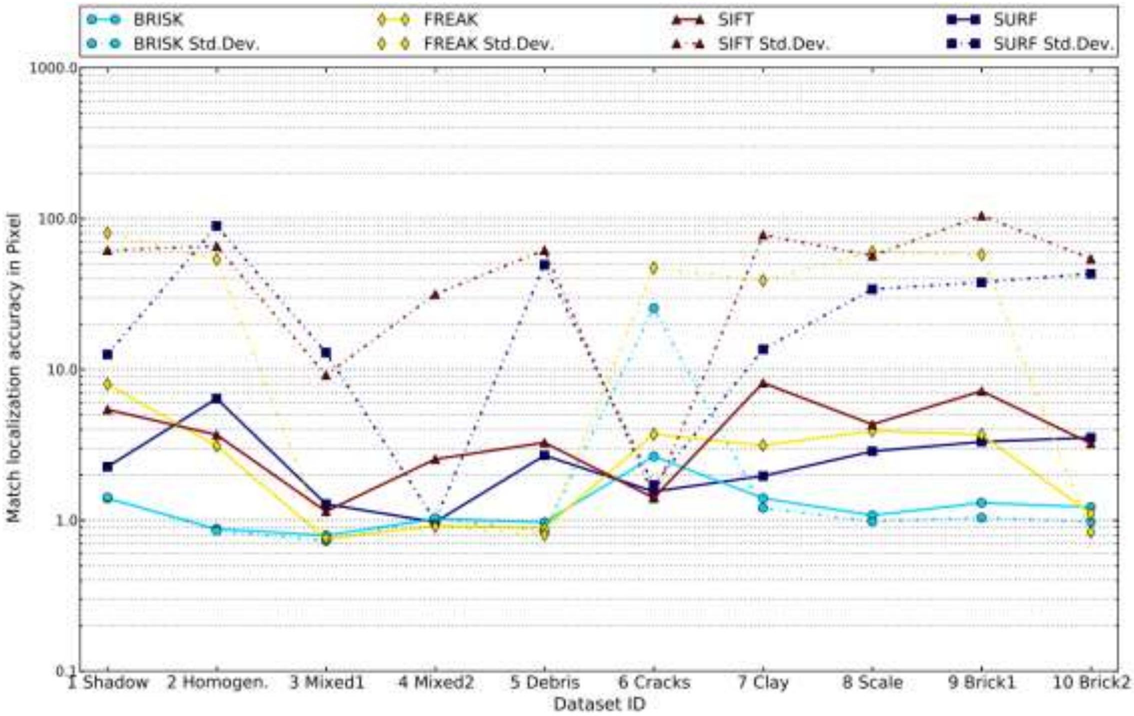

Figure 98: Matching accuracy on DoH features

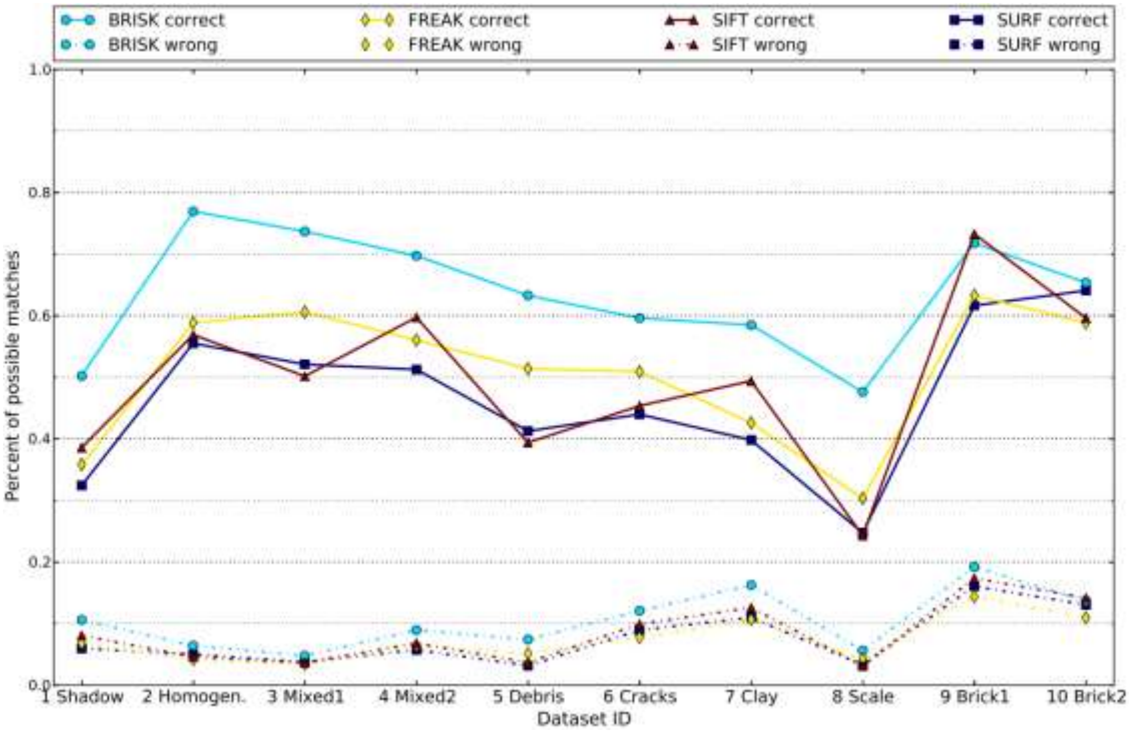

Figure 99: Matches from descriptors on DoH features with additional outlier filtering

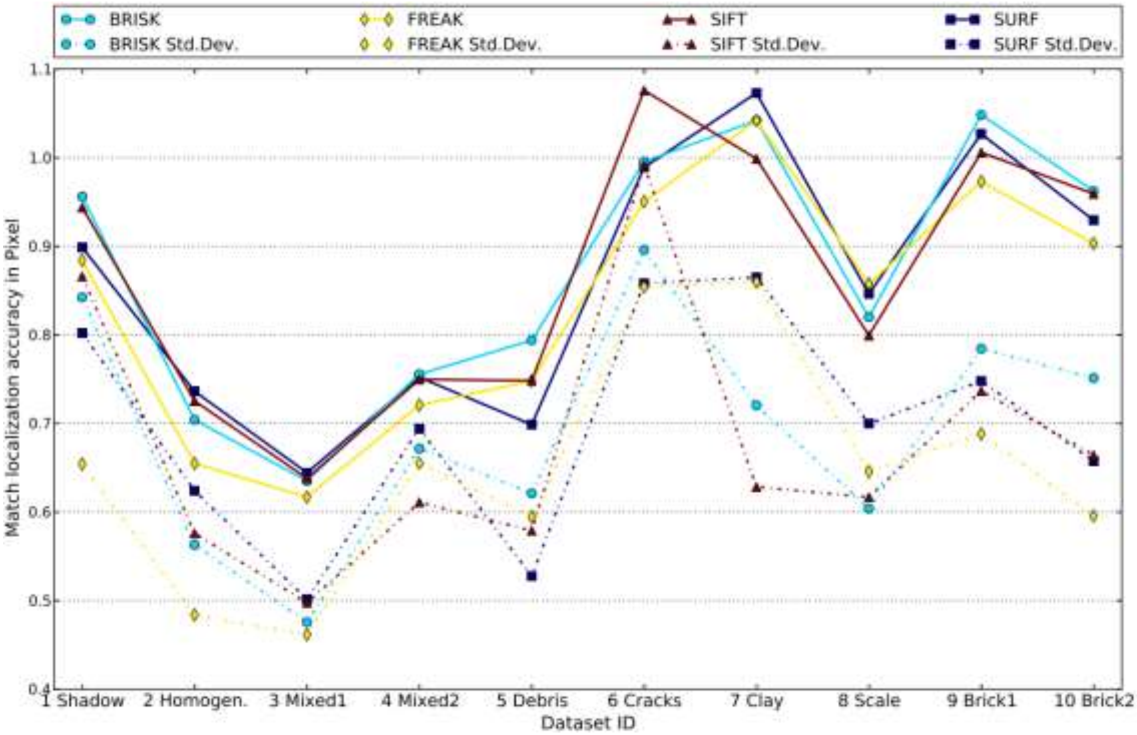

Figure 100: Matching accuracy on DoH features with additional outlier filtering

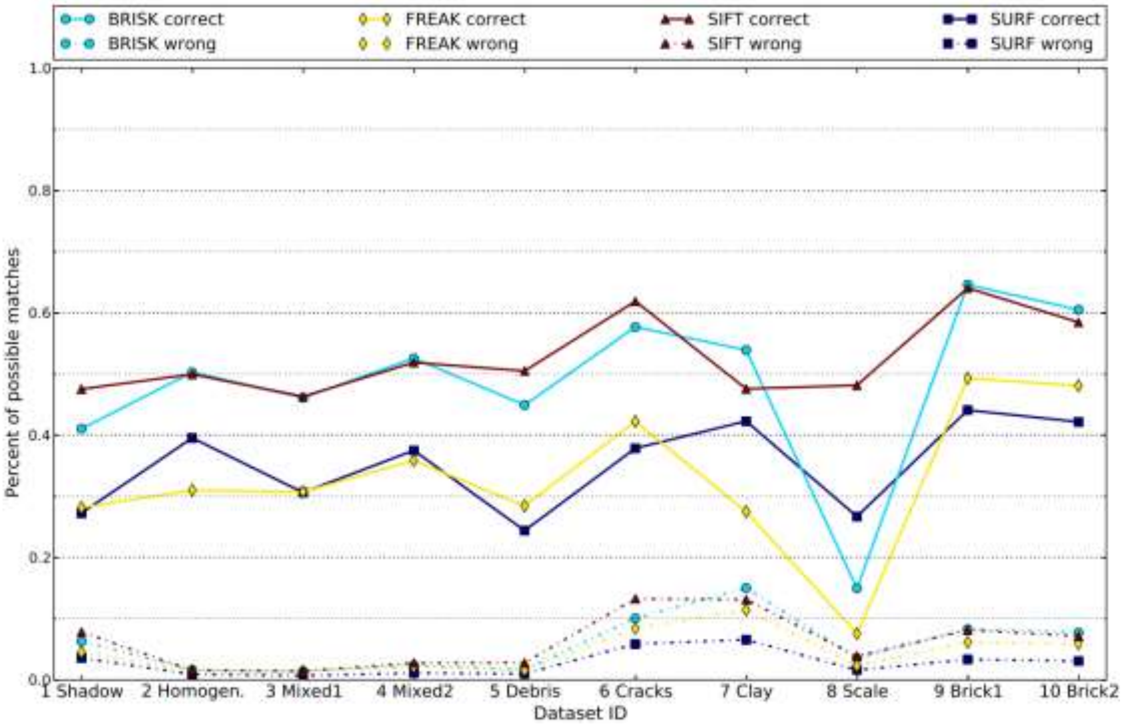

Figure 101: Matches on DoG features

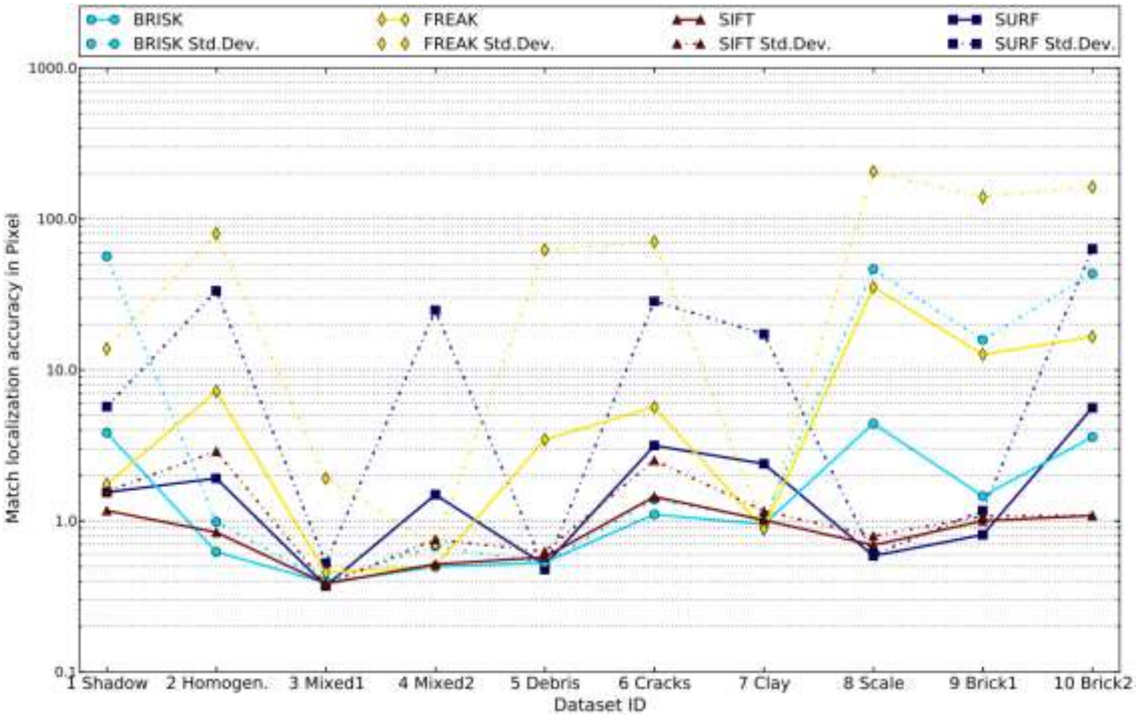

Figure 102: Matching accuracy on DoG features

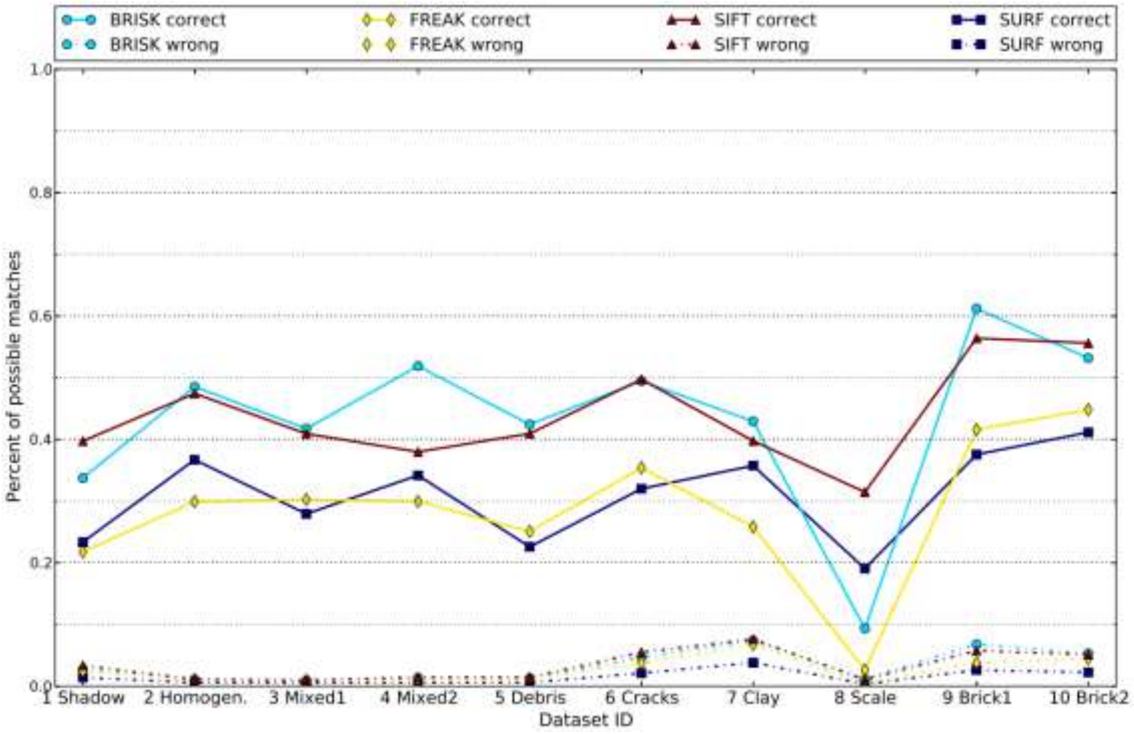

Figure 103: Matches on DoG features with additional outlier filtering

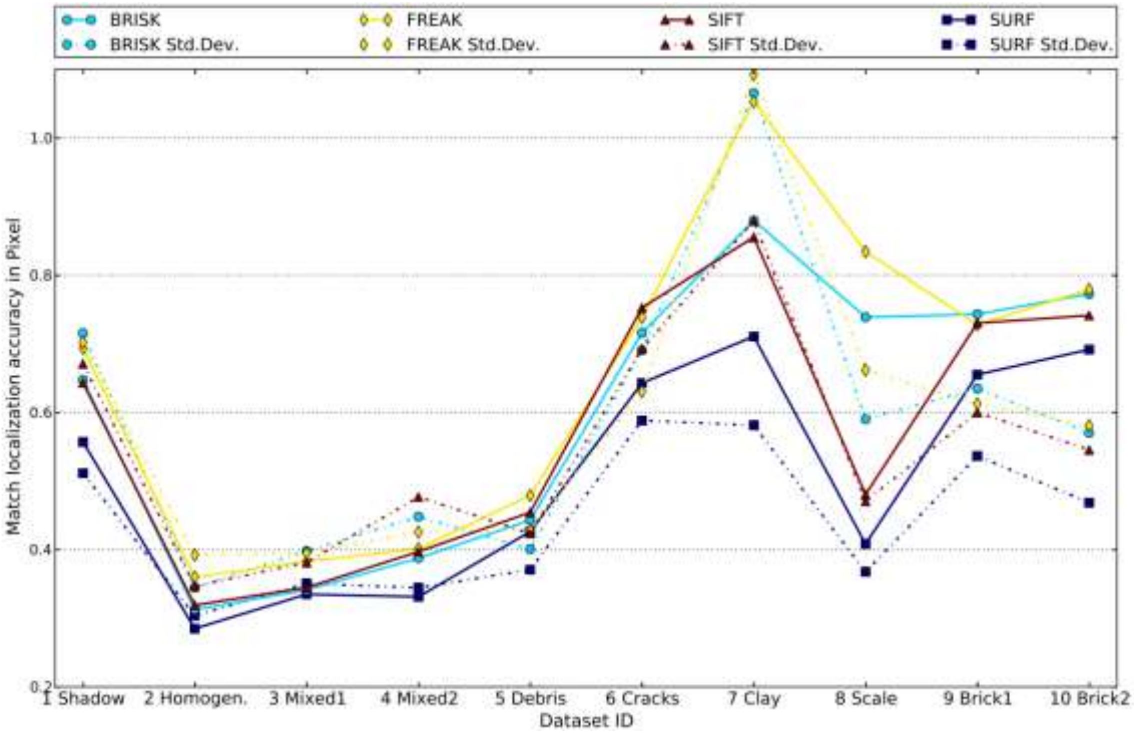

Figure 104: Matching accuracy on DoG features with additional outlier filtering

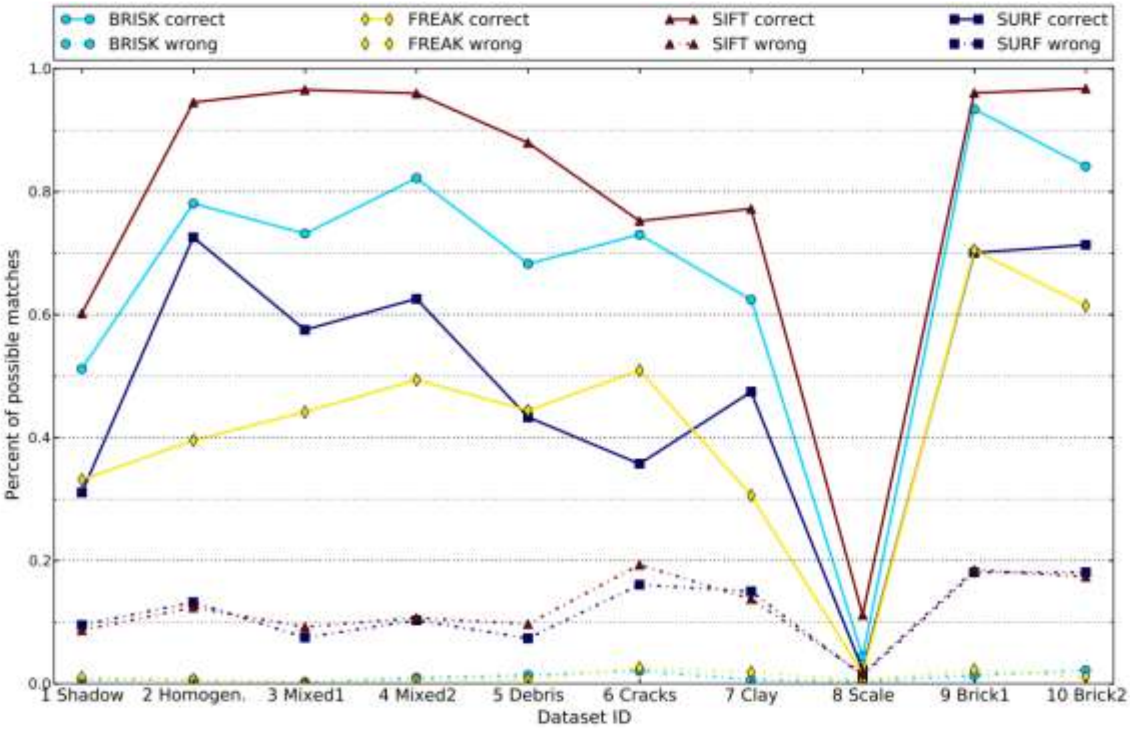

Figure 105: Matches from descriptors on FAST features

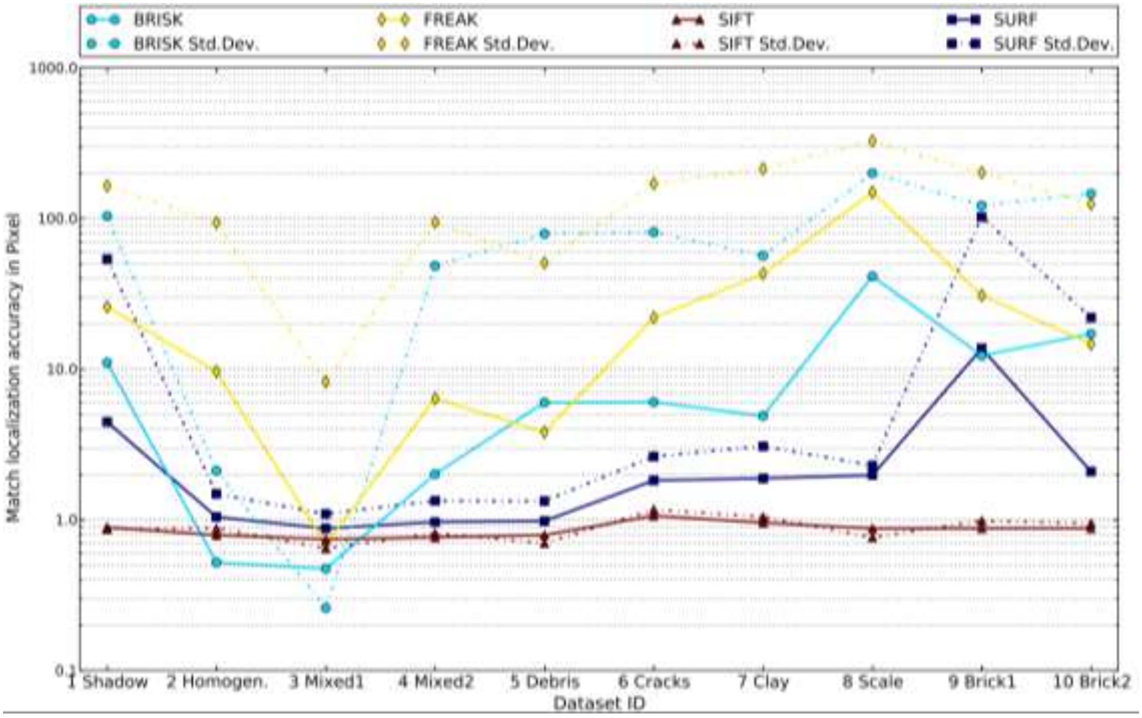

Figure 106: Matching accuracy on FAST features

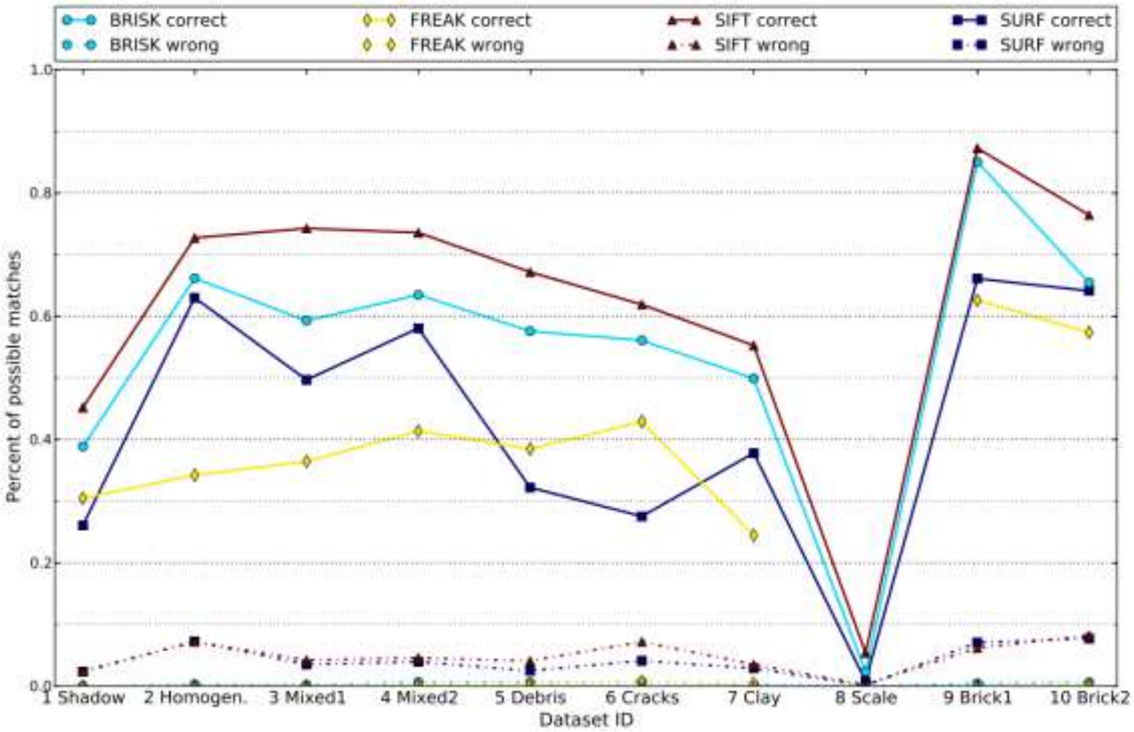

Figure 107: Matches on FAST features with additional outlier filtering

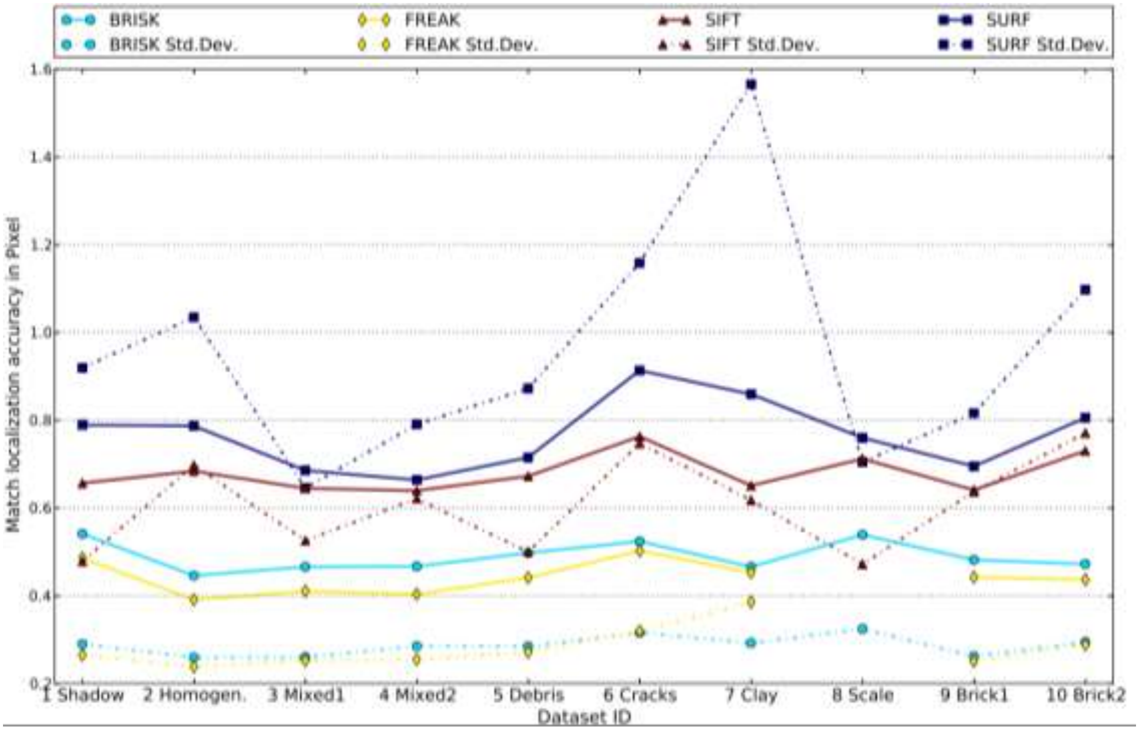

Figure 108: Matching accuracy on FAST features with additional outlier filtering

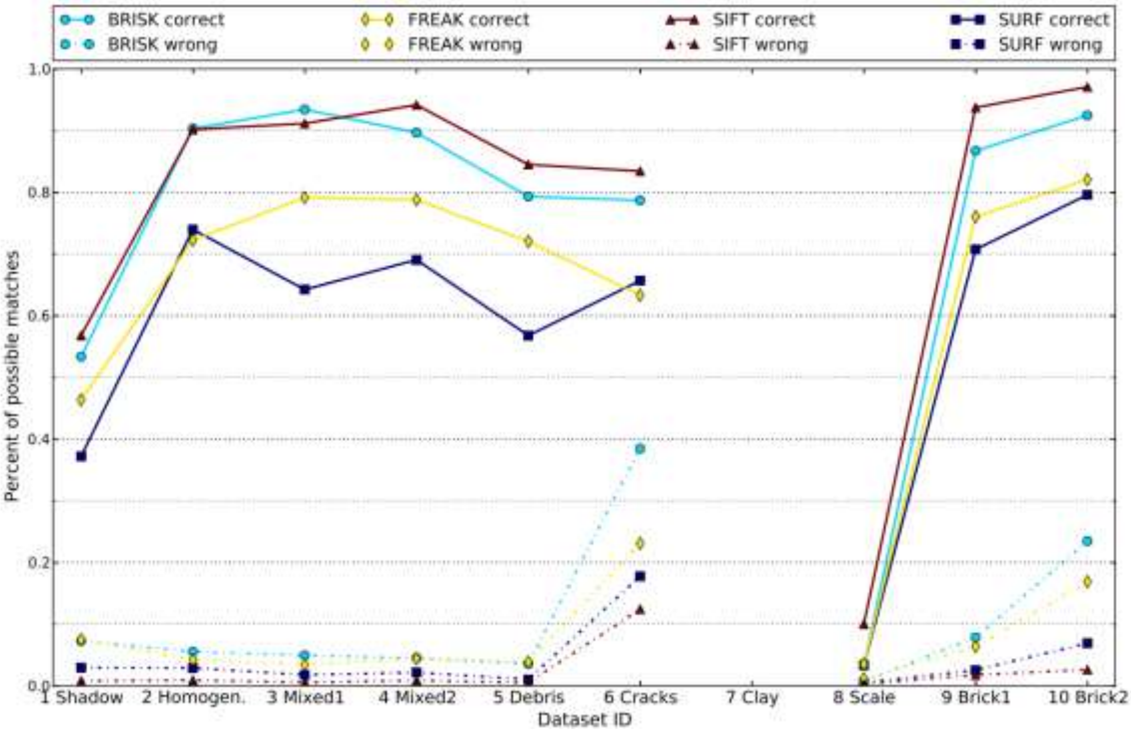

Figure 109: Matches on HARRIS features

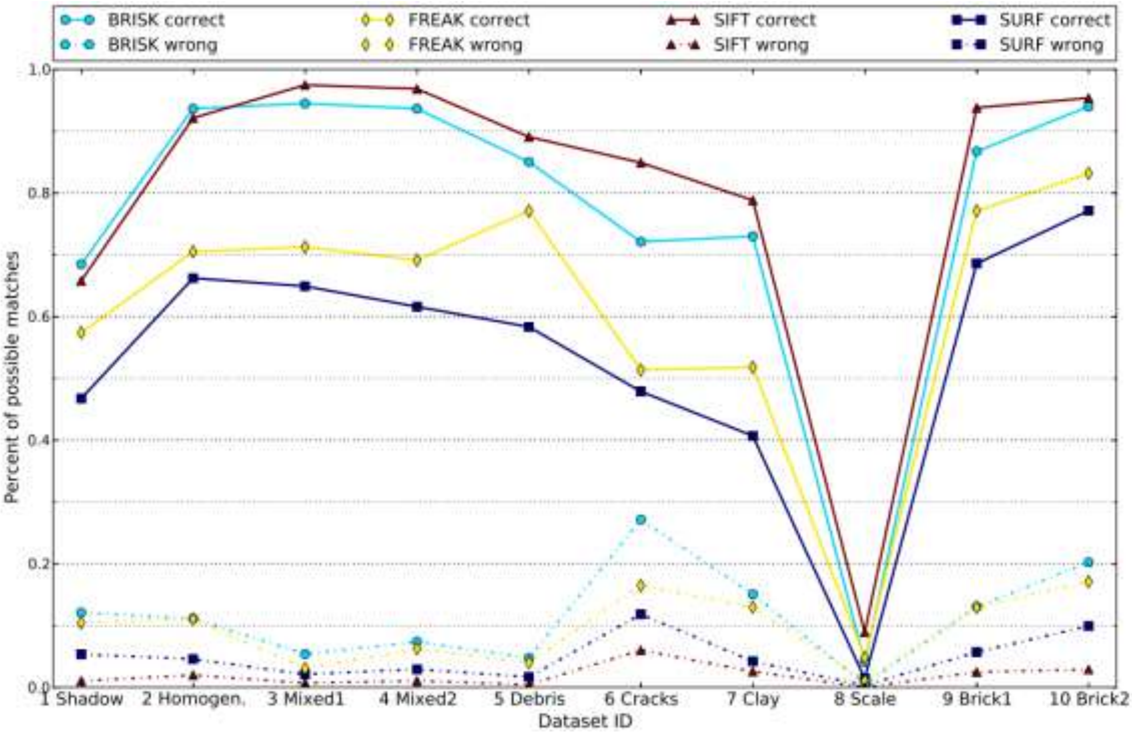

Figure 110: Matches on HARRIS features computed with AHE pre-processing

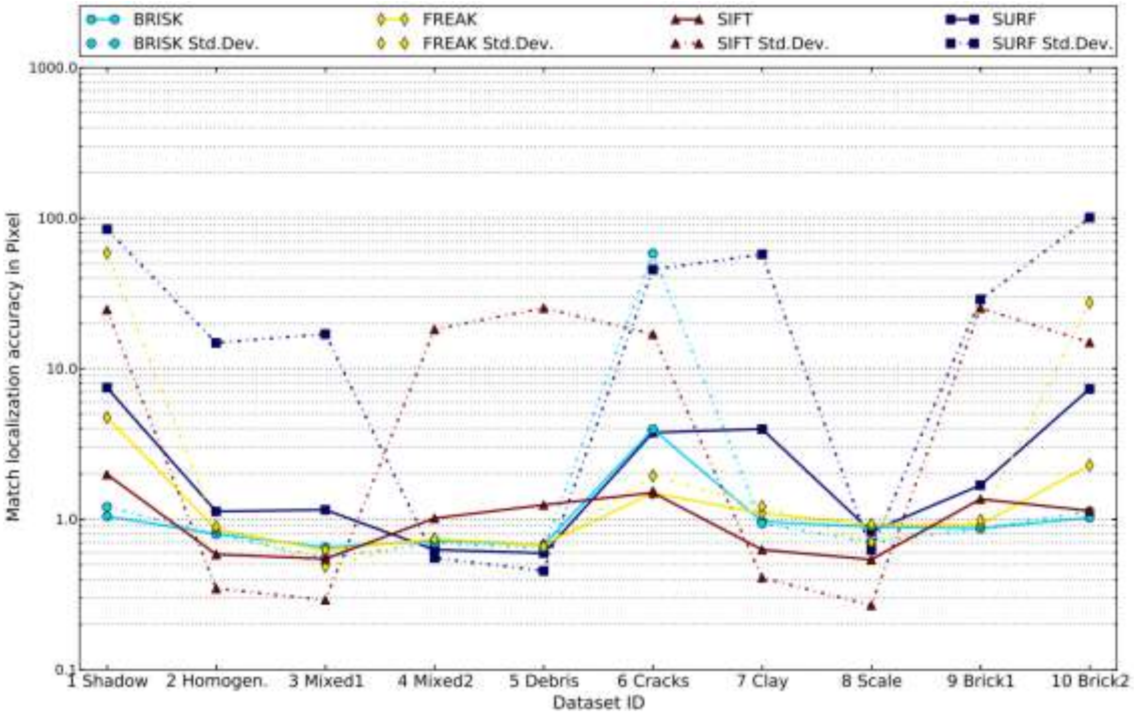

Figure 111: Matching accuracy on HARRIS features computed with AHE pre-processing

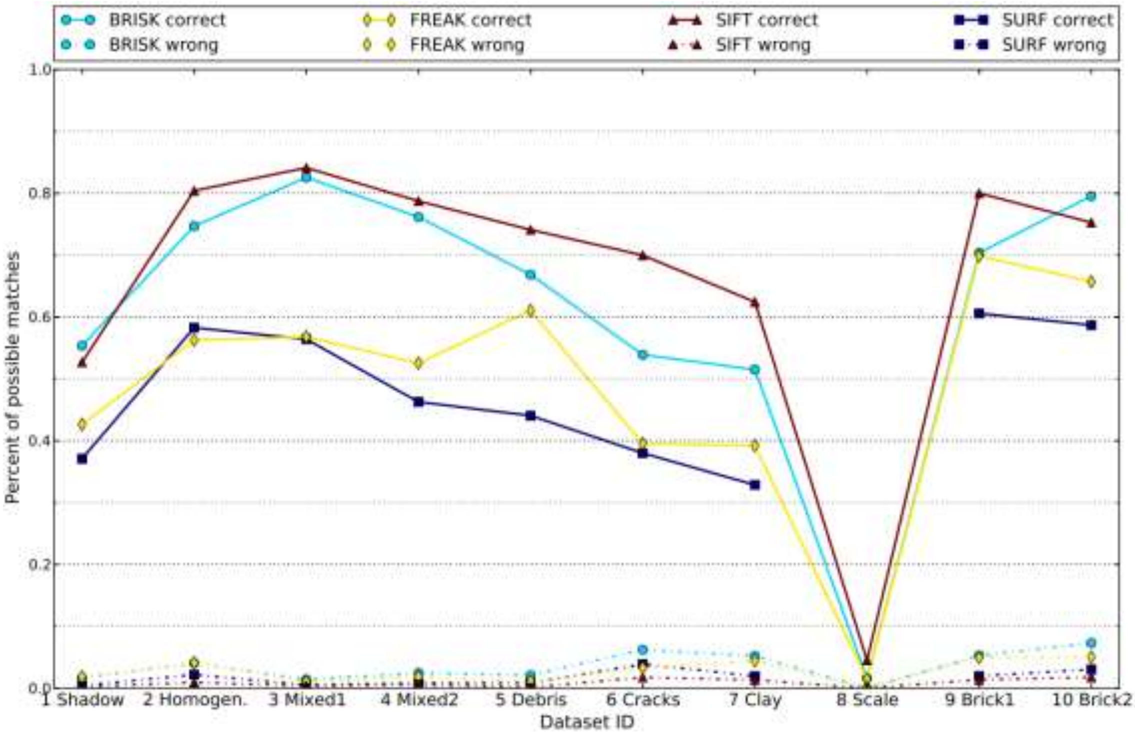

Figure 112: Matches on HARRIS features computed with AHE with outlier filtering

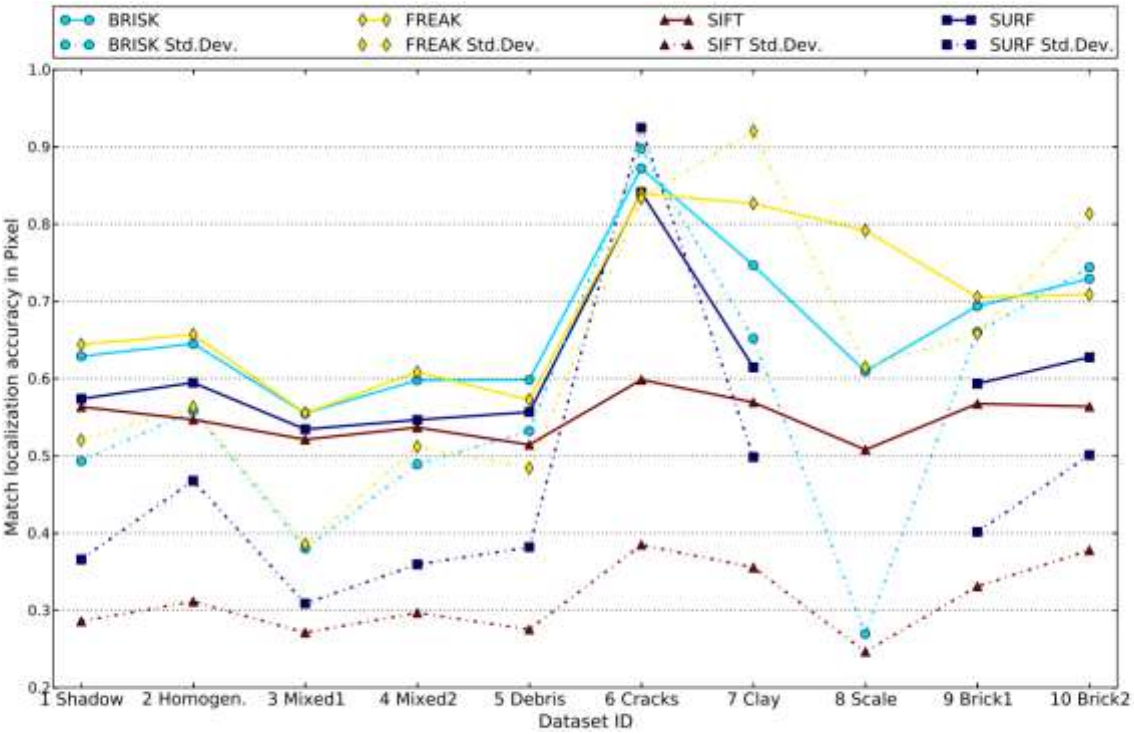

Figure 113: Matching accuracy on HARRIS features computed with AHE with outlier filtering

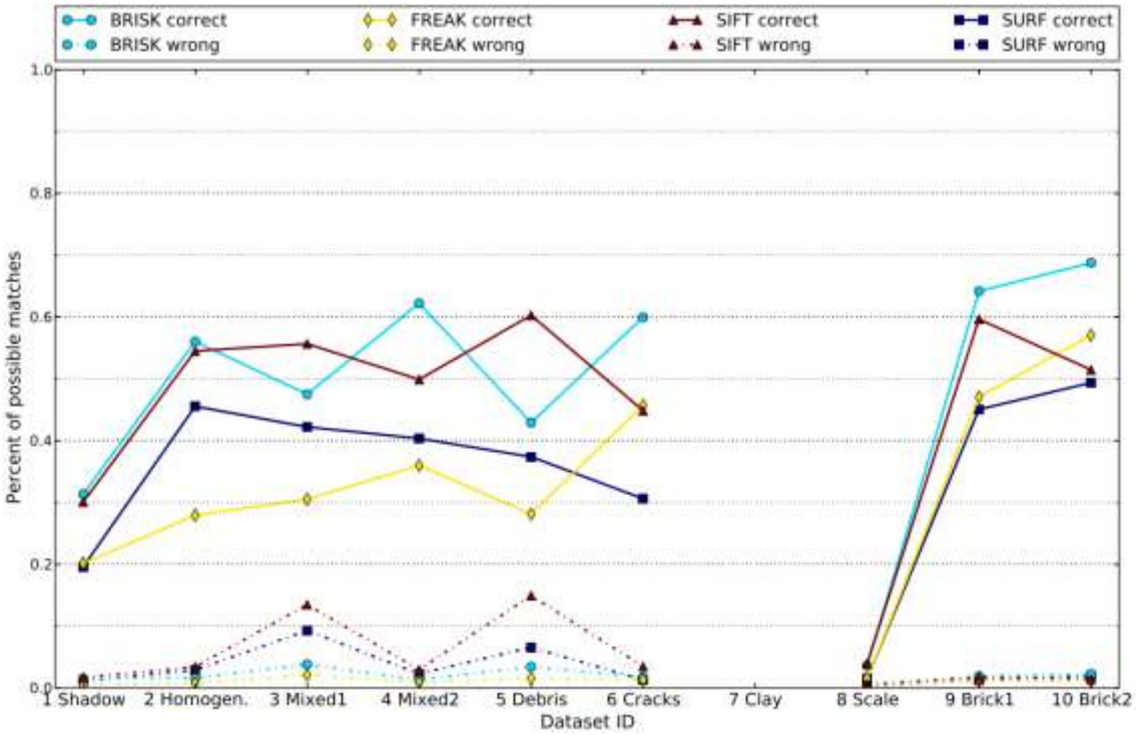

Figure 114: Matches on Pyramid HARRIS features with additional outlier filtering

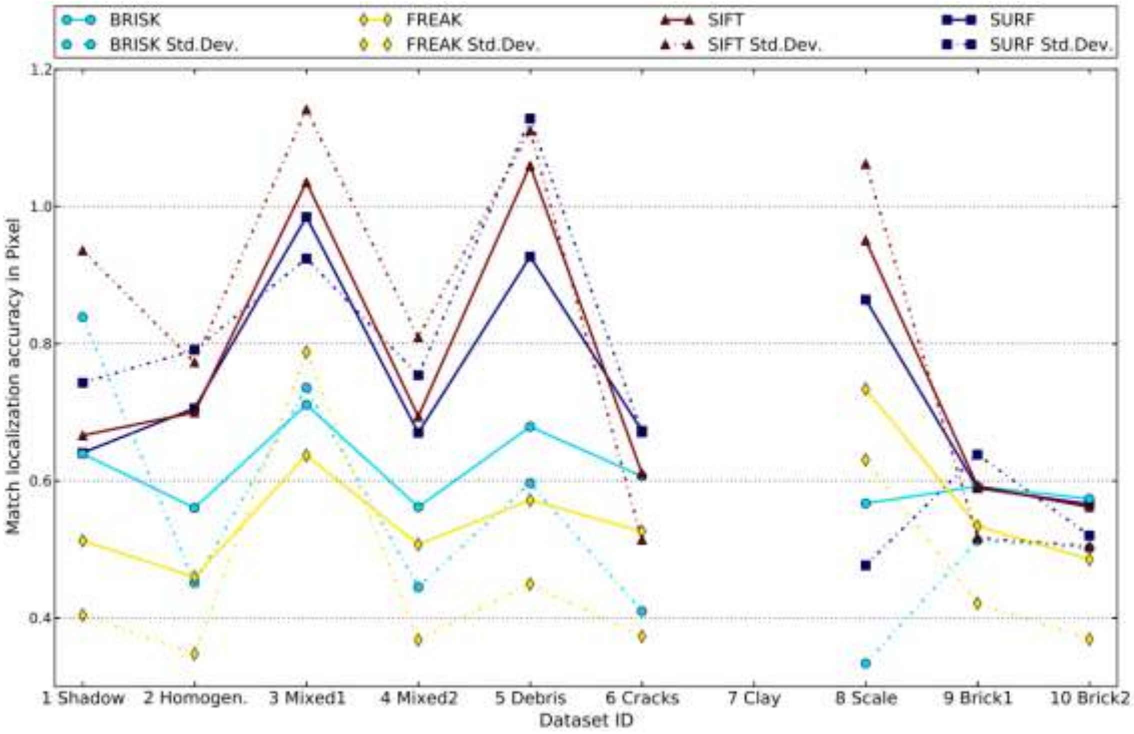

Figure 115: Matching accuracy on Pyramid HARRIS features with additional outlier filtering

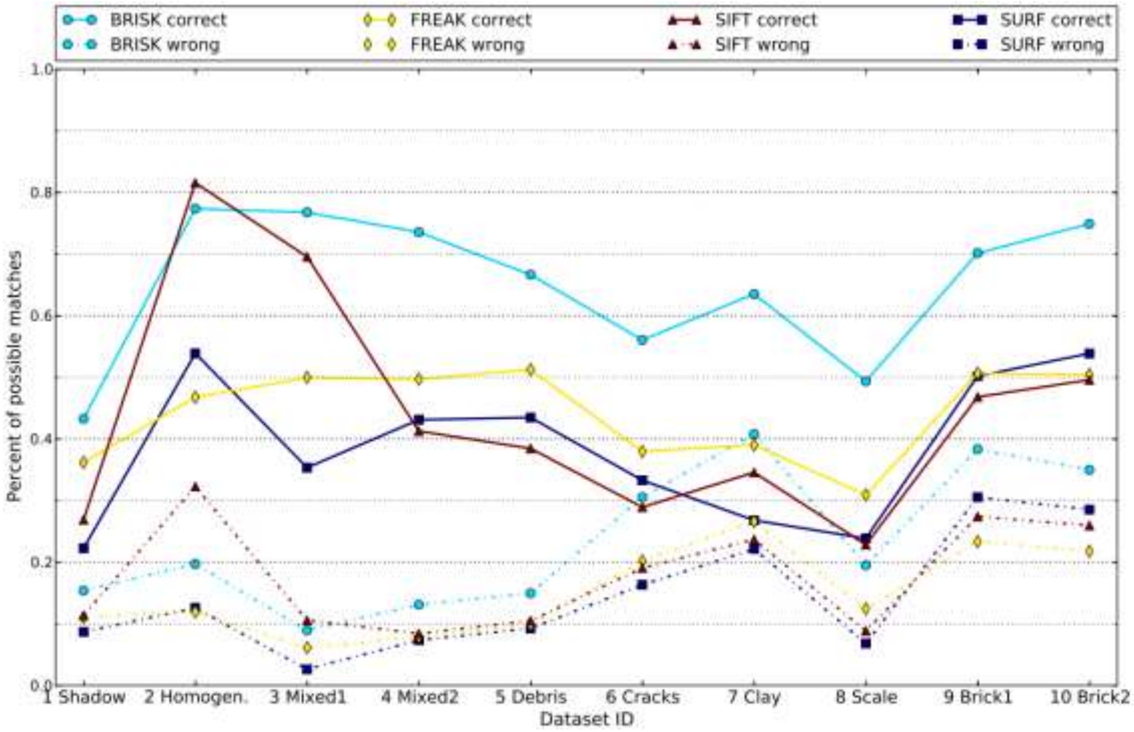

Figure 116: Matches on MSER features computed with AHE

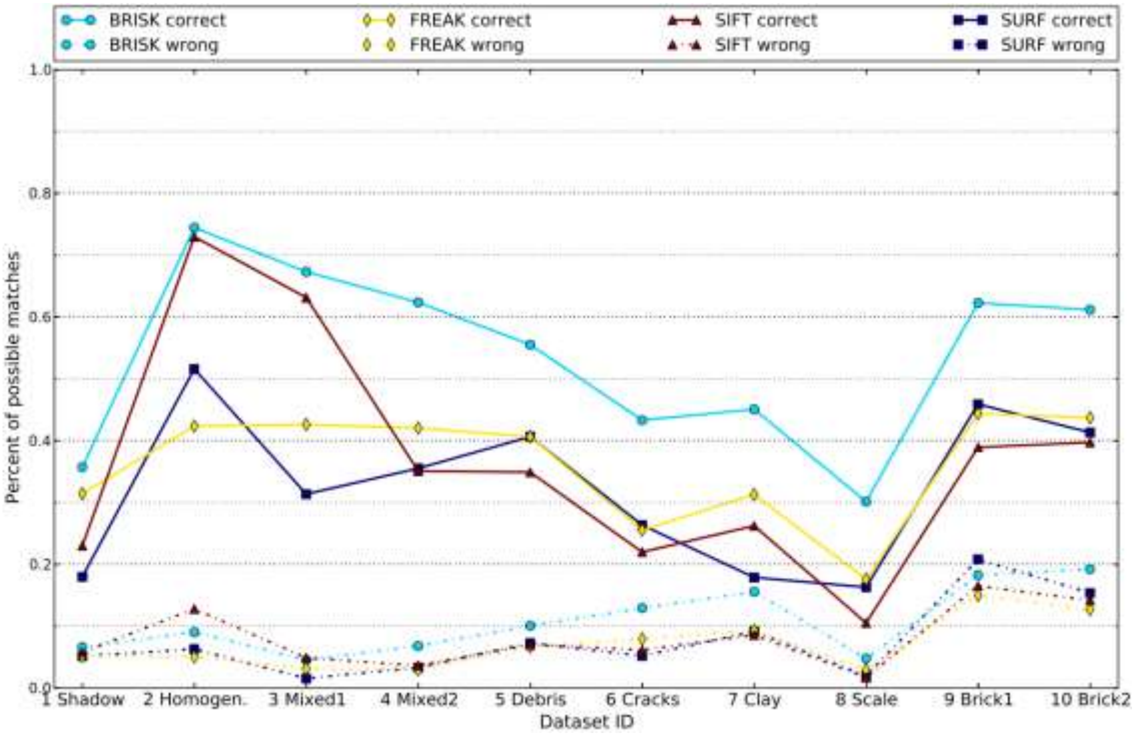

Figure 117: Matches on MSER features computed with AHE with outlier filtering

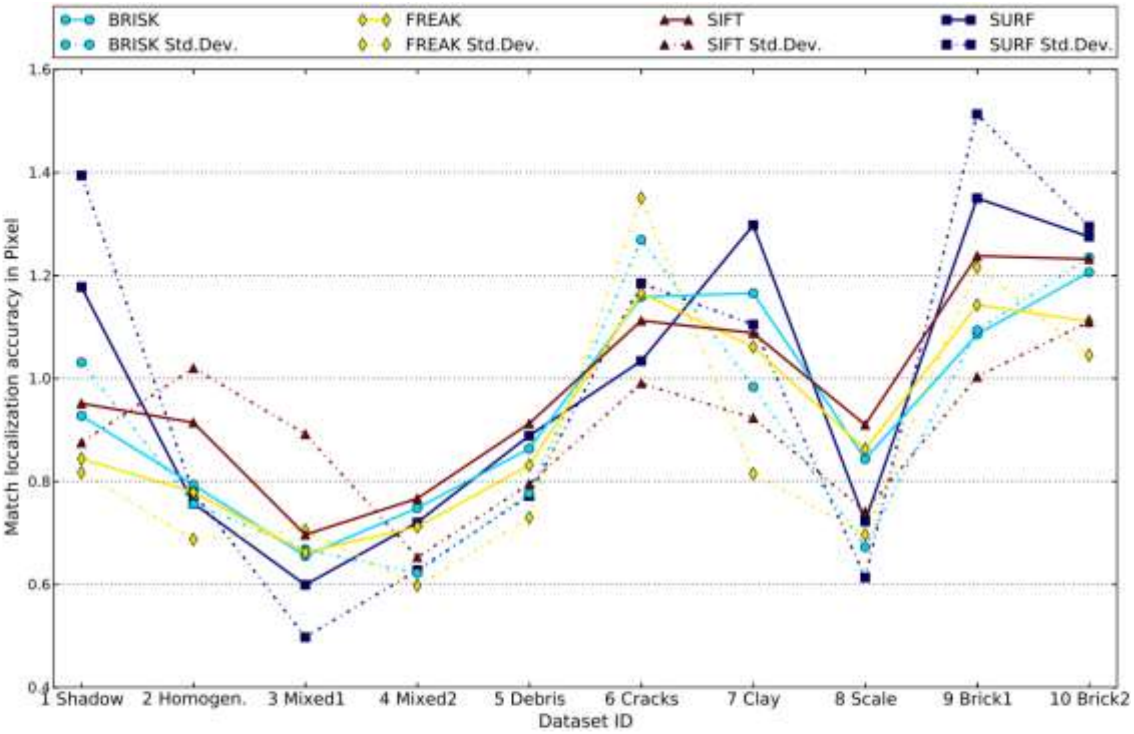

Figure 118: Matching accuracy on MSER features computed with AHE with outlier filtering

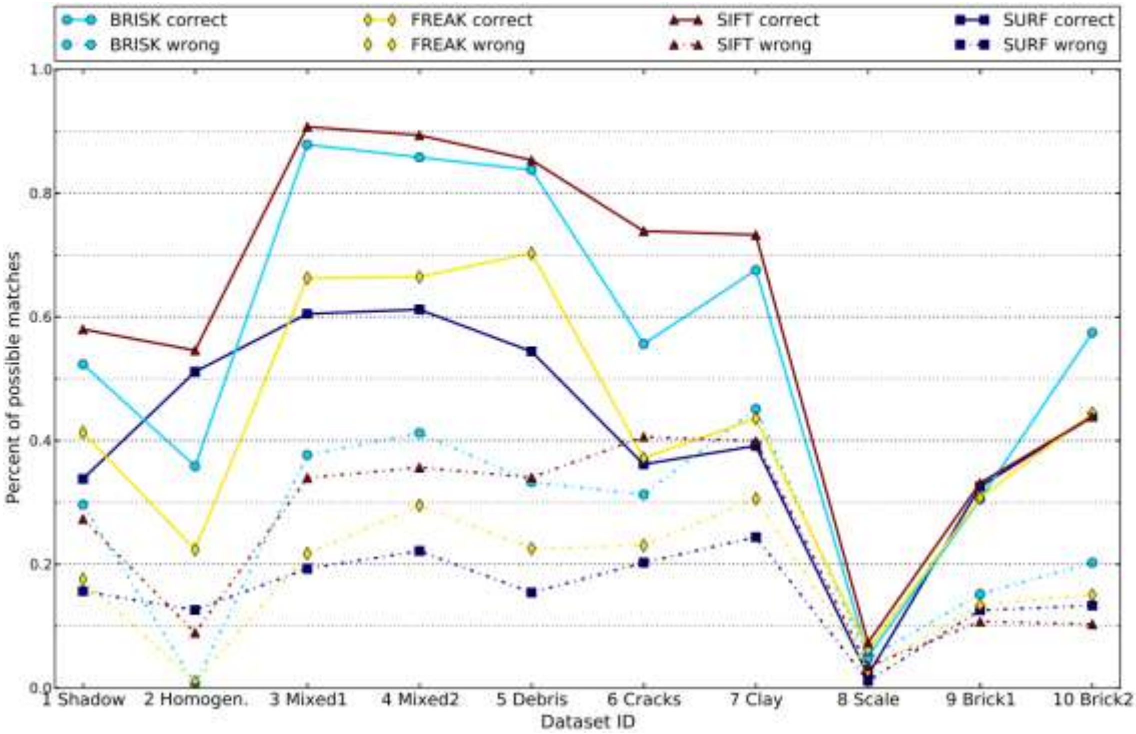

Figure 119: Matches on ORB features computed with AHE

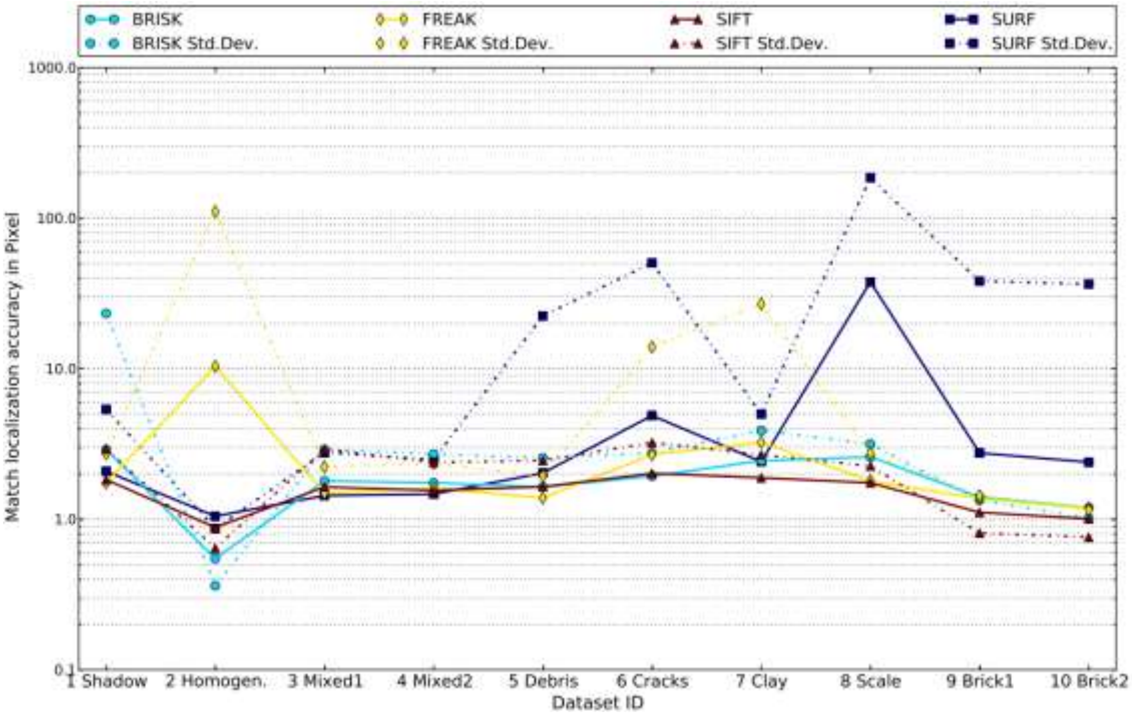

Figure 120: Matching accuracy on ORB features computed with AHE

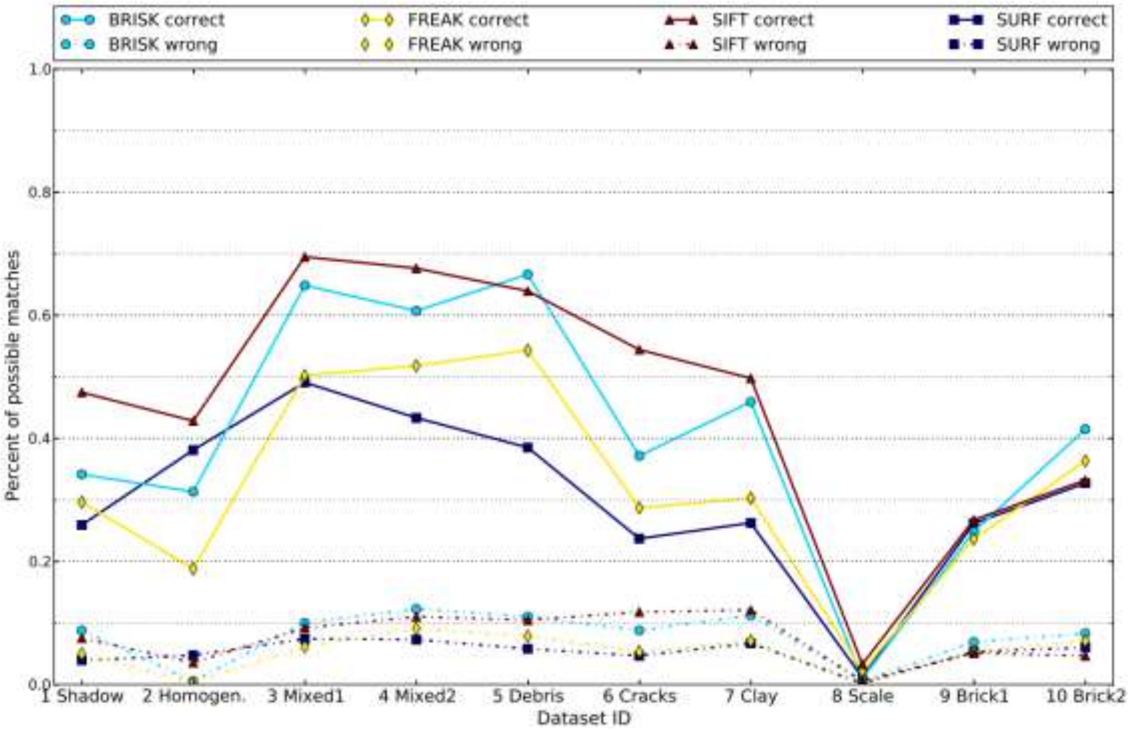

Figure 121: Matches on ORB features computed with AHE with outlier filtering

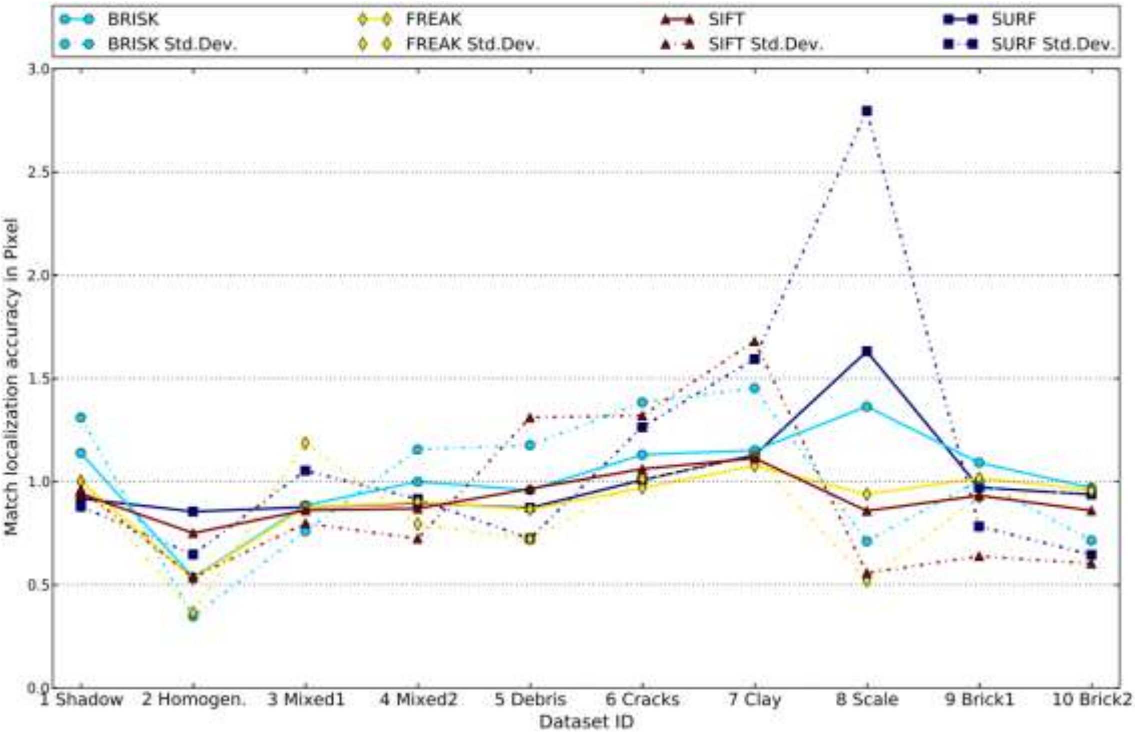

Figure 122: Matching accuracy on ORB features computed with AHE with outlier filtering

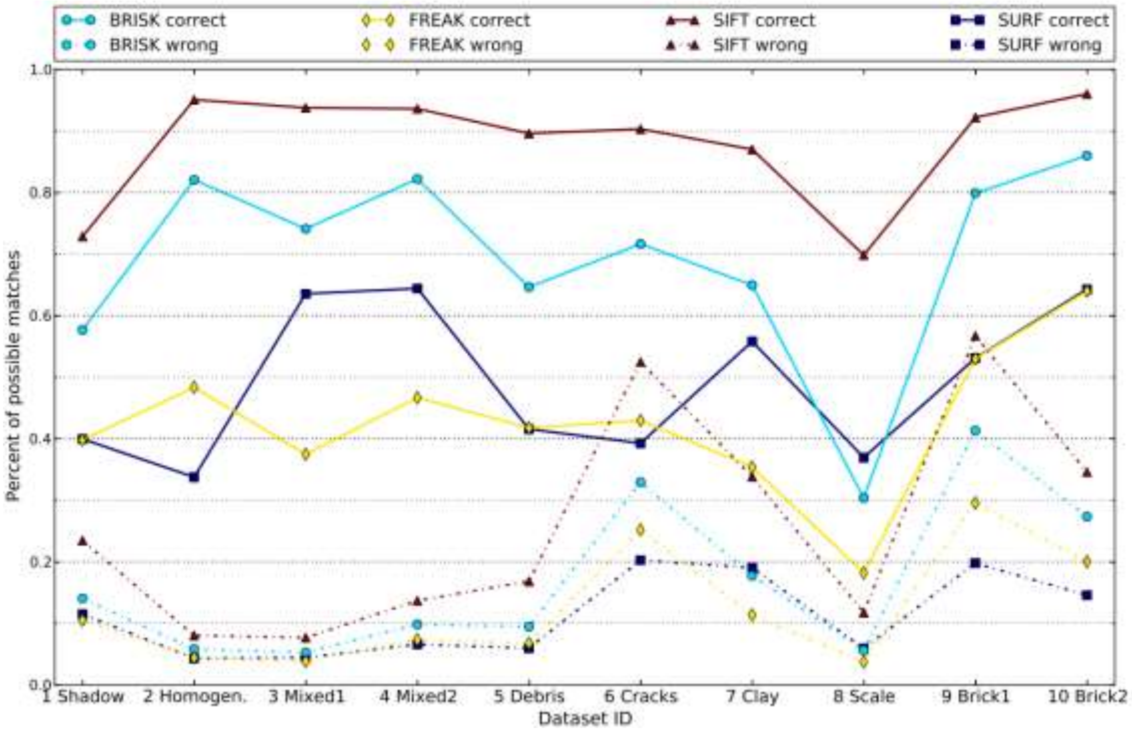

Figure 123: Matches on STAR features

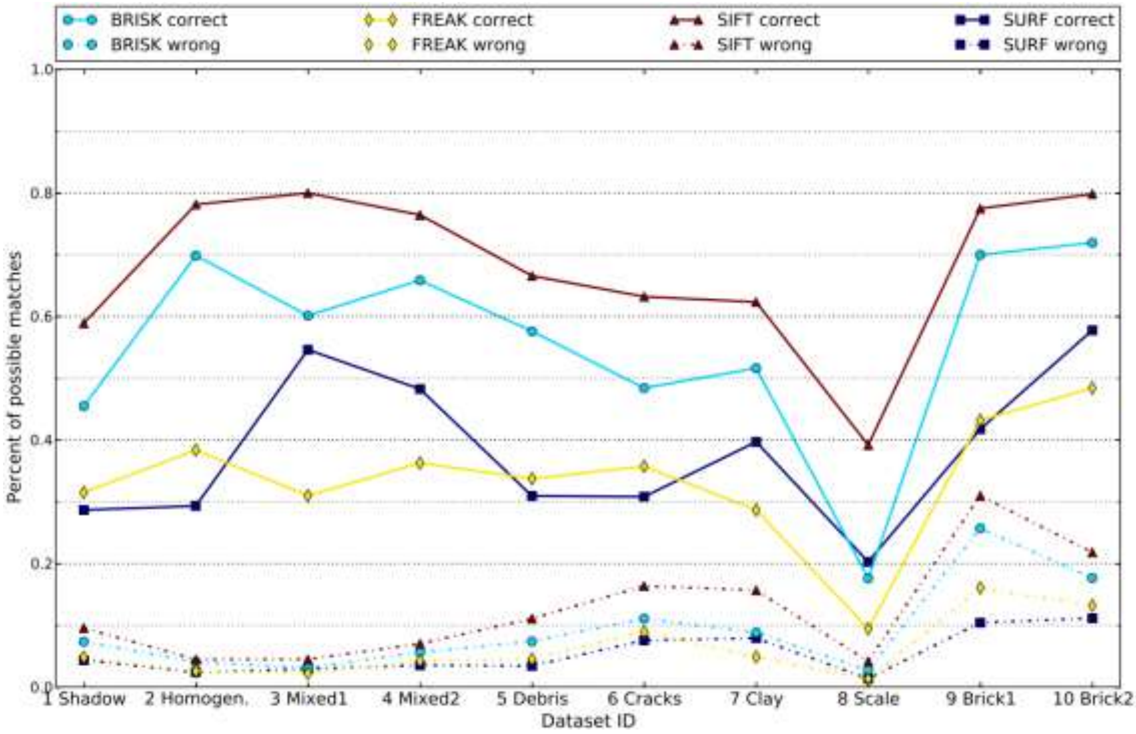

Figure 124: Matches on STAR features with additional outlier filtering

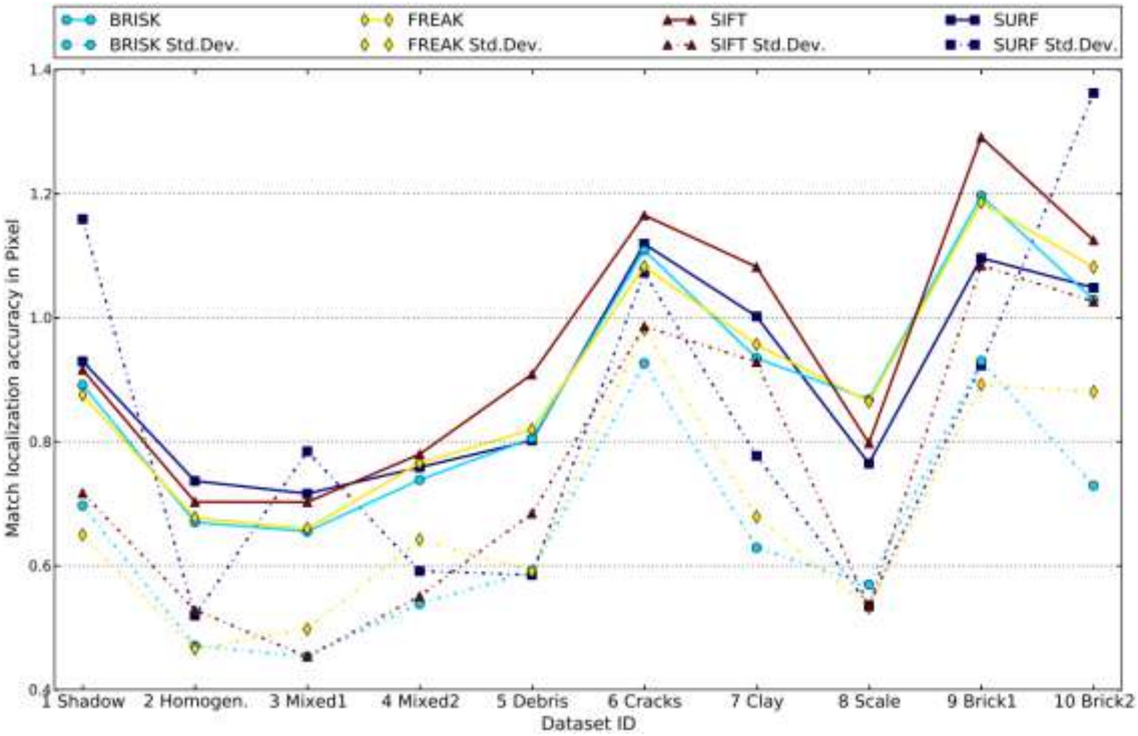

Figure 125: Matching accuracy on STAR features with additional outlier filtering

E. Least squares refinement performance evaluation charts

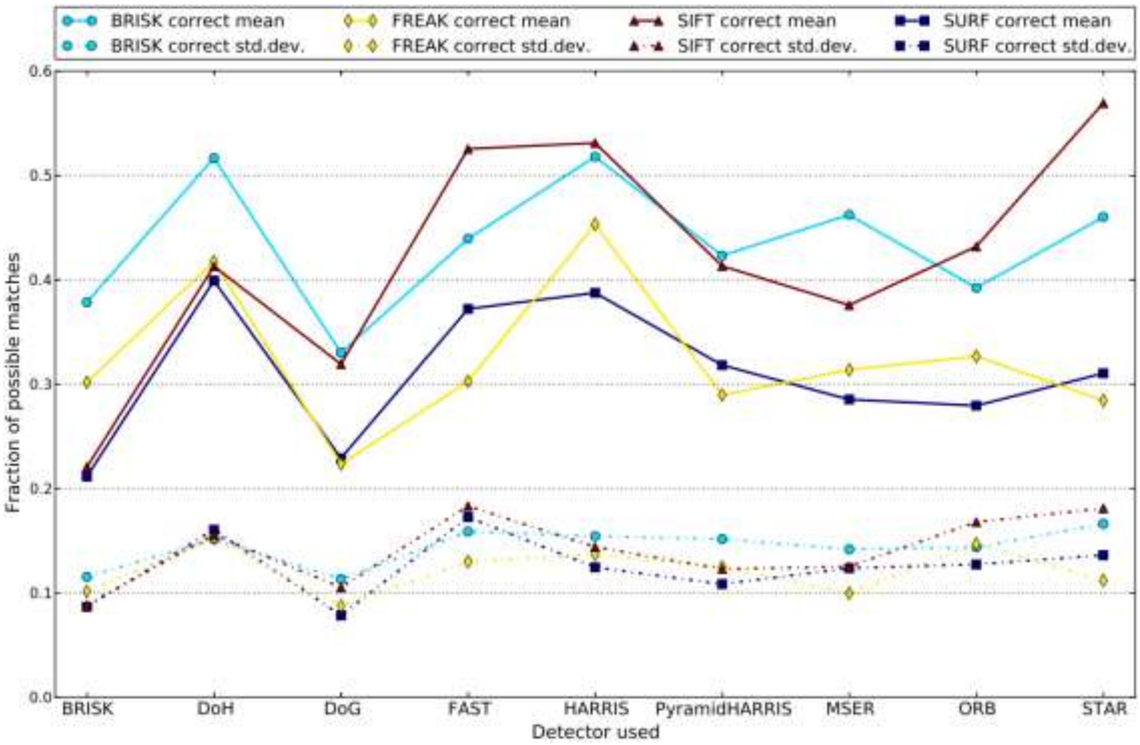

Figure 126: Average fraction of filtered correct matches of possible matches from LSM with a window size of 9 pixels applied to all detectors and descriptors over all data-sets including standard deviation

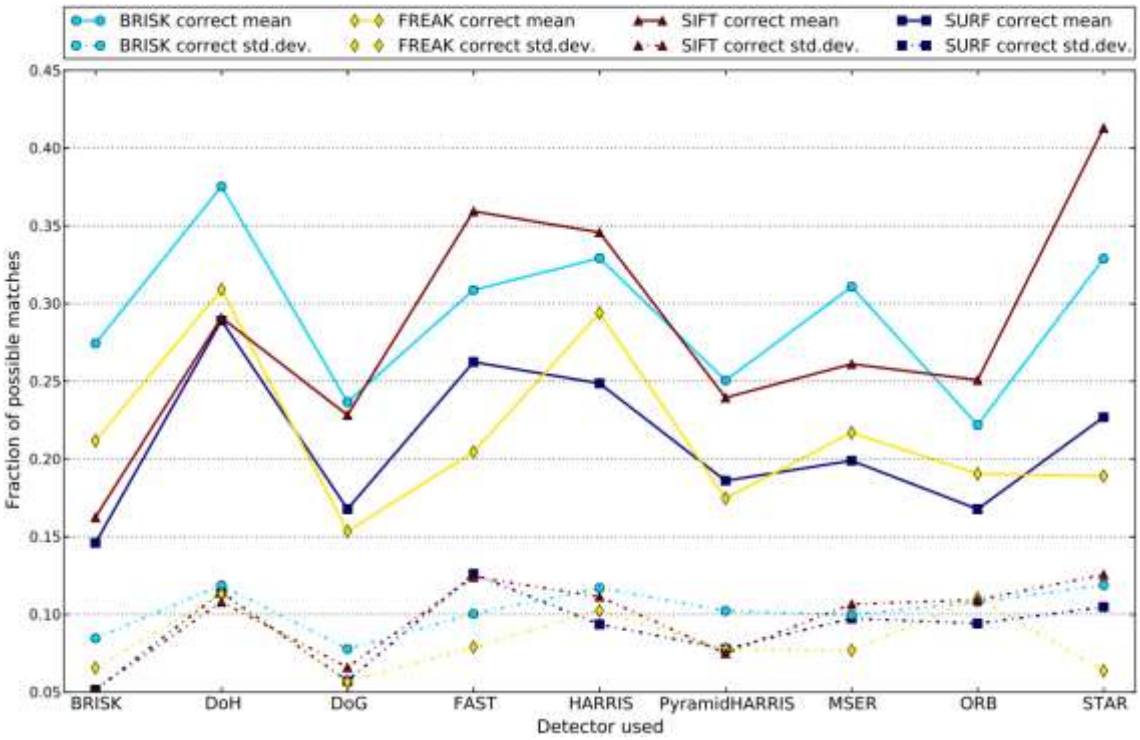

Figure 127: Average fraction of filtered correct matches of possible matches from LSM with a window size of 9 pixels and a back-matching distance of less than 0.1 pixel

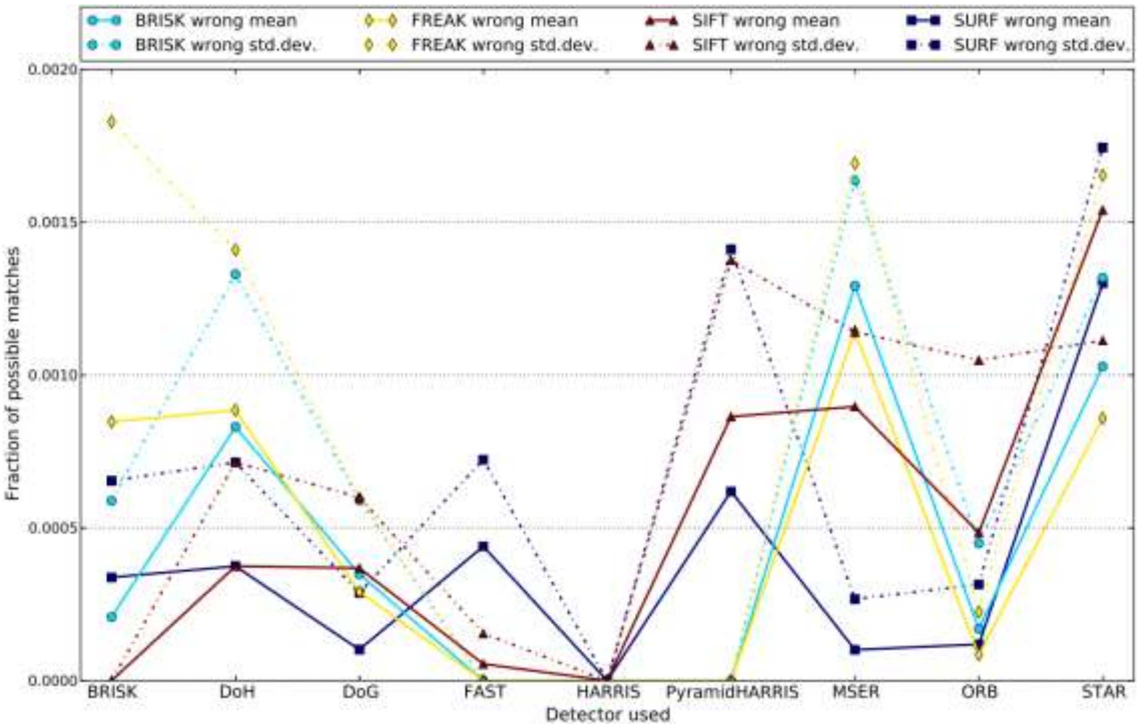

Figure 128: Average fraction of filtered incorrect matches of possible matches after LSM with window size of 9 pixels

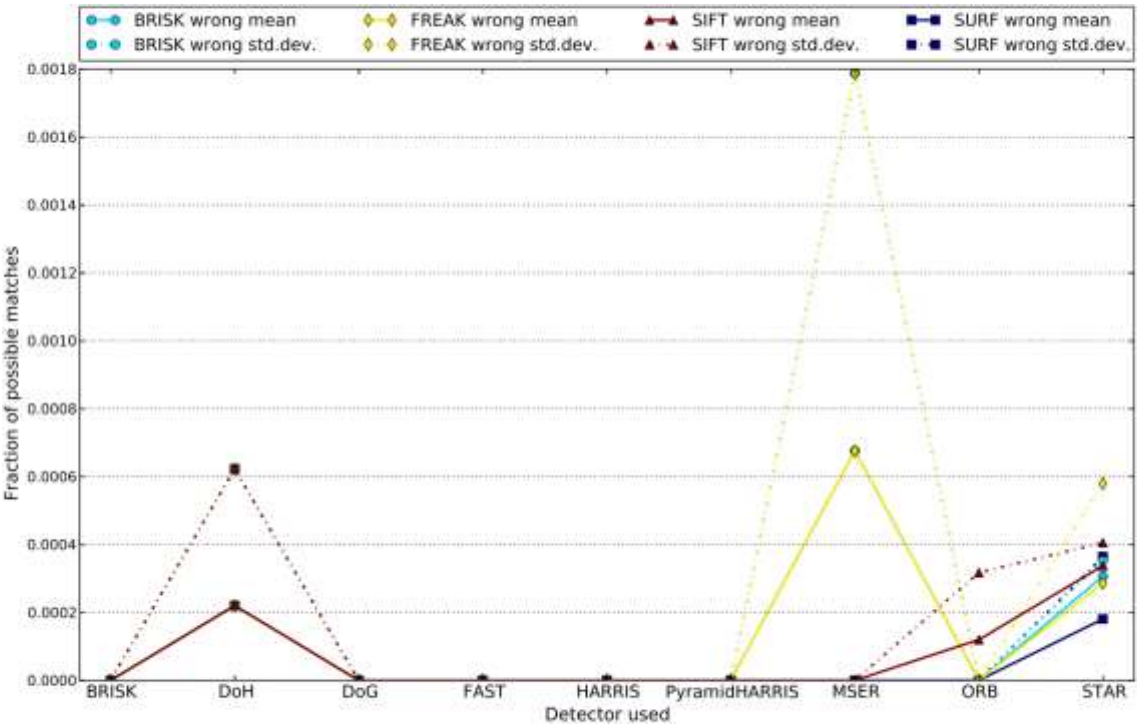

Figure 129: Average fraction of filtered incorrect matches of possible matches after LSM with window size of 9 pixels and a back-matching distance of less than 0.1 pixel

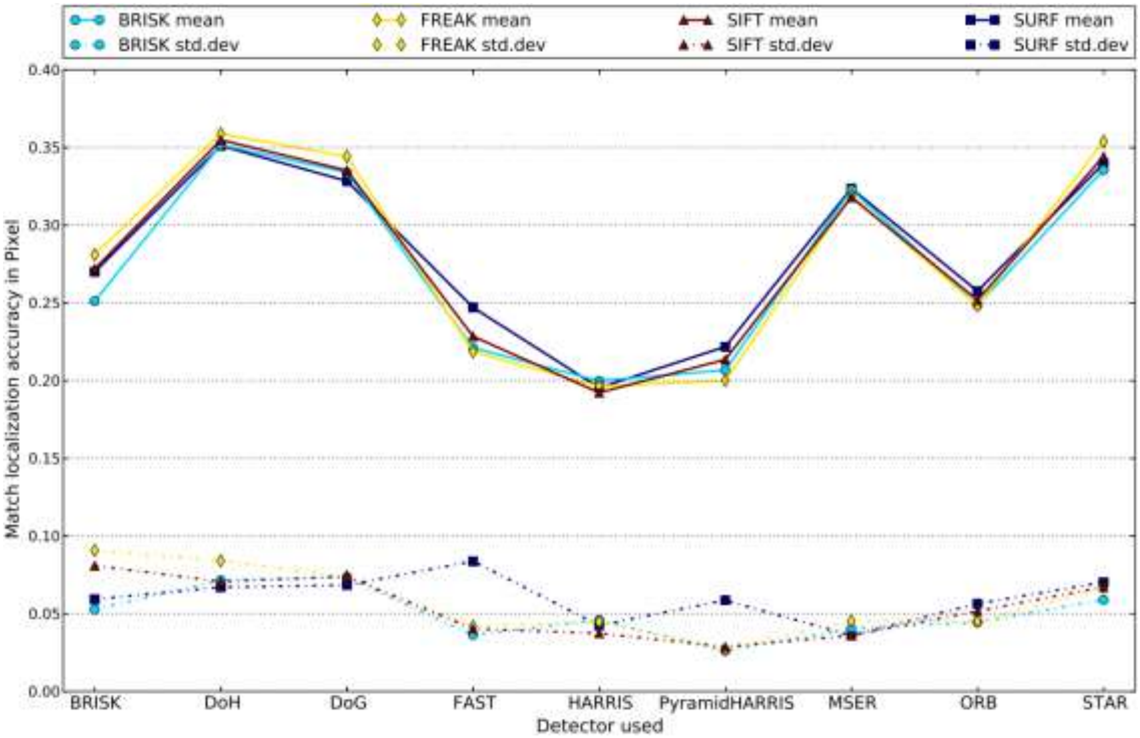

Figure 130: Average LSM accuracy for all detectors and descriptors over averages of all datasets including standard deviation with an LSM window size of 9 pixels

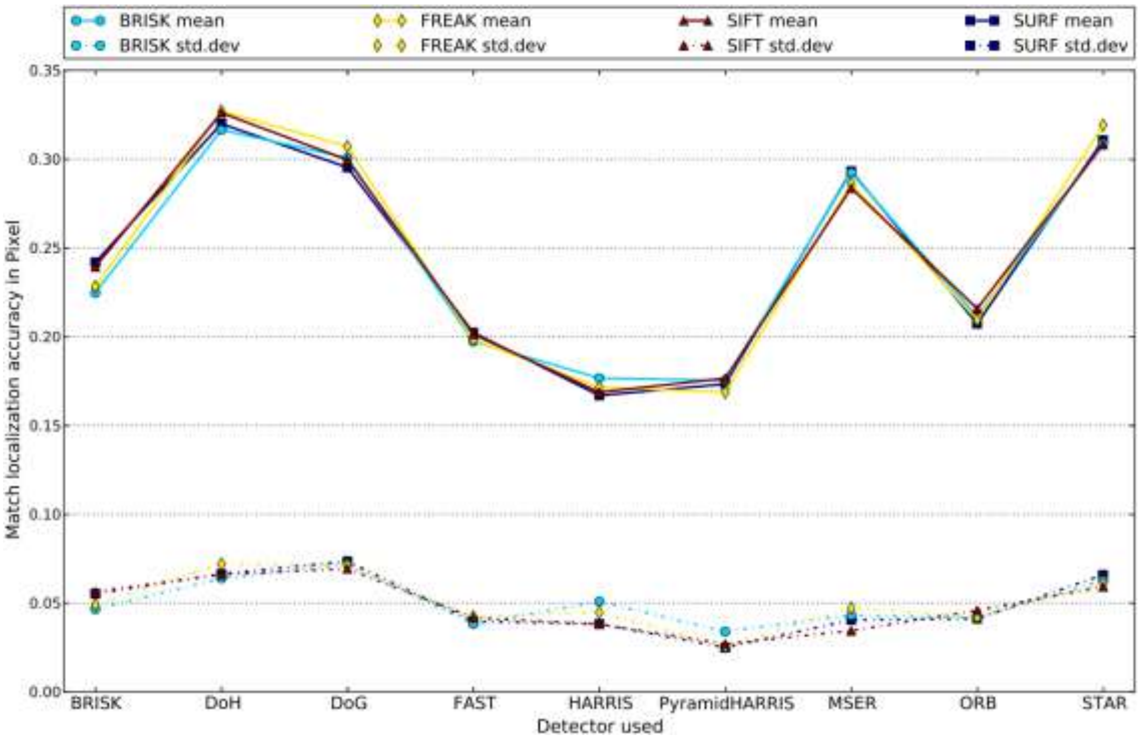

Figure 131: Average LSM accuracy for all detectors and descriptors over averages of all datasets including standard deviation with an LSM window size of 9 pixels and a back-matching distance of less than 0.1 pixel

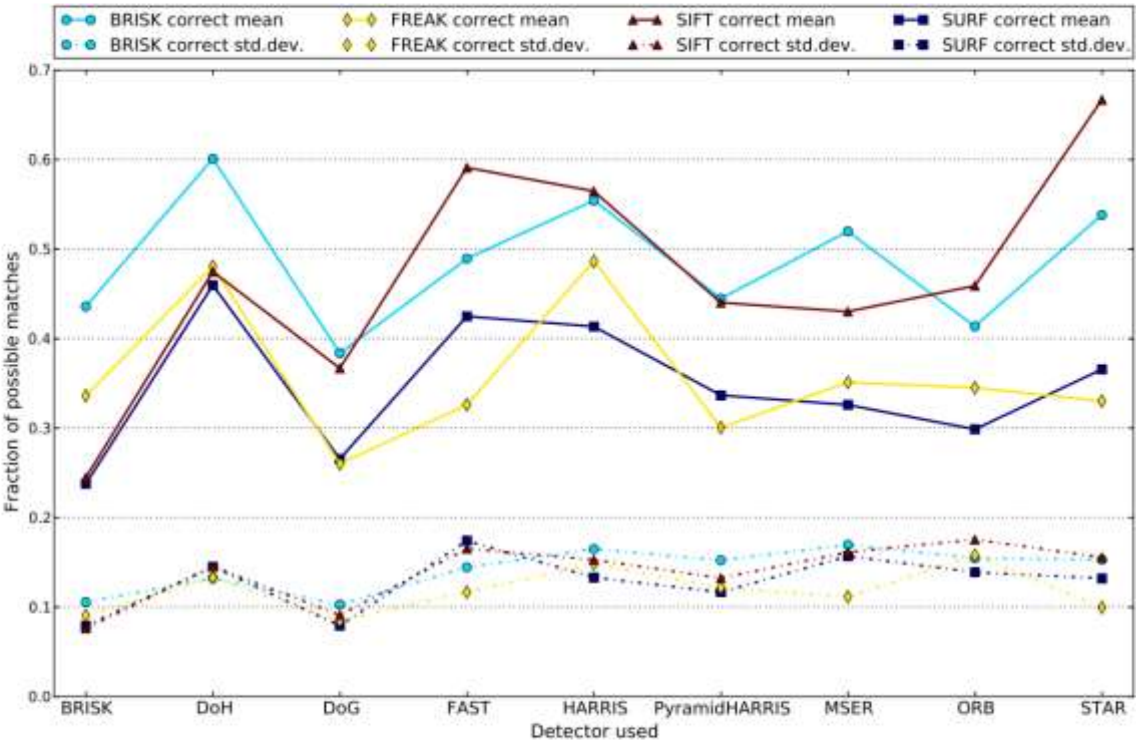

Figure 132: Average fraction of filtered correct matches of possible matches from LSM with a window size of 31 pixels

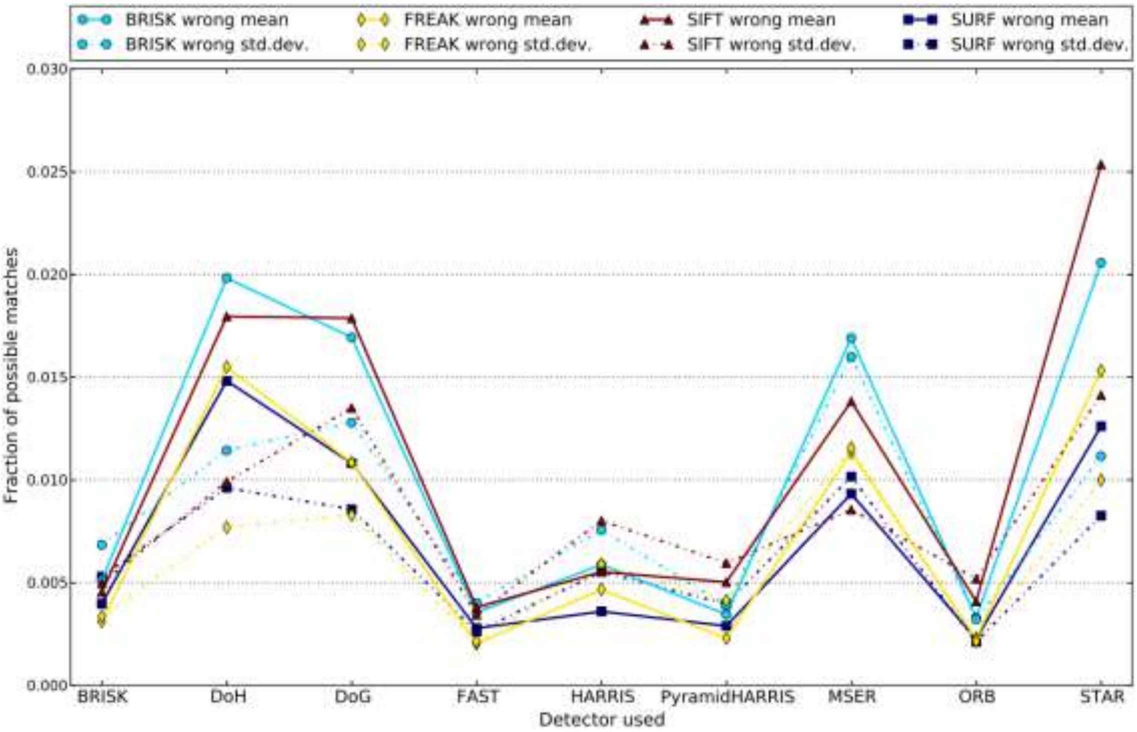

Figure 133: Average fraction of filtered incorrect matches of possible matches after LSM with window size of 31 pixels

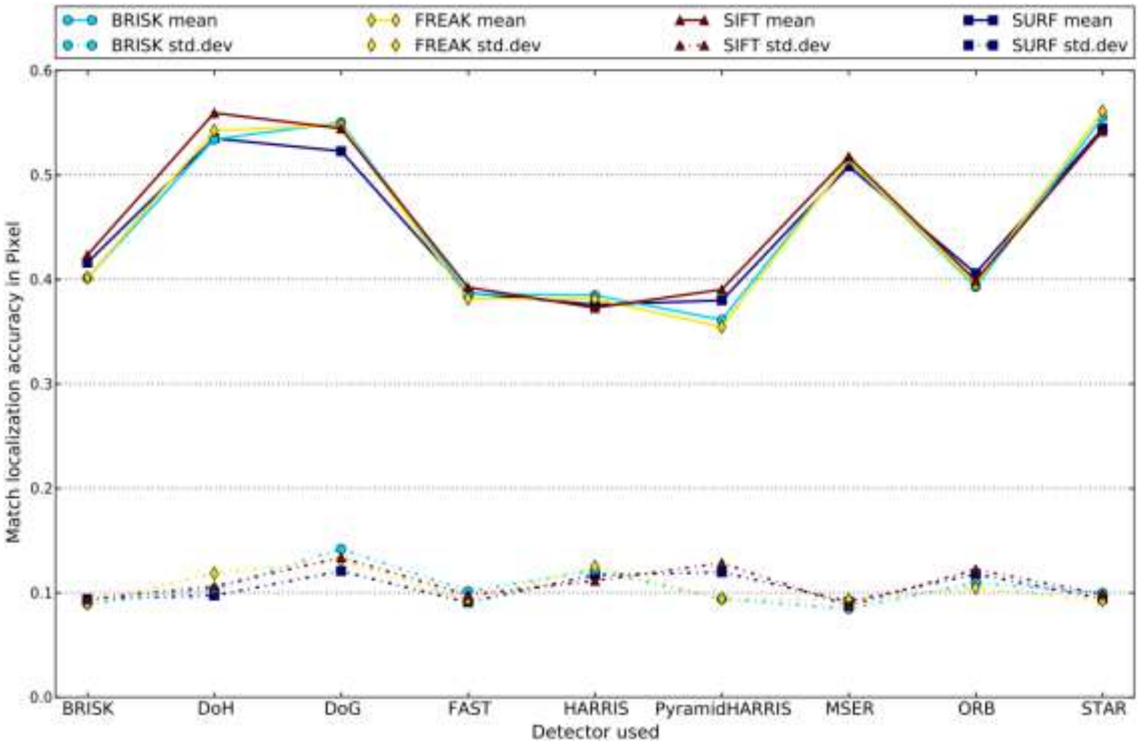

Figure 134: Average LSM accuracy with an LSM window size of 31 pixels

F. Adjustment report (TPS network Pellheim)

```
Errors of coordinates ==============================================================

A,B = Semi-major, semi-minor axis of the confidence ellipse (S = 95.00 %)
Phi = Bearing of the semi-major axis

Point number ----- Position --- Phi ------- A ------- B ---- Height -----------
                     (m)       (gon)       (m)       (m)      (m)
       IATS0        0.0024     33.08     0.0057    0.0020    0.0009
       IATS1        0.0025     47.09     0.0058    0.0024    0.0010
      IATS10        0.0078     40.25     0.0170    0.0090    0.0026
      IATS11        0.0099     64.01     0.0220    0.0106    0.0015
        K1.1        0.0027      0.82     0.0059    0.0031    0.0019
        K1.2        0.0027      0.26     0.0059    0.0031    0.0019
        K1.3        0.0028      1.98     0.0059    0.0037    0.0019
        K1.4        0.0027      1.05     0.0059    0.0031    0.0019
        K2.1        0.0027      0.80     0.0059    0.0031    0.0019
        K2.2        0.0028      2.12     0.0059    0.0037    0.0019
        K2.3        0.0027      0.80     0.0059    0.0031    0.0019
        K2.4        0.0028      2.12     0.0059    0.0037    0.0019
        K3.1        0.0028      1.96     0.0059    0.0037    0.0019
        K3.2        0.0027      1.02     0.0059    0.0031    0.0019
        K3.3        0.0028      1.94     0.0059    0.0037    0.0019
        K3.4        0.0027      0.99     0.0059    0.0031    0.0019
        K4.1        0.0027      0.84     0.0059    0.0031    0.0019
        K4.2        0.0028      2.15     0.0059    0.0037    0.0019
        K4.3        0.0027      0.85     0.0059    0.0031    0.0019
        K4.4        0.0027      1.07     0.0059    0.0031    0.0019
          LS        0.0043    180.77     0.0103    0.0021    0.0015
        MS50        0.0023    199.84     0.0055    0.0011    0.0008
          T1        0.0027      5.27     0.0060    0.0029    0.0008
          T2        0.0033     19.52     0.0075    0.0031    0.0014
          T3        0.0031    196.83     0.0071    0.0031    0.0014
          T4        0.0025      2.83     0.0055    0.0029    0.0008
     TPS1200        0.0024    133.06     0.0050    0.0030    0.0009
```

G. Adjustment report (TPS network Weinzödl)

```
Errors of coordinates =========================================================

A,B = Semi-major, semi-minor axis of the confidence ellipse (S = 95.00 %)
Phi = Bearing of the semi-major axis

Point number ----- Position --- Phi ------- A ------- B ---- Height -----------
                     (m)        (gon)    (m)      (m)        (m)
              1       0.0059   112.50   0.0108   0.0094     0.0098
              3       0.0060   160.54   0.0113   0.0096     0.0097
              4       0.0060     8.81   0.0111   0.0096     0.0097
              5       0.0055    53.07   0.0101   0.0089     0.0097
              6       0.0061   153.53   0.0114   0.0098     0.0098
              7       0.0065   156.72   0.0116   0.0103     0.0102
              8       0.0040   155.00   0.0074   0.0063     0.0062
              9       0.0061   158.59   0.0114   0.0098     0.0097
             10       0.0069   164.12   0.0138   0.0118     0.0113
             11       0.0066   166.28   0.0123   0.0105     0.0104
             12       0.0045   167.93   0.0083   0.0071     0.0071
             13       0.0035   168.10   0.0067   0.0055     0.0054
             14       0.0045   170.22   0.0083   0.0071     0.0071
          IATS0       0.0045   167.55   0.0102   0.0044     0.0043
          IATS1       0.0059    59.61   0.0109   0.0049     0.0047
```

H. Pointmatching tool command line options

```
---------------------------------------------------------------------------------------
   mandatory parameters:
    -l leftinputfile -r rightinputfile
   optional parameters :
    -thres threshold feature detection response threshold          (Default = 0.001)
    -u run in rotation variant mode                                (Default = false)
    -o octaves number of octaves to calculate                      (Default = 3)
    -i intervals number of intervals per octave                    (Default = 4)
    -is sampleStep sampling step in pixel                          (Default = 2)
    -Detector det The interest point detector to be used, options are
           FAST, ORB, BRISK, HARRIS, PyramidHARRIS, SURF, SIFT     (Default SURF)
    -Descriptor desc The interest point descriptor to be used, options are
           ORB, BRISK, FREAK, SIFT and SURF                        (Default SURF)
    -Matcher The matching method to be used, options are
           FlannBased, BruteForce, BruteForce-Hamming              (Default BruteForce)
    -loadOrientation L.cal R.cal dMaxEpipolarLineDeviation
           Use known image orientation to restrict matching        (Default = false)
    -matchRatio ratio similarity ration for matching               (Default = 0.65)
    -he global histogram 191qualization                            (Default = false)
    -ahe nTilesX nTilesY fContrastLimit if no additional parameter specified: 8 8 none
      (contrast limited) adaptive local histogram equal.           (Default = false)
    -outList list path to save output list                         (Default = Corresp.ip)
    -savePts save index scale & response additionally for points   (Default = false)
    -outImgCorrespList list path to save corresp. List             (Default = ImgCorr.ip)
    -descriptors img1.desc img2.desc feature vectors of matches    (Default = no save)
    -maxPoints 100 keep only the best N matches                    (Default = 100000)
    -outImg img1 img2 path to save output images                   (Default = no save)
    -lines 1 draw lines btw. stereo-points with given width [px]   (Default = false)
    -allPoints draw all found points                               (Default = false)
    -windows draw descriptor windows for points                    (Default = false)
    -montage img path of images stereo montage                     (Default = no save)
    -divide 4 divide into tiles. pts are detected separately       (Default = none)
    -outlierDetect 20 fundamental matrix for outlier detection     (Default = none)
           plus restrictivity of RANSAC [px] if no number given LMEDS will be used
    -h use Homography for outlier detection (with outlierDetect)   (Default = false)
    -filterEpipolarSpikes nNeighbors fMaxDeviationFactor filter points by comparing
      horizontal position on epipolar line with nNeighbors. Allow a deviation from mean
      1(+/-)fMaxDeviationFactor. Defaults: nNeighbors = 5, fMaxDeviationFactor = 0.2
    -filterDisparityDistanceAndAngle KNN fMaxDistDev fMaxAngle filter by deviation of
      disparity distance or angle within KNN neighbors mean: 1(+/-)DistDev & Angle[gon]
      defaults: KNN = 5, fMaxDistDev = 0.2 fMaxAngle = 10.0
    -addMatches add points consistent w. outlier model as matches  (Default = false)
    -multipleHomographies calculate multiple homographies with outliers of previous ones
    -MatchCorners nCornerSize estimate homography from first matching run, then match
      patches of given sidelength [px] around corners of calculated image intersection a
      second time. No patch size given, imagesize/3 will be used   (Default = false)
    -RANSAConly corresp.ip load points and perform RANSAC          (Default = no)
    -batch write image correspondences                             (Default = no)
    -loadPts corresp.ip load ip coordinates from file              (Default = no)
    -loadDescs img1.desc img2.desc load descriptors from file. if only 1 is given it
      will be matched with image2. format as -descriptors          (Default = no)
    -templateBasedOutlierDetection 4 10 0.5 scale up factor of batches, batch size
      factor, accuracy                                             (Default = no)
    -mask strMaskImage pixels with value 0 in mask will be excluded
    -LSM enter (all in one line - surrounded with quotes):         (Default = no)
           "windowsize iterations converge numparams BMdist bLSMOnly"
```