

MASARYK UNIVERSITY
FACULTY OF INFORMATICS



Imaging Distortion Correction in Panorama Stitching in SEM

MASTER'S THESIS

Bc. Ján Bella

Brno, Spring 2017

MASARYK UNIVERSITY
FACULTY OF INFORMATICS



Imaging Distortion Correction in Panorama Stitching in SEM

MASTER'S THESIS

Bc. Ján Bella

Brno, Spring 2017

Replace this page with a copy of the official signed thesis assignment and the copy of the Statement of an Author.

Declaration

Hereby I declare that this paper is my original authorial work, which I have worked out on my own. All sources, references, and literature used or excerpted during elaboration of this work are properly cited and listed in complete reference to the due source.

Bc. Ján Bella

Advisor: doc. RNDr. Pavel Matula Ph.D.

Acknowledgement

I would like to thank Pavel Matula for his patience and precious advice, Vojtěch Filip for providing the data and answering all my questions related to operation of Scanning Electron Microscopes, and my family for the lifetime support.

Abstract

The aim of the thesis was to define a distortion model of the imaging process of Scanning Electron Microscope, estimate the parameters of the model given image tiles and knowledge of the overlapping regions, and apply the model while stitching image tiles into a panorama. In comparison to standard distortion correction methods, the resultant panoramas should correspond more accurately to the scanned objects in terms of their shapes and sizes.

Keywords

scanning electron microscopy, image registration, panorama stitching, distortion correction, feature detection, feature matching, constrained non-linear optimization, polynomial optimization, simulated annealing

Contents

1	Introduction	1
2	Scanning electron microscope	3
2.1	<i>Layout</i>	3
2.2	<i>Distortion</i>	6
3	Panorama stitching in digital photography	7
3.1	<i>Lens distortion in digital photography</i>	7
3.2	<i>Stitching</i>	8
4	Keypoint detection and matching	11
4.1	<i>Scale Invariant Feature Transform overview</i>	12
4.2	<i>Oriented FAST and Rotated BRIEF</i>	14
4.3	<i>Keypoint matching</i>	15
5	Mathematical description	19
5.1	<i>Coordinate systems and stitching</i>	19
5.2	<i>The global distortion correction transform</i>	19
5.3	<i>Stage shifts and stitching</i>	24
5.4	<i>The optimization cost function</i>	25
5.5	<i>The optimization constraints</i>	29
5.6	<i>Properties and considerations regarding the optimization cost function</i>	31
5.7	<i>Considerations regarding stitching</i>	33
6	Optimization	35
6.1	<i>Optimization via semi-definite programming relaxations</i>	35
6.2	<i>Simulated annealing</i>	36
6.3	<i>Local methods</i>	39
7	Experiments and results	41
7.1	<i>Suitability of SIFT and ORB keypoints</i>	43
7.2	<i>Stitching by simple tile alignment</i>	44
7.3	<i>The optimization results</i>	45
7.3.1	Results obtained by enforcing the expected bounds	45
7.3.2	Best results by using alternative bounds	46
7.3.3	Comparison of model T_1^{-1} vs. T_2^{-1}	47
7.3.4	One-step vs. two-step optimization	49
7.3.5	Normality of stage shift errors	50
7.4	<i>Stability of solutions</i>	50
7.5	<i>Optical distortion estimation from a subset of tiles</i>	51

8	Discussion	53
8.1	<i>Distortion parameters</i>	53
8.2	<i>Precision of stitched panoramas</i>	53
8.3	<i>Note on quality evaluation</i>	54
9	Conclusion	55
	Index	57
	Bibliography	57
A	Stitched panoramas	61
B	Q-Q plots for stage shift errors	67
C	Brief description of the application framework	69

1 Introduction

A common task in electron microscopy is to acquire a scan of a specimen at large magnification and high resolution. However, the specimen might be too large to fit in one image at the given magnification, so it is necessary to obtain several images of different parts of the specimen (image tiles) and to combine them (stitch them) together to obtain the final image.

In a perfect world, if the images acquired by microscope were not affected by any distortions, it would be sufficient to align the tiles next to one another. Yet in reality, the acquisition device (microscope) is not perfect. Due to the physical limitations of the device construction, the electron optics gives rise to a small distortion, which can be characterized mostly as a combination of rotation, magnification and non-orthogonality transforms. Moreover, when the specimen stage shifts between two positions, a small random positioning error is introduced. As a result, if one simply aligns the neighbouring tiles next to each other, the resulting panorama will contain visible discontinuities along the edges or repeated patterns close to the borders of the tiles. Therefore, a global distortion correction transform needs to be applied prior to the tile alignment. The tile alignment should correspond to specimen stage positioning errors.

The aim of this thesis is to mathematically describe the problem of distortion correction and panorama stitching for images acquired by the Scanning Electron Microscope (SEM), find parameters of the model and apply it to stitch panoramas without visual defects. During the search for the unknown parameters of the model, the estimates of the intervals for each parameter are taken into the account. As a consequence, the resulting panoramas should be metrologically precise, which means that all objects in the panorama should have the correct shape and size.

Until April 2017, the author of this thesis is not aware of any other work trying to solve the same problem. The closest approach was described by Kaynig et al. in [1]. However, there are plural differences between this work and the work of Kaynig et al. First of all, they focus on images acquired by the Transmission Electron Microscope (TEM), which has a distortion dominated by different types of transforms.

1. INTRODUCTION

Also, they estimate the general affine transform followed by a non-linear transform, but in this work the global transform must be strictly composed of rotation, scaling and non-orthogonality. Secondly, the authors do not consider the stage shift imprecisions. Lastly, they expect their algorithm to work on the tiles with at least 50% overlap, but the approach described here is shown to work with datasets with much smaller overlaps (down to 5%).

The thesis is structured into eight main chapters. The first chapter describes the SEM optics to understand where the specific types of distortions are generated. The second chapter explains the differences between this work and panorama stitching in digital photography. The third chapter outlines the method used for the keypoint detection and matching in order to establish stable and correct correspondences between the keypoints. The fifth chapter contains the description of the mathematical model. The sixth chapter suggests optimization methods suitable to find the parameters of the model for the given dataset. Finally, the seventh and eight chapter evaluate and discuss the outcomes.

2 Scanning electron microscope

Scanning electron microscope (SEM) is a device, which "permits the observation and characterization of heterogeneous organic and inorganic materials on a nanometre (nm) to micrometre (μm) scale" [2].

As it is hinted by the device name, the image is formed by scanning (rastering) the surface of a specimen line-by-line using a ray of electrons. The main advantage over Transmission Electron Microscopes (TEM) is the possibility to observe thick specimens, because the electrons do not need to be transmitted through them. The advantage over common light microscopes is the possibility to achieve more detail.

The concept of the SEM was first described by Knoll in 1935 [3], but the device was first constructed by Zyworkin in 1942. The following works by Oatley, Smith, and Everhart and Thornley proposed improvements to Zyworkin's prototype, which resulted in the first commercially successful instrument called SEM V, developed by Pease and Nixon in 1963 [2] [4]. Since then, the main microscope parts remain similar with minor advancements.

2.1 Layout

The microscope construction can be divided into three main parts:

- the specimen chamber,
- the electron optics system,
- the acquisition system, electronics.

The specimen chamber contains the stage, on top of which the specimen is placed. The stage is able to move, generally in five directions: translation in x , y and z axes, pitch, and yaw (so called 5-axis SEM stage). The optical system is placed above the specimen and the electron ray is generated and focused there. The acquisition system consists of a set of detectors which collect a specific type of signal (for example the back-scattered electrons) and convert the signal to pixel intensities. All the components operate in high vacuum, because gas atmosphere would attenuate the electron beam and the scan could fail.

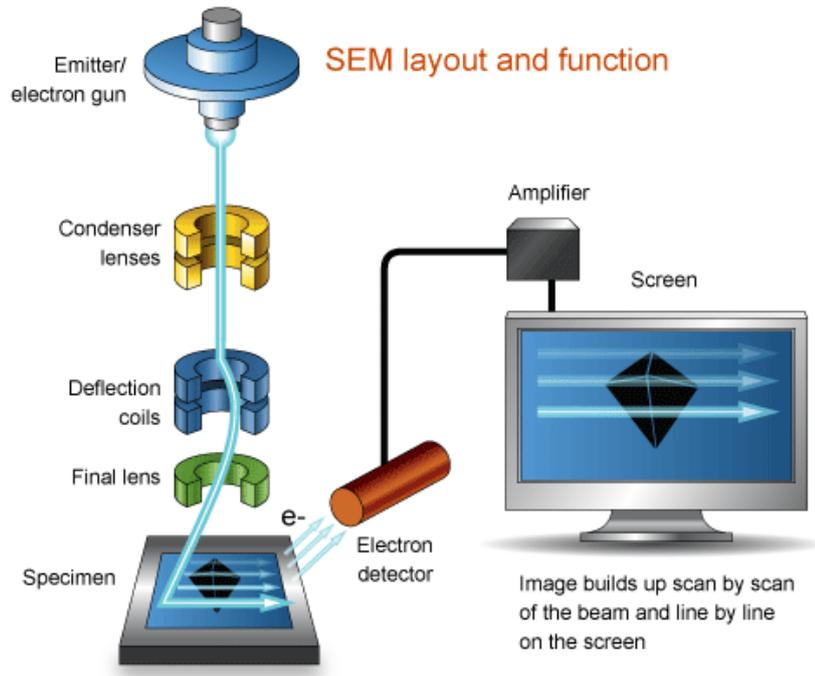


Figure 2.1: Layout of the SEM. Adopted from [5].

The electron optics is the most important part for this work. As it is illustrated in the SEM schema in Figure 2.1, it is composed of electron gun, condenser lens, deflection system, and objective lens.

The electron gun is a component which generates the electron beam by accelerating the electrons in a specific range of energy. The amount of energy given to the electrons affects the brightness of the final image. The raw beam itself needs to be demagnified and skewed to hit a set of discrete positions of the specimen to enable the line-by-line pixel-wise scanning.

Demagnification is an operation which reduces the diameter of the ray. It is provided by the first electromagnetic lens under the gun, which is called the condenser lens (also referred to as the electron lens). If demagnification lens was not included, the acquired image would be either blurry, or one would have to use a very small resolution.

The deflection (scanning) coils are the second in the stack. Their purpose is to deflect the beam over a set of discrete positions, so that at each position, the corresponding pixel intensity can be recorded. The

operation of the deflection coils affects magnification, as it depends on the step size at which the specimen is scanned.

The last component in the simple scheme is the final (or objective) lens, which serves for a similar purpose as the condenser lens: focus the beam onto the sample. When the beam is generated, it is naturally diverging to all directions. Despite the condenser lens focuses the beam, its convergence is inverted before it hits the sample, so there needs to be another lens to obtain the converging ray at the end of the optical system.

When the beam hits the specimen, it interacts with it and the electrons are scattered within, either elastically (interaction with atomic nuclei), or inelastically (interaction with electrons of atoms). The electrons that are ejected from the specimen as a result of inelastic scattering are called secondary electrons and can be recorded to form an image of the surface of the scanned object (due to the low energy they carry, they are able to escape only from the top levels of the specimen). Inelastically scattered electrons carry more energy, have larger interaction volume and form an image which carries the information about the material composition of the specimen [6].

When acquiring the panoramic image, the above described procedure must be repeated for each tile. In between the scans, the stage comes into operation, as it must shift from one position to another.

The stage shifting is not formally defined in the literature, so it depends on the computer software for the microscope control. Both size of the overlaps, and the scan orders are options, which need to be defined by the microscope operator. Generally, meander scanning is a good option, because it requires the least stage shifts, but the standard forward or backward line scans are possible. The size of the overlap affects the quality of the stitching. In larger overlaps, the correlation of pixel intensities is more stable, or the number of keypoint correspondences in the overlapping image parts is larger. However, it is also affected by the character of the specimen itself.

Further information on the layout, operation of SEM and advanced topics can be found in [2].

2.2 Distortion

Optical system distortion From a technical point of view, the lenses are composed of one or more coils. At deflection coils, the scanning is realised by exciting a pair of coils, each with different energy. To avoid the distortion, the coils excitation would have to be so precise, that every change of power in both directions would produce the same step. However, because the mechanical accuracy is limited, as well as the accuracy of the electronic components, it is hardly feasible to achieve such behaviour. If the step size is different in the x and y directions, it can be perceived in form of zoom distortion. At the same time, non-orthogonality is produced, for example if a ray is deflected a bit more in one of the $x - y$ directions at each step of the rastering process. The electromagnetic field at the electron lens pushes the ray into its centre. Naturally, the ray takes the spiral trajectory down the coil. Especially at low magnifications, the rotation of the final image can be seen [7]. Even though it affects both condenser and objective lenses, when modelling the global distortion, it can be assumed that the rotation appears after the zoom and non-orthogonality. Distortion generated by the optical system is identical for each tile captured with the same settings. It will be further referred as the global distortion.

Distortion by stage shifts The positioning of the specimen stage has only a limited mechanical precision. The difference between the requested position and the position at which the specimen stage stopped can be observed in form of a translation distortion, which is different for each pair of stage shift. Distant positioning might generate bigger distortion compared to adjacent shifts.

3 Panorama stitching in digital photography

Probably the first question that one asks themselves is whether it is necessary to develop a new method and whether it is not sufficient to use any of the available tools for digital photography. Although some concepts are the same, generally these are two different problems.

Digital cameras (both point-and-shoot and digital single-lens reflex) use light optics. The light is focused by a set of glass lenses to the image sensor, which converts photons to electric charge (scheme illustrated in Figure 3.1). Further details on the image acquisition process in light cameras are omitted as they are not important for this work; a curious reader might find more information in [8].

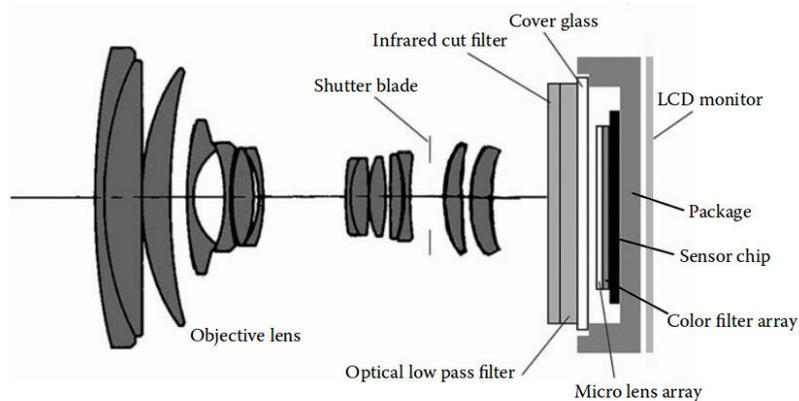


Figure 3.1: Point-and-shoot camera layout. Adopted from [8].

3.1 Lens distortion in digital photography

Glass lenses produce different type of distortion compared to the one described before. In cameras, it has the form of barrel, pincushion or moustache distortion; such distortions cause straight lines to appear as curves (see Figure 3.2). It is connected to spherical aberration, a phenomena at which the image of the object appears to be out of focus due to the spherical shape of the lens. To avoid spherical aberration, an aperture stop is placed next to the lens to block some of the rays

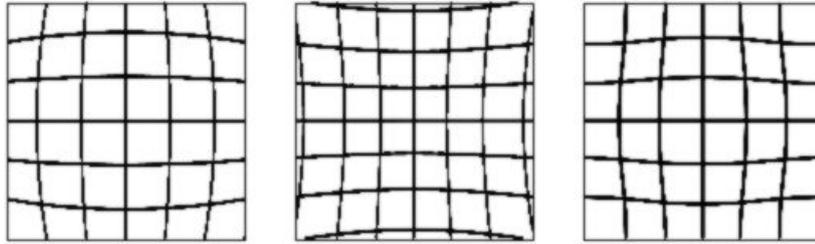


Figure 3.2: Illustration of barrel, pincushion and moustache distortion.

and minimize the circle of confusion. However, when the chief ray (ray passing through the centre of the aperture) refracts, the image becomes distorted [9].

A non-linear model has to be considered to correct the global distortion. Camera calibration can be performed in order to find the correction transform, usually by photographing an image containing straight lines, for example a chequerboard. The calibration transform can be applied to captured images later in the post-processing.

3.2 Stitching

The primary goal in digital photography is to acquire a beautiful image without any visual deformations. In case of panorama stitching, the primary focus is on registering the tiles and hiding the seam. Little effort if any is spent on estimating and correcting the exact distortion generated by the optical system. Instead, image tiles are simply aligned one to another in order to create the resulting panoramic image. Tile-by-tile alignment is also a common practice in panorama stitching in microscopy. There are two general approaches to find the correct tile alignment: pixel-based methods and keypoint-based methods.

Keypoint based methods first find a set of keypoints in each tile and match them. After the first (reference) tile is positioned, other tiles aligned to it by applying homographies. A homography is generally a projective transform mapping a set of keypoints found in one tile to a set of corresponding keypoints in the other tile. Each tile can be aligned to the rest of already stitched panorama by a different transform – the homography does not have to be shared by all pairs of tiles. In this

work, keypoints are utilized to determine and correct the distortion generated by SEM. Further information on keypoint detection and matching are provided in the next chapter.

Pixel based approaches work directly with pixel intensities. The aim is to find a translation that for example minimizes the squared differences between the intensities in the overlapping regions, or maximizes their correlation. The main advantage of this approach is its simplicity, because no prior computation (unlike in case of keypoint detection methods) is necessary. The main disadvantage is that it is only possible to find the best translation between two neighbouring tiles, which might not hide the seam perfectly. While the optical system in microscopy is stable and the specimen stage shifts in a plane in between tile scans, in digital photography the photographer usually moves the camera while the scene is static. As a consequence, the coordinates of such panorama are not flat. For example, if a scene was captured by standing at the same spot and rotating the camera, the acquired images copy the shape of the inner surface of a cylinder. If the camera motion is known, it is possible to apply a different coordinate mapping to each of the image tiles, which enables the tile alignment by translations.

The best translation between the tiles can be found either hierarchically or by using one of the direct methods. Hierarchical methods try and evaluate different alignments in each iteration in order to find the best relative positioning of the tiles, but such approach can be inefficient and tedious. The phase correlation method is a representative of direct pixel-based methods. It uses the fact that a translation in the spatial domain corresponds to a shift of phase in the frequency domain. Moreover, correlation can be efficiently computed in the frequency domain thanks to the convolution theorem: a convolution in the spatial domain corresponds to pixel-wise multiplication in the frequency domain; and correlation is nothing else but the convolution of an image with a reverse kernel. Therefore, after transforming both images to the frequency domain, the cross-power spectrum is computed by taking point-wise multiplications between the first frequency image and the complex conjugate of the second frequency image, normalized by the magnitudes. Normalized cross-correlation matrix is obtained by applying the inverse Fourier transform to the cross-power

spectrum. The peak in the cross-correlation matrix corresponds to the shift between the images.

After having the image tiles properly registered, the last step is to compose the final panoramic image. Another coordinate transform can be utilized to map the panorama to a desired composition surface, some pixels might be cropped; however, the most important part of the composition is blending in the seams to handle moving objects appearing only in one of the tiles, and to compensate different exposure to hide the seam. Because no attention is paid to blending in this work, please find more information in *Image Alignment and Stitching: A Tutorial* [10]. The paper provides a great overview of available methods for all steps of panorama stitching in digital photography, as well as many references to further sources.

In this work, it is desired to visualise the misalignments in the overlaps and therefore each pixel in an overlap is set to the maximum of the intensities of overlapping images.

4 Keypoint detection and matching

Keypoint-based methods are an alternative to pixel-based methods in panorama stitching. Pairs of corresponding pixels in the neighbouring images enable the computation of homographies required for tile alignment. In this work, selection of keypoints and a faultless matching are essential for computation of the global distortion correction transform and determination the stage positioning errors.

A keypoint can be any important point in the image. It is desired for a keypoint of an object to be detectable regardless of the zoom, rotation or point of view of the camera (as long as it is visible). Keypoints are often accompanied by descriptors which usually carry information about keypoints' neighbourhood. Applying an affine transform, or adding noise to image should not affect the descriptors, so that keypoint matching between two images was be possible.

There are many keypoint detection methods to choose from. Harris corner detector and its derivatives are a basic solution; yet they have a problem of not being invariant to scale changes. A break-through invention in the field of feature detection was the invention of SIFT (Scale Invariant Feature Transform) [11]. The method is able to produce large numbers of stable features, and also provides a way to compute the feature descriptors, which are important for matching. The main drawback of the method is its computational complexity. Many authors tried to accelerate SIFT computation while preserving or improving its properties. SURF (Speeded-up Robust Features) [12] managed to achieve better computational performance by using box filters to approximate LoG (Laplacians of Gaussians), and by employing Haar wavelets in descriptor computation, but with a price of loosing the rotational invariance. CenSurE detector (Center Surround Extrema) [13] retains SURF advantages while approximating LoG by center-surround bi-level filters. ORB detector (Oriented FAST and Rotated BRIEF) [14] claims to be an effective replacement of SIFT, although it rather belongs to Harris family of detectors. Still, in contrast to SIFT and SURF, it is not patented and can be used also in commercial projects. The implementation is freely available as a part of OpenCV project [15].

Both SIFT and ORB were tested on the SEM images and are provided as options in the stitching application. Their comparison is included in the Results chapter. Even though SIFT generally outperformed ORB in quality of found keypoints and their matching, the speed-up of ORB is so large that it deserves a consideration. Descriptions of both methods follow in the next sections.

4.1 Scale Invariant Feature Transform overview

SIFT is a method that was invented and presented by David Lowe [11] and it is patented by University of British Columbia. It consists of four main stages:

1. Scale-space extrema detection,
2. Keypoint localization,
3. Orientation assignment,
4. Keypoint description.

Scale-space extrema detection In the first stage, the input image $I(x, y)$ is extended to the Gaussian, and the Difference of Gaussians (DoG) pyramids. In particular, DoG serves as a good approximation to the scale-normalized Laplacian, whose extrema are proven to be stable and invariant across scales. Moreover, DoG ($D(x, y, \sigma)$) can be computed from the Gaussian pyramid ($L(x, y, \sigma)$) by subtraction of the nearby scales, separated by the constant factor k :

$$\begin{aligned}L(x, y, \sigma) &= G(x, y, k\sigma) * I(x, y), \\D(x, y, \sigma) &= (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y), \\&= L(x, y, k\sigma) - L(x, y, \sigma), \\G(x, y, \sigma) &= \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2}.\end{aligned}$$

The pyramid is composed of octaves and intervals. In computer vision jargon, an octave refers to a set of images, which have the same size, and differ by the blur level. The images inside an octave are

referred as intervals. Number of intervals per octave is one of the parameters of SIFT detector. By setting $k = 2^{1/numberIntervals}$, the blur level of an image doubles after each *numberIntervals* steps. An image with double blur level starts a new octave; it can be downsampled to quarter size while preserving the sampling accuracy to blur level ratio, which reduces the computation time.

The keypoint candidates are detected as the minima and maxima in the DoG pyramid by comparing each pixel value with its 26 neighbours (8-connectivity in the local and adjacent scales).

Keypoint localization The second stage aims to improve the localization of the extrema to achieve the sub-pixel precision, and to filter them to discard unstable responses. Sub-pixel precision is achieved by using Taylor expansion of $D(x, y, \sigma)$ up to quadratic terms. The extrema are located at the points where the first derivatives of D are equal to 0. Unstable extrema are considered those having low contrast and those being localized along edges. The first category instabilities are removed by thresholding the function values at interpolated locations (with typical threshold value 0.003). For the second category instabilities, the ratio of the principal curvatures at the extrema is considered and only the extrema having the ratio small (typically < 10) are maintained.

Orientation assignment Assigning the orientation to the keypoints ensures the property of rotational invariance. Gaussian smoothed image at scale σ is selected, and the gradient magnitudes and orientations are computed for each pixel within a region around the keypoint. The orientation histogram is composed of 36 bins, each representing a different fraction of the full rotational angle (2π). The intensities weighted by the gradient magnitudes are added to their bins based on the gradient orientations. The peaks in the histogram are assigned as the orientations of the keypoints. If there are more peaks within the 80% of the highest peak, the keypoint is copied and added to the set of keypoints repeatedly with different orientations.

Keypoint description In the final stage, the SIFT descriptor is computed. Like before, a neighbourhood of a keypoint in the correspond-

ing Gaussian blurred image is considered, but with the fixed neighbourhood of 16×16 pixels. This neighbourhood is further split into four 4×4 regions and for each small region, a weighted orientation histogram is computed similarly as in the previous stage. The four histograms are concatenated, resulting in 128-bin feature vector. To achieve position invariance of the features, it is recommended to normalize the feature vector to unit length.

Ordinal descriptor Toews and Wells [16] suggest an alternative to the standard SIFT descriptor, which according to their results improves the matching precision. Ordinal descriptor can be obtained by replacing each value in the standard SIFT descriptor by its rank:

$$r_i = |x_k : x_k \leq x_i|.$$

No further normalization of the descriptor is performed. Instead, the authors suggest using Spearman or Kendall correlation coefficients to evaluate the goodness of match.

4.2 Oriented FAST and Rotated BRIEF

ORB incorporates and slightly improves the ideas from two other works: FAST detector (Features from accelerated segment test) [17] and BRIEF descriptor (Binary Robust Independent Elementary Features) [18]. Like before, this section is not an exhaustive description of the method, but rather an overview of the main concepts.

Ordinary FAST FAST detector has three main steps. In the first step, a segment test is performed at a pixel. It considers a Bresenham circle of 16 pixels to determine corner candidates. If either n contiguous pixels are brighter than the central pixel's intensity $+t$, or all of them are darker than the central pixel's intensity $-t$ (t is a predefined intensity threshold), the central pixel is considered a keypoint candidate. In the next step, each circle pixel of each candidate is assigned into three sets based on the relative brightness. A machine learning is employed to decide which candidates are corners (by minimizing the entropy of the set of selected corners). The third step is non-maximal

suppression. From a pair of adjacent candidates, the one with a lower sum of absolute differences between the center and the circle pixels is discarded.

FAST improvements In ORB, a Harris corner measure is employed to filter the corners to avoid edge responses. Also, FAST is separately applied to each level of a scale space pyramid to obtain multi-scale keypoints. The orientation is assigned by using the intensity centroid method:

$$\theta = \arctan\left(\sum_{x,y} yI(x,y), \sum_{x,y} xI(x,y)\right).$$

BRIEF with location Standard BRIEF descriptor is defined as a binary string of length n , built by comparing the intensities in a smoothed image patch p at n (x, y) location pairs:

$$f_n(p) = \sum_{1 \leq i \leq n} 2^{i-1} \tau(p, x_i, y_i),$$

$$\tau(p, x_i, y_i) = \begin{cases} 1 & \text{if } p(x_i) < p(y_i), \\ 0 & \text{otherwise.} \end{cases}$$

Steered BRIEF can be obtained by using the same definition, but the location pairs are rotated according to the patch rotation, which was computed in the FAST improvement step. The authors of ORB suggest discretizing the patch rotation angle to 30 values (step $\frac{2\pi}{30}$) and to create a lookup table of precomputed BRIEF patterns. The original BRIEF has a property of large variance and mean ≈ 0.5 . However, the property is lost after applying rotation. To recover it, a greedy search is performed to find the tests (coordinate pairs) satisfying the original properties. The output of ORB are keypoint positions accompanied with their binary descriptors.

4.3 Keypoint matching

Keypoint matching is no less important than the keypoint detector, because false matches might spoil the estimate of the transform sig-

nificantly. Keypoint matching techniques are generally of two types: they are either brute-force, testing all options to match a keypoint, or they partition the search space and solve the matching as a combinatorial problem. Avoiding brute-force matching can reduce the computation time, especially when matching larger sets. In this work, the image overlaps are of rather small sizes and produce quite small sets of features (at most ~ 1000 elements), so the brute-force matching is sufficient. Moreover, the overhead building the search data structures has much larger impact on the computation time when matching small sets.

A suitable metric must be selected to evaluate the goodness of match between two features. A Hamming distance (count of different bits) is used in case of ORB method. SIFT descriptors are evaluated using the sum of square differences, both in case of standard and ordinal descriptors. In case of ordinal descriptor, Spearman coefficient evaluation of matches was also tested, but it resulted in higher number of false matches than while using the sum of square distances.

Because naïve brute-force matching results in large amount of false matches (see Figure 4.1a), Lowe [11] suggested considering two best matches for each feature and keep the best match only if the second-best is worse by at least 25% of the best match. The effect of Lowe's filtering is illustrated in Figure 4.1b.

To improve over Lowe's approach, the fact that the distortion produces nearly uniform displacements is used. After applying the Lowe's filtering, the remaining correspondences are sorted into four sets based on the slope of the line segment connecting the positions of the keypoints in the overlaps. Let $K_1 = (x_1, y_1)$ and $K_2 = (x_2, y_2)$ be the positions of two keypoints and let $\Delta = (dx, dy) = K_1 - K_2$. Then the four sets contain correspondences which:

1. have positive dx and positive dy ,
2. have positive dx and negative dy ,
3. have negative dx and positive dy ,
4. have negative dx and negative dy .

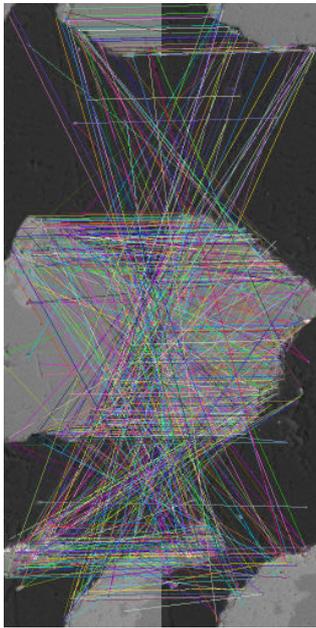
Under the assumption that most of the correspondences are correct, the set that has the highest cardinality is preserved and the rest

correspondences are discarded. In different applications, or if a weaker feature detector and descriptor were used, such assumption could be incorrect, although for the purpose of this work it showed good results. The effect is illustrated in Figure 4.1c.

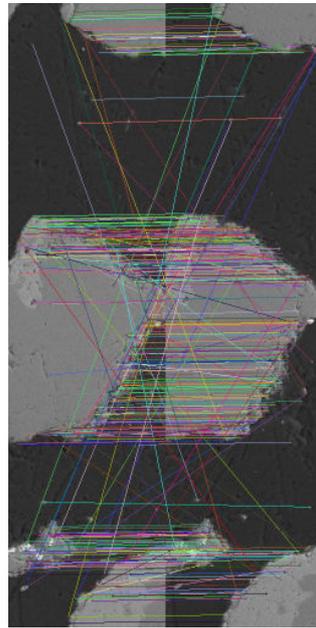
The final step to remove the false matches is based on RANSAC algorithm (Random Sample Consensus) [19]. In each iteration, the algorithm randomly selects three pairs of matched keypoints and computes a homography transforming the three keypoints from the first to the second overlap. After applying the homography to the first overlap, the keypoints that are transformed close to their pairs are considered inliers. The set of inliers is recorded in each iteration. After the last iteration, the set containing the most elements is returned as the final set of correspondences. In this work, the number of iterations for RANSAC filtering was set to 10000 and the distance threshold for inliers was set to 10 pixels. The result is illustrated in Figure 4.1d.

Note that RANSAC should be able to cope with the result given by the brute-force directly without the need of performing additional filtering steps. However, the filtering steps enable to use less RANSAC iterations while producing equally stable results.

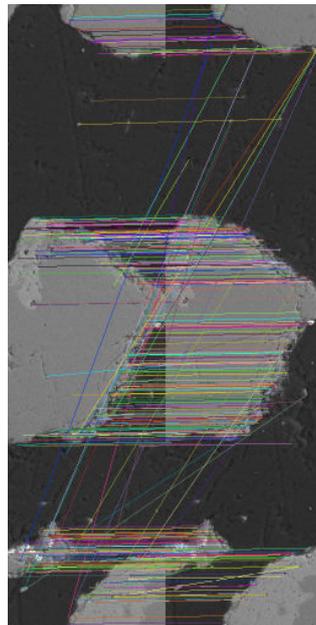
4. KEYPOINT DETECTION AND MATCHING



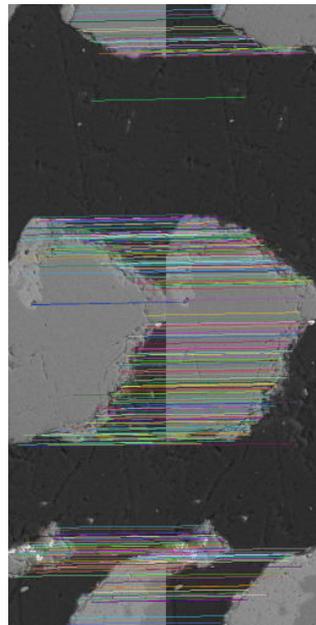
(a) Naïve brute-force matching.



(b) Lowe's filtering.



(c) Highest cardinality set.



(d) Filtering by RANSAC.

Figure 4.1: Different stages of keypoint matching performed in one of the overlaps.

5 Mathematical description

5.1 Coordinate systems and stitching

Recalling the aim of this work, the distortion correction transform and a set of stage shift errors need to be estimated such that the panoramic image, stitched together from the scanned tiles, does not contain any visual defects.

The panorama stitching works with three different coordinate systems:

1. tile local coordinate system,
2. global coordinate system,
3. coordinates of tiles.

The local coordinate system has its origin in the top-left corner of the tile, with horizontal axis denoted x and the vertical axis denoted y . The global distortion correction transform is applied to each tile separately in its local coordinates.

The global coordinate system can be defined such that its origin is located in the upper-left corner of the upper-left tile. Panorama stitching is basically a transform converting the local tile coordinates to the global coordinates.

Tile coordinates in the panorama are also important for stitching, because they determine the shift size between the local and the global coordinates of a tile. Each tile is assigned coordinates (r, c) , where r stands for the tile row and c for tile column. Tile coordinates are illustrated in image 5.3.

5.2 The global distortion correction transform

The global distortion correction transform serves for correcting the distortion generated by the optical system of SEM. It is designed such that it respects the construction of the electron column of SEM and the distortion types that emerge there. Recall that the main distortion types are zoom error, non-orthogonality, and rotation error. The first

5. MATHEMATICAL DESCRIPTION

two are generated first and together, while the rotation error arises later.

Denote N the operator causing the non-orthogonality, S the zooming error and R the rotation error. The order of non-orthogonality and zoom error cannot be determined, but the order of the operations mathematically matters because matrix multiplication is not commutative. As a result, two models for the distortion transform have to be considered:

$$\begin{aligned} T_1 &= R \cdot (S \circ N), \\ T_2 &= R \cdot (N \circ S). \end{aligned}$$

The aim is to find the inverse transform:

$$\begin{aligned} T_1^{-1} &= (N^{-1} \circ S^{-1}) \cdot R^{-1}, \\ T_2^{-1} &= (S^{-1} \circ N^{-1}) \cdot R^{-1}. \end{aligned}$$

S^{-1} and R^{-1} are simple scaling and rotation matrices of the forms:

$$\begin{aligned} S^{-1} &= \begin{pmatrix} s_x & 0 & 0 \\ 0 & s_y & 0 \\ 0 & 0 & 1 \end{pmatrix}, \\ R^{-1} &= \begin{pmatrix} \cos(\varphi) & -\sin(\varphi) & -c_x \cdot \cos(\varphi) + c_y \cdot \sin(\varphi) + c_x \\ \sin(\varphi) & \cos(\varphi) & -c_x \cdot \sin(\varphi) - c_y \cdot \cos(\varphi) + c_y \\ 0 & 0 & 1 \end{pmatrix}. \end{aligned}$$

Non-orthogonality has the form of rhomboidal skewness. The situation is illustrated in the Figure 5.1. Suppose each point has non-orthogonal coordinates (X', Y') . To compute its coordinates (X, Y) in the orthogonal coordinate space, the transform of the form

$$N^{-1} = \begin{pmatrix} \cos(\psi_1) & \sin(\psi_2) & 0 \\ \sin(\psi_1) & \cos(\psi_2) & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

can be used.

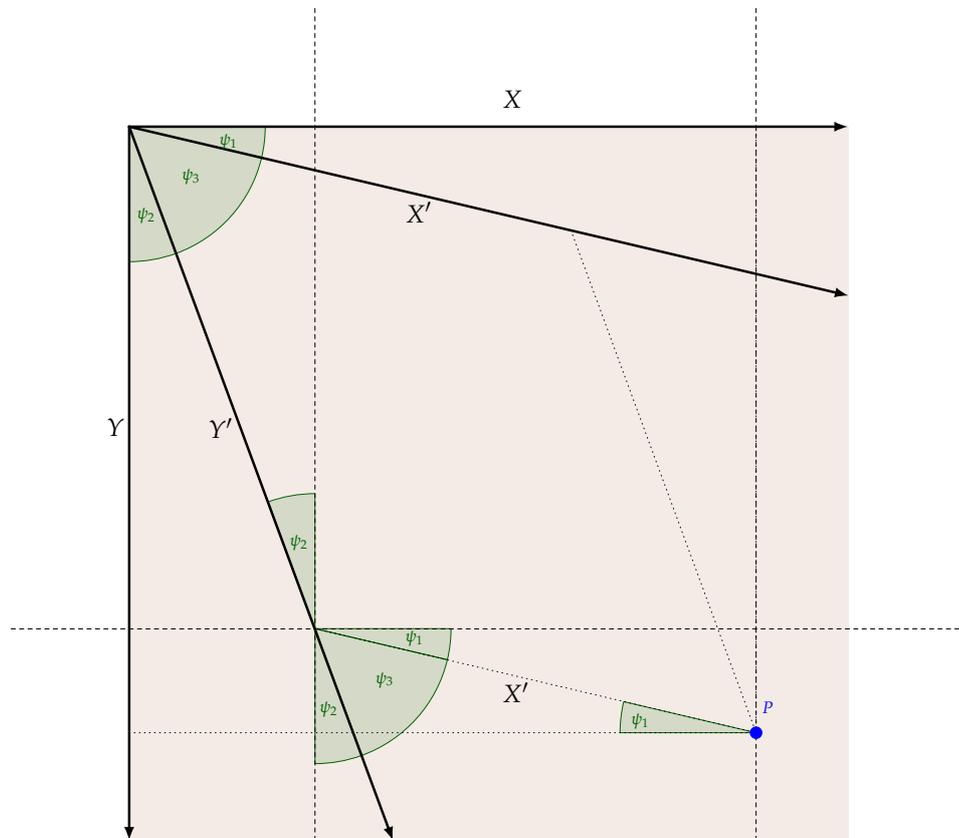


Figure 5.1: Non-orthogonality: The point P has non-orthogonal coordinates (X', Y') and orthogonal coordinates (X, Y) .

All the parameters can only take values from the limited intervals. The angles for both rotation and non-orthogonality are expected to be close to zero. As a result, sine and cosine functions can be substituted for their approximations for small angles:

$$\begin{aligned}\sin(\varphi) &= \varphi, \\ \cos(\varphi) &= 1 - 0.5\varphi^2.\end{aligned}$$

5. MATHEMATICAL DESCRIPTION

This yields:

$$N^{-1} = \begin{pmatrix} 1 - 0.5\psi_1^2 & \psi_2 & 0 \\ \psi_1 & 1 - 0.5\psi_2^2 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

$$R^{-1} = \begin{pmatrix} 1 - 0.5\varphi^2 & -\varphi & 0.5\varphi^2 c_x + c_y \cdot \varphi \\ \varphi & 1 - 0.5\varphi^2 & 0.5\varphi^2 c_y - c_x \cdot \varphi \\ 0 & 0 & 1 \end{pmatrix}.$$

The final global correction matrix has the form:

$$T_1^{-1} = (N^{-1} \circ S^{-1}) \cdot R^{-1} = \begin{pmatrix} A_1 & B_1 & C_1 \\ D_1 & E_1 & F_1 \\ 0 & 0 & 1 \end{pmatrix},$$

$$A_1 = s_x(1 - 0.5\varphi^2 - 0.5\psi_1^2 + 0.25\varphi^2\psi_1^2) + s_y\varphi\psi_2,$$

$$B_1 = -s_x(\varphi - 0.5\varphi\psi_1^2) + s_y(\psi_2 - 0.5\psi_2\varphi^2),$$

$$C_1 = 0.25s_x\varphi(2c_x\varphi + 4c_y - c_x\varphi\psi_1^2 - 2c_y\psi_1^2) + s_y\varphi\psi_2(0.5c_y\varphi - c_x),$$

$$D_1 = s_x(\psi_1 - 0.5\varphi^2\psi_1) + s_y(\varphi - 0.5\varphi\psi_2^2),$$

$$E_1 = -s_x\varphi\psi_1 + s_y(1 - 0.5\varphi^2 - 0.5\psi_2^2 + 0.25\varphi^2\psi_2^2),$$

$$F_1 = s_x\varphi\psi_1(0.5c_x\varphi + c_y) + 0.25s_y\varphi(2c_y\varphi - 4c_x - c_y\psi_2^2\varphi + 2c_x\psi_2^2).$$

Or alternatively:

$$T_2^{-1} = (S^{-1} \cdot N^{-1}) \circ R^{-1} = \begin{pmatrix} A_2 & B_2 & C_2 \\ D_2 & E_2 & F_2 \\ 0 & 0 & 1 \end{pmatrix},$$

$$\begin{aligned}
 A_2 &= s_x(1 - 0.5\varphi^2 - 0.5\psi_1^2 + 0.25\varphi^2\psi_1^2 + \varphi\psi_2), \\
 B_2 &= s_x(-\varphi + 0.5\psi_1^2\varphi + \psi_2 - 0.5\varphi^2\psi_2), \\
 C_2 &= 0.25s_x\varphi(2c_x\varphi + 4c_y - c_x\varphi\psi_1^2 - 2c_y\psi_1^2 + 2c_y\varphi\psi_2 - 4c_x\psi_2), \\
 D_2 &= s_y(\psi_1 - 0.5\varphi^2\psi_1 + \varphi - 0.5\varphi\psi_2^2), \\
 E_2 &= s_y(-\varphi\psi_1 + 1 - 0.5\varphi^2 - 0.5\psi_2^2 + 0.25\varphi^2\psi_2^2), \\
 F_2 &= 0.25s_y\varphi(2c_x\varphi\psi_1 + 4c_y\psi_1 + 2c_y\varphi - 4c_x - c_y\psi_2^2\varphi + 2c_x\psi_2^2).
 \end{aligned}$$

The effect of application of the global correction transform to a rectangular tile is illustrated in Figure 5.2. Scaling was omitted in the illustration for the sake of simplicity.

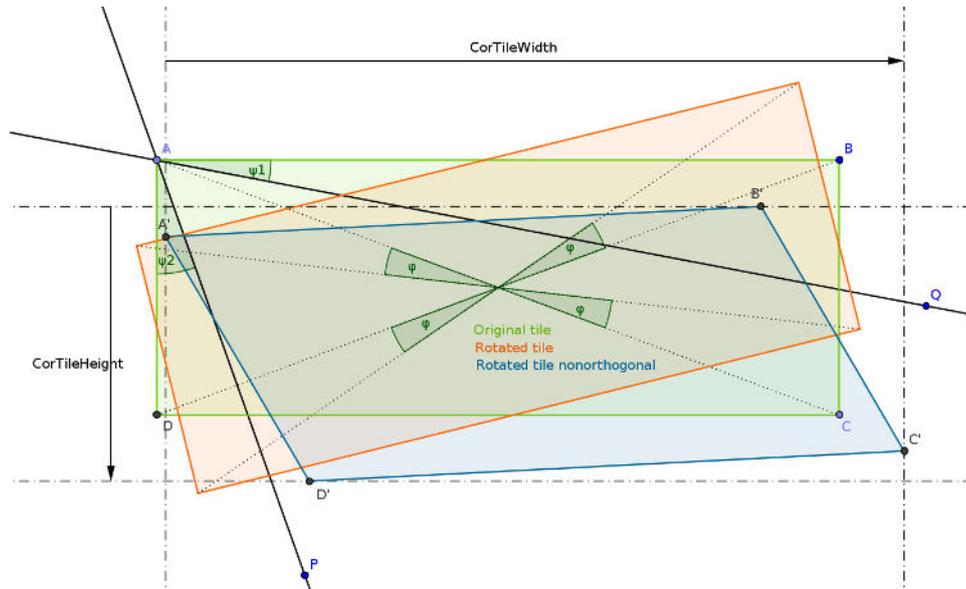


Figure 5.2: The effect of global distortion on tile shape and size. The original tile is in green colour, orange is after applying rotation and blue after applying non-orthogonality.

Note the new size of a tile is defined as a distance between the left-most and the right-most pixel, or the top-most and the bottom-most pixel respectively. In Figure 5.2, it is denoted by $CorTile_{width}$ and $CorTile_{height}$.

5.3 Stage shifts and stitching

Everytime a scan of a tile is complete, the specimen stage is given an instruction to move to a new position. If no error (shifting nor global) was introduced, the stage shift would be of size $\vec{\sigma}$. The size of shift is equal to:

- $(Tile_{width} - Overlap_{width}, 0)$ in case of a shift in a row,
- $(0, Tile_{height} - Overlap_{height})$ in case of a shift in a column.

Both values, the dimensions of a tile and the size of overlap, are given in pixels in the metadata, which makes them sensitive to zoom error. Recall that zoom error is introduced due to unequal step size in both principal directions when rastering the tile with an electron beam. The global distortion correction imposes scaling to cope with the zoom error; however, scaling effectively adds or removes pixels of a tile and so it alters the expected stage shift.

Moreover, the stage positioning is not precise. The variables denoting shift errors between neighbouring positions in a row are denoted $\vec{\delta}$ and are indexed the same way as the tile on the left, $\vec{\epsilon}$ denote shift errors between neighbouring positions in a column and are indexed like the tile on the top. For clarification, see Figure 5.3.

The definition of shifts allows establishing the transform from the tiles' local coordinate systems to the global coordinate system. For the sake of simplicity, suppose that tiles in each dataset have uniform sizes and that the size of stage shift between two neighbouring positions is also uniform. Indeed, this assumption holds in all tested datasets and it is presumed to be true in the vast majority of practical cases.

One of the ways to write the change of the local coordinate system to the global coordinate system is

$$\begin{aligned} p_{global_x} &= T^{-1}p_{local_x} + (i - 1) \cdot s_x\sigma_{row_x} + \sum_{l=0}^j \delta_{0,l_x} + \sum_{l=0}^i \epsilon_{l,j_x}, \\ p_{global_y} &= T^{-1}p_{local_y} + (j - 1) \cdot s_y\sigma_{col_y} + \sum_{l=0}^j \delta_{0,l_y} + \sum_{l=0}^i \epsilon_{l,j_y}; \end{aligned} \quad (5.1)$$

where p_{global} and p_{local} are the global, respectively local coordinates of a point in $Tile_{i,j}$. The equations above imply considering $Tile_{1,1}$ as

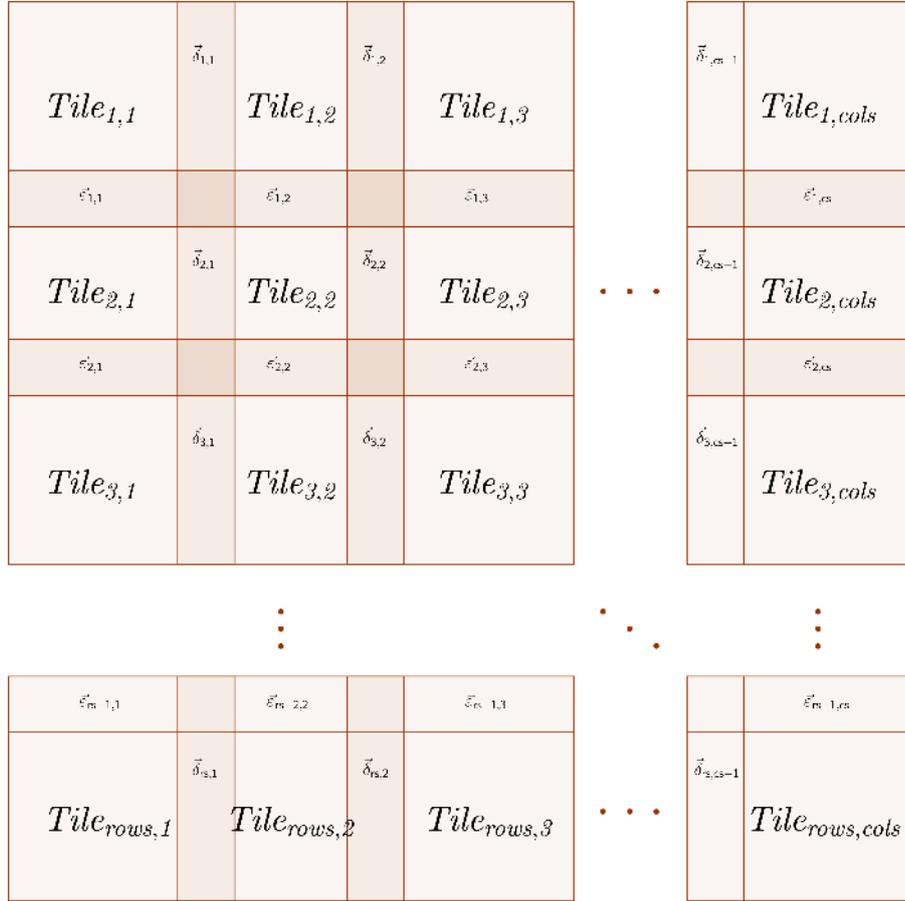


Figure 5.3: Panorama coordinate system.

the reference and aligning the rest of panorama towards it. Also, the transform is not unique. Any sequence of shifts starting in the upper left corner and ending at position of $Tile_{i,j}$ can be used.

5.4 The optimization cost function

The definition of the optimization cost function is straightforward with respect to the desired result. In the correctly stitched panorama, for each corresponding points P_1, P_2 it must hold that:

5. MATHEMATICAL DESCRIPTION

- $\forall (P_1, P_2), P_1 \in \text{Tile}_{i,j} \wedge P_2 \in \text{Tile}_{i,j+1}$:

$$T^{-1}P_1 = T^{-1}P_2 + S^{-1}\vec{\sigma}_{row} + \vec{\delta}_{i,j},$$

- $\forall (P_1, P_2), P_1 \in \text{Tile}_{i,j} \wedge P_2 \in \text{Tile}_{i+1,j}$:

$$T^{-1}P_1 = T^{-1}P_2 + S^{-1}\vec{\sigma}_{col} + \vec{\epsilon}_{i,j}.$$

Like before, T^{-1} stands for the global distortion correction transform, S^{-1} denotes scaling part of T^{-1} , $\vec{\sigma}$ is the expected table shift between two neighbouring tiles and $\vec{\delta}_{i,j}$, $\vec{\epsilon}_{i,j}$ are the stage shift errors.

For the optimization, consider the sum of squared distances between all corresponding points in the panorama (L2 norm). The aim is to minimize the cost function \mathcal{C} :

$$\begin{aligned} \mathcal{C} &= \sum_{i=1}^{rows} \sum_{j=1}^{cols-1} \sum_{(P_1, P_2) \in \text{RowCorr}_{i,j}} \left\| T^{-1}P_1 - T^{-1}P_2 - S^{-1}\vec{\sigma}_{row} - \vec{\delta}_{i,j} \right\|_2^2 \\ &+ \sum_{i=1}^{rows-1} \sum_{j=1}^{cols} \sum_{(P_1, P_2) \in \text{ColCorr}_{i,j}} \left\| T^{-1}P_1 - T^{-1}P_2 - S^{-1}\vec{\sigma}_{col} - \vec{\epsilon}_{i,j} \right\|_2^2. \end{aligned}$$

In the equation above, $\text{RowCorr}_{i,j}$ denotes the set of corresponding points in row i between columns j and $j + 1$ (tiles $\text{Tile}_{i,j}$ and $\text{Tile}_{i,j+1}$), and $\text{ColCorr}_{i,j}$ denotes the set of corresponding points in column j between rows i and $i + 1$ (tiles $\text{Tile}_{i,j}$ and $\text{Tile}_{i+1,j}$).

The inner formulas can be simplified. Let $\Delta = P_1 - P_2$. Then:

$$\begin{aligned} &\left\| T^{-1}P_1 - T^{-1}P_2 - S^{-1}\vec{\sigma} - \vec{\delta}_{i,j} \right\|_2^2 = \left\| T^{-1}\Delta - S^{-1}\vec{\sigma} - \vec{\delta}_{i,j} \right\|_2^2 \\ &= \left\| \begin{array}{c} A\Delta_x + B\Delta_y - s_x\sigma_x - \delta_{ij_x} \\ D\Delta_x + E\Delta_y - s_y\sigma_y - \delta_{ij_y} \\ 0 \end{array} \right\|_2^2 \\ &= \left(A\Delta_x + B\Delta_y - s_x\sigma_x - \delta_{ij_x} \right)^2 + \left(D\Delta_x + E\Delta_y - s_y\sigma_y - \delta_{ij_y} \right)^2. \end{aligned}$$

The expanded form of the cost function is:

$$\begin{aligned}
 \mathcal{C} = & \sum_{i=1}^{rows} \sum_{j=1}^{cols-1} \sum_{\substack{\Delta=P_1-P_2 \\ (P_1, P_2) \in RowCorr_{i,j}}} ((A\Delta_x - B\Delta_y - s_x\sigma_{row_x} - \delta_{ij_x})^2 \\
 & + (D\Delta_x + E\Delta_y - s_y\sigma_{row_y} - \delta_{ij_y})^2) \\
 & + \sum_{i=1}^{rows-1} \sum_{j=1}^{cols} \sum_{\substack{\Delta=P_1-P_2 \\ (P_1, P_2) \in ColCorr_{i,j}}} ((A\Delta_x - B\Delta_y - s_x\sigma_{col_x} - \varepsilon_{ij_x})^2 \\
 & + (D\Delta_x + E\Delta_y - s_y\sigma_{col_y} - \varepsilon_{ij_y})^2).
 \end{aligned}$$

The column shift is expected to be zero ($\sigma_{row_y} = 0$) for each pair of neighbouring tiles in a column, respectively the row shift should be 0 ($\sigma_{col_x} = 0$) for a pair of keypoints in the neighbouring tiles in a row. By substitution it yields:

$$\begin{aligned}
 \mathcal{C} = & \sum_{i=1}^{rows} \sum_{j=1}^{cols-1} \sum_{\substack{\Delta=P_1-P_2 \\ (P_1, P_2) \in RowCorr_{i,j}}} ((A\Delta_x - B\Delta_y - s_x\sigma_{row_x} - \delta_{ij_x})^2 \\
 & + (D\Delta_x + E\Delta_y - \delta_{ij_y})^2) \\
 & + \sum_{i=1}^{rows-1} \sum_{j=1}^{cols} \sum_{\substack{\Delta=P_1-P_2 \\ (P_1, P_2) \in ColCorr_{i,j}}} ((A\Delta_x - B\Delta_y - \varepsilon_{ij_x})^2 \\
 & + (D\Delta_x + E\Delta_y - s_y\sigma_{col_y} - \varepsilon_{ij_y})^2)
 \end{aligned} \tag{5.2}$$

The above formula can be further expanded to a form suitable for implementation:

$$\begin{aligned}
 \mathcal{C} = & \sum_{i=1}^{rows} \sum_{j=1}^{cols-1} \left(\sum_{\substack{\Delta=P_1-P_2 \\ (P_1, P_2) \in RowCorr_{i,j}}} \Delta_x^2 \right) (A^2 + D^2) - 2 \left(\sum_{\substack{\Delta=P_1-P_2 \\ (P_1, P_2) \in RowCorr_{i,j}}} \Delta_x \right) \\
 & \cdot (As_x \sigma_{row_x} + A\delta_{i,j_x} + D\delta_{i,j_y}) + 2 \left(\sum_{\substack{\Delta=P_1-P_2 \\ (P_1, P_2) \in RowCorr_{i,j}}} \Delta_x \Delta_y \right) (AB + DE) \\
 & - 2 \left(\sum_{\substack{\Delta=P_1-P_2 \\ (P_1, P_2) \in RowCorr_{i,j}}} \Delta_y \right) (Bs_x \sigma_{row_x} + B\delta_{i,j_x} + E\delta_{i,j_y}) + \left(\sum_{\substack{\Delta=P_1-P_2 \\ (P_1, P_2) \in RowCorr_{i,j}}} \Delta_y^2 \right) \\
 & \cdot (B^2 + E^2) + |RowCorr_{i,j}| (s_x^2 \sigma_{row_x}^2 + 2s_x \sigma_{row_x} \delta_{i,j_x} + \delta_{i,j_x}^2 + \delta_{i,j_y}^2) \\
 & + \sum_{i=1}^{rows-1} \sum_{j=1}^{cols} \left(\sum_{\substack{\Delta=P_1-P_2 \\ (P_1, P_2) \in ColCorr_{i,j}}} \Delta_x^2 \right) (A^2 + D^2) - 2 \left(\sum_{\substack{\Delta=P_1-P_2 \\ (P_1, P_2) \in ColCorr_{i,j}}} \Delta_x \right) \\
 & \cdot (A\varepsilon_{i,j_x} + Ds_y \sigma_{col_y} + D\varepsilon_{i,j_y}) + 2 \left(\sum_{\substack{\Delta=P_1-P_2 \\ (P_1, P_2) \in ColCorr_{i,j}}} \Delta_x \Delta_y \right) (AB + DE) \\
 & - 2 \left(\sum_{\substack{\Delta=P_1-P_2 \\ (P_1, P_2) \in ColCorr_{i,j}}} \Delta_y \right) (B\varepsilon_{i,j_x} + Es_y \sigma_{col_y} + E\varepsilon_{i,j_y}) + \left(\sum_{\substack{\Delta=P_1-P_2 \\ (P_1, P_2) \in ColCorr_{i,j}}} \Delta_y^2 \right) \\
 & \cdot (B^2 + E^2) + |ColCorr_{i,j}| (s_y^2 \sigma_{col_y}^2 + 2s_y \sigma_{col_y} \varepsilon_{i,j_y} + \varepsilon_{i,j_x}^2 + \varepsilon_{i,j_y}^2).
 \end{aligned}$$

The main advantage of this form is the possibility to store just six numbers for each overlap: the sums of Δx^2 , Δx , Δxy , Δy , Δy^2 and number of correspondences, compute them in advance and not iterate through all the correspondences in each iteration of the optimization method.

5.5 The optimization constraints

The minimization of the cost function must be performed with respect to a set of constraints, which should reflecting the physical limits of the microscope. Expected intervals for the distortion parameters were given by an expert, yet they need to be confirmed or corrected by the outcomes of the optimization.

Let Err_{op} denote the size of error of operation op . Let \widetilde{Err}_{op} be the expert's estimate of the size of the error. The optimization has to be performed subject to the following constraints:

1. zoom error should be up to 1%:

- $s_x \in [1.0 - Err_{zoom}, 1.0 + Err_{zoom}]$,
- $s_y \in [1.0 - Err_{zoom}, 1.0 + Err_{zoom}]$, $\widetilde{Err}_{zoom} \approx 0.01$,

2. rotation error should be up to 1° :

- $\varphi \in [-Err_{rotation}, Err_{rotation}]$, $\widetilde{Err}_{rotation} \approx 1^\circ$,

3. non-orthogonality error should be up to 1° :

- $\psi_1 \in [-Err_{nonortho}, Err_{nonortho}]$,
- $\psi_2 \in [-Err_{nonortho}, Err_{nonortho}]$,
- $\psi_1 + \psi_2 \in [-Err_{nonortho}, Err_{nonortho}]$, $\widetilde{Err}_{nonortho} \approx 1^\circ$,

4. stage positioning error should be random and more precise when performing close shifts:

- $\forall \delta_{ij_x} : \delta_{ij_x} \in [-Err_{shift}, Err_{shift}]$,
- $\forall \delta_{ij_y} : \delta_{ij_y} \in [-Err_{shift}, Err_{shift}]$,

5. MATHEMATICAL DESCRIPTION

- $\forall \varepsilon_{ij_x} : \varepsilon_{ij_x} \in [-Err_{shift}, Err_{shift}]$,
- $\forall \varepsilon_{ij_y} : \varepsilon_{ij_y} \in [-Err_{shift}, Err_{shift}]$, $\widetilde{Err}_{shift-near} \approx 1\mu m$,
 $\widetilde{Err}_{shift-far} \approx 5\mu m$,

5. Panorama size should not deviate from the expected size more than given by $Err_{shift-far}$.

The constraints on the panorama size require more explanation. The expected panorama size is effectively a size, which a panorama would have if no correction was applied and tiles were aligned at overlaps:

$$\begin{aligned} ExpPanorama_{width} &= Tile_{width} + (cols - 1)\sigma_{row_x}, \\ ExpPanorama_{height} &= Tile_{height} + (rows - 1)\sigma_{col_y}. \end{aligned}$$

Define $(Panorama_{width}, Panorama_{height})$ as the size of a stitched panorama after applying corrections. $Panorama_{width}$ corresponds to the distance between the left-most and the right-most pixel of the resulting image, $Panorama_{height}$ is the distance between the top-most and bottom-most pixel. Panorama size constraints are then:

$$\begin{aligned} ExpPanorama_{width} - Err_{shift-far} &\leq Panorama_{width} \\ &\leq ExpPanorama_{width} + Err_{shift-far}, \\ ExpPanorama_{height} - Err_{shift-far} &\leq Panorama_{height} \\ &\leq ExpPanorama_{height} + Err_{shift-far}. \end{aligned}$$

$ExpPanorama_{width}$ and $ExpPanorama_{height}$ are computed from the dataset and do not consider the scaling error. It might seem problematic, however, if $Err_{shift-far}$ is large enough, it can be neglected from the formula.

Because the expected bounds on the distortion parameters are rough, some of the actual distortion parameters can be larger than the expert's estimates. Therefore, the optimization method should have all Err_{op} set to $Err_{op} = \widetilde{Err}_{op} + d_{op}$, for some d_{op} large enough.

5.6 Properties and considerations regarding the optimization cost function

The optimization cost function has the following properties:

- it is a polynomial of highest total degree 10,
- it is nonnegative,
- it is a sum of squares,
- it is a function of $5 + 2 \cdot (rows - 1)(cols) + 2(rows)(cols - 1)$ variables.

Most of the optimization constraints are linear in the parameters of the distortion model. The only exception are the constraints on panorama size. Their analytic expression is not provided, but the size of panorama depends on the size of each tile, which depends on the global distortion correction transform, which is polynomial in its parameters.

Also, note that the global distortion correction transform is usually not sufficient to perfectly align both coordinates of corresponding keypoints. Moreover, if allowed stage shift error is large, there is an ambiguity concerning what part of misalignment should be corrected by the shift and what part should be corrected by the global distortion correction transform. The situation is illustrated in Figure 5.4.

Figure 5.4b shows the maximal alignment, that can be achieved by global distortion correction. Any portion of distortion not corrected by the global distortion correction must be corrected by one of the stage shift correction parameters. In the global distortion correction matrix, the keypoint alignment is achieved mostly thanks to the rotation and non-orthogonality. Scaling mostly controls the size of the resulting panorama, but in the model T_1^{-1} it also weights the effects of rotation and non-orthogonality.

The optimization cost function does not distinguish between the global distortion correction and the stage shifts. From the mathematical point of view, it is valid to assign values to all stage shift parameters and set the global distortion correction transform to identity, if it does not break the optimization constraints. To avoid such results and in

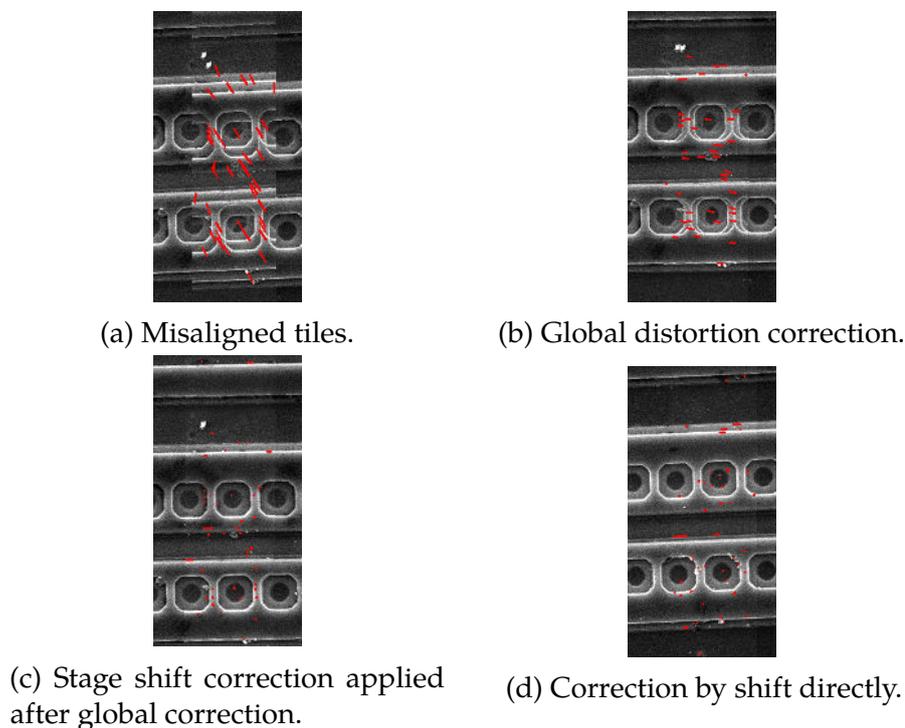


Figure 5.4: Illustration of different outcomes when preferring different types of distortion correction. Red line segments connect matched keypoints.

order to maximize the effect of global distortion correction, a two-step optimization can be considered. In the first step, the optimization method can consider only the global distortion correction parameters and all δ_{i,j_y} and ε_{i,j_x} . The optimal solution for such setup corresponds to the situation, where all stage shifts were performed in a row or in a column and did not generate errors in the other coordinate (for example, if stage moved in a row, no distortion was generated in a column). In the second step, the parameters found by the first step are fixed and the optimal assignment for parameters δ_{i,j_y} and ε_{i,j_x} is searched.

The optimization cost function sums corresponding points' distances for all overlaps in the panorama. The correctness of the keypoint matching should be ensured by RANSAC, but if there are little correspondences at the input of RANSAC filtering step, it can still output

a wrong matching. As a result, the optimization method might estimate the distortion parameters incorrectly. To avoid this behaviour, only the overlaps with more matches shall be considered. A threshold should depend also on the nature of the image data and quality of the keypoint detection, but generally a threshold value of at least 6 – 10 correspondences per overlap seems reasonable. Note that variables c_x and c_y denoting the centre of the rotation, which act in both global distortion models, got cancelled in the optimization cost function. However, it can be presumed that the centre of rotation is always placed in the centre of the tile. There is no physical interpretation for its displacement.

5.7 Considerations regarding stitching

If some of the overlaps are not considered in the optimization, the associated shift parameters cannot be determined and the stitching algorithm must be able to adapt to such situations. Moreover, even if all shift errors were estimated but some of them were not exact, the stitching based on the equation 5.1 would not be the best option. Each inexact parameter creates a visible misalignment in the panorama and the errors can accumulate by additions. Hence, it is desired to use as few stage shift error variables as possible. This can be achieved by the algorithm described in Algorithm 1, based on breadth-first-search strategy.

In Algorithm 1, *shiftCor* is a pair of values in which each tile stores its stage shift correction, so that it can be properly aligned to the rest of the panorama. Tile B is *reachable* from tile A if the shift between the tiles A and B was determined (the correspondences in the overlap between tiles A and B were considered in the optimization). The last if statement serves for cases where no overlaps of a tile were used in the optimization and therefore the tile is unreachable.

Algorithm 1 Stitching based on BFS

```

1: function STITCHING
2:    $Q \leftarrow \text{emptyQueue}$ 
3:    $(\text{Tile}_{\text{rows}/2, \text{cols}/2}).\text{shiftCor} \leftarrow (0, 0)$ 
4:    $\text{enqueue}(Q, \text{Tile}_{\text{rows}/2, \text{cols}/2})$ 
5:   while  $Q$  not empty do
6:      $\text{Tile} \leftarrow \text{dequeue}(Q)$ 
7:     Insert every pixel:
8:     for each pixel with coordinates  $P_{\text{local}}$  do
9:       Compute the global coordinates:  $P_{\text{global}} = T^{-1}P_{\text{local}} +$ 
           $\left( \text{Tile.coords} - \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right) \cdot S \begin{pmatrix} \sigma_{\text{row}_x} \\ \sigma_{\text{col}_y} \end{pmatrix} + \text{Tile.shiftCor}$ 
10:
11:       Set the intensity at  $P_{\text{global}}$ 
12:     end
13:     if there is a noninserted reachable tile above then
14:        $(\text{Tile}_{\text{above}}).\text{shiftCor} \leftarrow \text{Tile.shiftCor} - \varepsilon_{\text{Tile}_{\text{above}}.\text{coods}}$ 
15:        $\text{enqueue}(Q, \text{Tile}_{\text{above}})$ 
16:     if there is a noninserted reachable tile below then
17:        $(\text{Tile}_{\text{below}}).\text{shiftCor} \leftarrow \text{Tile.shiftCor} + \varepsilon_{\text{Tile.coods}}$ 
18:        $\text{enqueue}(Q, \text{Tile}_{\text{below}})$ 
19:     if there is a noninserted reachable tile on the left then
20:        $(\text{Tile}_{\text{left}}).\text{shiftCor} \leftarrow \text{Tile.shiftCor} - \delta_{\text{Tile}_{\text{left}}.\text{coods}}$ 
21:        $\text{enqueue}(Q, \text{Tile}_{\text{left}})$ 
22:     if there is a noninserted reachable tile on the right then
23:        $(\text{Tile}_{\text{right}}).\text{shiftCor} \leftarrow \text{Tile.shiftCor} - \delta_{\text{Tile.coods}}$ 
24:        $\text{enqueue}(Q, \text{Tile}_{\text{right}})$ 
25:     end
26:     if not all tiles were inserted then
27:       Find a non-inserted tile neighbouring with an inserted tile
28:       Insert the tile, its shiftCor = neighbour's shiftCor + shift
          estimated using phase correlation
29:       Add the tile's neighbours to  $Q$  like in the while loop
30:       Repeat the While loop

```

6 Optimization

Now that the model is defined and the stitching algorithm was presented, the only missing step is the minimization of the cost function. The aim is to find its global minimum. The choice of the optimization method depends on the properties of the problem. Generally, the problem is considered to be non-convex. To show whether the objective function is convex or not, two theorems could be used:

1. Sum of convex functions over a convex domain is again convex,
2. A function is convex if and only if its Hessian matrix (square matrix of its partial derivatives) is positive semidefinite.

Each of the terms inside the sums could be processed individually, producing a Hessian matrices of dimensions $m \times m$, which is too difficult to handle even for a small panoramas (for example, a panorama of 5×5 tiles has $m = 85$ variables).

Non-convex functions might have several local minima and local optimization methods could converge to a non-optimal solution. The following two sections describe the methods suited to this kind of problem: the first one is a deterministic method based on convex relaxations, using semi-definite programming. However, the size of the problem was found to be too large to be handled by a common computer. The second method described here is simulated annealing, a probabilistic method. Yet, the probabilistic methods are not guaranteed to find the exact solution to the problem. The local optimization methods are therefore employed after simulated annealing to improve the probabilistic guess.

6.1 Optimization via semi-definite programming relaxations

Polynomial optimization is a field that currently receives much of attention from the researchers [20]. When optimizing a polynomial using any of the common methods, the objective function's property of being polynomial is overlooked. In many cases, standard methods either fail to reach the global optimum, or are not deterministic. It

is desired to use the specific properties of the polynomial to design better optimization methods to achieve reliable convergence.

One of the most promising ideas in the polynomial optimization is the relaxation to a hierarchy of semi-definite programs. If given a polynomial defined over a compact semi-algebraic set and the polynomial is positive, the minimum of the polynomial is the largest positive number that can be subtracted from the polynomial such that the polynomial stays non-negative. One of the certificates of positivity of the polynomial is for example being a sum of squares (SOS) polynomial. If the polynomial problem is defined up to a set of polynomial constraints, it can be expressed for example by using Putinar's Positivstellensatz. Then, checking whether a polynomial is SOS is equivalent to solving a semi-definite program.

Because the field of polynomial optimization is very large and much depends complex, further information is omitted in this these. More details can be found in the literature, the most relevant are those by Parrilo [21] and Lasserre [22]. However, available software tools enable users to use the method without further understanding of the principle.

The optimization of the cost function (Equation 5.2) was tried by using C++ version of ncpol2sdpa package [23]. It enables the user to specify the polynomial optimization problem using symbolic variables and generates an input for semi-definite programming solver, sdpa [24]. However, the relaxation to a semi-definite program failed to compute, most likely because of the size of the problem. Size of the matrix for semi-definite programming is $\binom{n+d}{d} \times \binom{n+d}{d}$ [25]. A small panorama of 5×5 tiles requires finding 85 parameters (or 45 if enforcing zeros for each δ_y and ε_x), which means the matrix would have dimensions $\binom{90}{5} \approx 44 \cdot 10^6 \times 44 \cdot 10^6$ (or $\binom{50}{5} \approx 2 \cdot 10^6 \times 2 \cdot 10^6$).

6.2 Simulated annealing

Simulated annealing is one of the standard probabilistic optimization methods well-suited for constrained global optimization. It was presented by Kirkpatrick et al. in 1983 [26]. The method is inspired by annealing in metallurgy, where materials are heated so that their

atoms can overcome energetic barriers and get to more stable states after cooling. The algorithm is given in pseudo-code in Algorithm 2.

Algorithm 2 Simulated annealing general scheme.

```

1: function SIMULATED ANNEALING(temp, coolRate, init)
2:   current  $\leftarrow$  init
3:   currentEnergy  $\leftarrow$  EVALUATE(current)
4:   best  $\leftarrow$  current
5:   bestEnergy  $\leftarrow$  currentEnergy
6:   while temp > 1 do
7:     new  $\leftarrow$  GENERATE NEW STATE(current)
8:     energy  $\leftarrow$  EVALUATE(new)
9:     if energy < bestEnergy then
10:      best  $\leftarrow$  new
11:      bestEnergy  $\leftarrow$  energy
12:     if PROBABILITY(current  $\rightarrow$  new) > RANDOM(0,1) then
13:       current  $\leftarrow$  new
14:       currentEnergy  $\leftarrow$  energy
15:     temp  $\leftarrow$  temp  $\cdot$  (1.0 - coolRate)
16:   return best

```

Algorithm 3 Metropolis acceptance probability.

```

1: function PROBABILITY(current  $\rightarrow$  new)
2:   if energy < currentEnergy then
3:     return 1.0
4:   else
5:     return  $\exp\left(\frac{\text{currentEnergy} - \text{energy}}{\text{temp}}\right)$ 

```

Algorithm 2 has three inputs: the initial temperature of the system, denoted by *temp*, the cooling rate (*coolRate*) and the initial solution, denoted *init*. In each iteration, a new solution is generated and evaluated. The acceptance of the new solution is determined by the Metropolis criterion which is given in Algorithm 3.

In Algorithm 2, EVALUATE function evaluates the cost of the solution (equation 5.2) and RANDOM function generates a random real number

6. OPTIMIZATION

from the given interval. Function `GENERATE NEW STATE` alters given state, trying to improve the cost of the solution, while respecting the constraints defined for the optimization.

The algorithm setup for minimization of the cost function was the following:

- *init* was always set to no transform ($s_x = s_y = 1$, other variables were set to 0),
- *temp* was set 500000,
- *coolRate* was set to 0.003.

Generation of new solution (function `GENERATE NEW STATE`) in each iteration altered one of the following:

- rotation angle φ ,
- non-orthogonality parameters ψ_1 and ψ_2 ,
- horizontal magnification s_x ; if changing s_x violated any of the constraints on the panorama width, corresponding δ_{i,j_x} were regenerated,
- vertical magnification s_y ; if changing s_y violated any of the constraints on the panorama height, corresponding ε_{i,j_y} were regenerated,
- a vector of δ_{i,j_x} ; if all ε_{i,j_x} were enforced to zeros, it meant setting δ_{i,j_x} in all rows at once, otherwise the generation was performed in a certain row i , modifying also corresponding ε_{i,j_x} to preserve the correctness,
- a vector of ε_{i,j_y} ; if all δ_{i,j_y} were enforced to zeros, it meant setting ε_{i,j_y} in all columns at once, otherwise the generation was performed in a certain column j , modifying also corresponding δ_{i,j_y} to preserve the correctness.

The choice from above-defined options was random, as well as the values assigned to all the variables. Nonetheless, it was ensured that the assignment conforms all the constraints.

6.3 Local methods

Local methods were employed to improve the potentially imprecise solution given by the simulated annealing. There are many local methods available for the given type of problem; together they can be referred as the non-linear programming methods. The two main categories of algorithms are penalty methods and interior-point methods.

NLopt optimization library [27] was selected for the local optimization. A committee of 4 methods was employed:

- Constrained Optimization by Linear Approximations [28],
- Sequential Quadratic Programming [29],
- Method of Moving Asymptotes [30],
- Augmented Lagrangian algorithm [31].

General ideas of the methods are described in referenced articles; however, according to the author of NLopt, the implementation is not exact and was often modified or improved. Because of that, details on the local methods are omitted in this text.

The above methods shared the same setup: the optimization was set to stop when the relative change in all optimization parameters was smaller than $1 \cdot 10^{-15}$, or if the local optimization took longer than 120 seconds. The constraint on execution time was imposed to speed-up the optimization process. The local optimization performed with some the datasets took several hours, while it is disputable whether new solutions were noticeably improving.

The result of simulated annealing was used as a starting point for all of the methods, giving four candidates for the best solution. Out of these, the one with the best quality (the lowest value of the cost function) was selected as the final result.

In case of two-step optimization, the search of parameters was restricted in the first step to consider all variables except all δ_{i,j_y} and ε_{i,j_x} . In the second step, all four methods were re-run starting with the previously found best solution, restricting the search only to all δ_{i,j_y} and ε_{i,j_x} variables. Similarly as before, the one with the best cost was selected as a result.

7 Experiments and results

There were six datasets available for testing the previously defined method. They differed in many properties, such as number of tiles, sizes of tiles, magnification level, overlap size and the acquisition device they were captured by. The dataset characteristics are summarized in Table 7.1.

Dataset number		1	2	3	4	5	6
Content		Mexican pesos	Chip	Metalurgy I	Metalurgy II	Scobs I	Scobs II
Dataset size	Rows	8	4	18	11	6	5
	Columns	8	5	19	7	6	5
Tile size [px]	Width	768	768	1024	768	768	768
	Height	768	768	1024	576	768	768
Overlap size [px] (%)	In row	115 (15 %)	115 (15 %)	256 (25 %)	76 (10 %)	38 (5 %)	38 (5 %)
	In col.	115 (15 %)	115 (15 %)	256 (25 %)	57 (10 %)	38 (5 %)	38 (5 %)
Pixel size [μm]		3.397363	0.368164	0.78125	6.1074	2.6042	3.8728
Magnification		86.694	800	346.00	46.197	108.34	72.852
SEM device name		VEGA TS 5136LM	LYRA3 GMU	VEGA3 LMU	MIRA3 XMU		

Table 7.1: Properties of datasets.

Several experiments were conducted on top of the datasets. First, the suitability of ORB and SIFT keypoints was compared in order to decide for the right keypoint detection method for the following experiments.

Then, image panoramas were stitched together without applying any correction and with respect to shifts between tiles found by the phase correlation method. The quality measures of these results were computed and serve as a reference for comparison of the results obtained by the proposed method.

In the third step, the optimization was performed in the datasets several times. In the first trial, the expert-given bounds on the parameters were enforced and the results were recorded. In the next optimization runs, the bounds were loosened in order to find real size of error of each distortion parameter and in order to compose panoramas that were as seamless as possible (recall that no blending is performed at the seams). The quality of models T_1^{-1} and T_2^{-1} is compared, as well as the outcomes of one-step and two-step optimization processes. Further experiments involved normality and randomness

testing of the best solutions, checking stability of the solution by splitting datasets into subsets and comparing the partial outcomes, and trying to determine minimal subset of tiles for which the partial outcome outputs the same optical distortion correction matrix as for the whole panorama.

The quality of the solutions were evaluated by using Structural Symmetry Index (SSIM) [32]. The similarity between two signals x and y is defined as

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{x,y} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)},$$

where μ denotes the mean value, σ denotes standard deviation and C_1, C_2 are small constants to avoid computational instabilities. Wang et al. suggest computing SSIM values locally for each pixel, by considering a window of 11×11 pixels. Moreover, the intensities in the window should be weighted by Gaussian with $\sigma = 1.5$ and normalized to unit sum. The per-pixel SSIMs are then averaged to give the final result, which is denoted as MSSIM (mean SSIM). Both SSIM and MSSIM were computed in the overlapping regions of the panorama (by comparing each overlapping part of a panorama with corresponding overlap in a tile). Per overlap SSIM and MSSIM values were averaged for each overlap to give the final SSIM and MSSIM values for the resulting panoramas. Besides similarity, other qualities used for comparison of the results were the cost function normalized by the number of used correspondences and distances between the keypoints in the stitched panorama. While SSIM and MSSIM evaluate the visual quality of the result, solution cost per correspondence and distances between keypoints tell more about suitability of the model and quality of the optimization with respect to (possibly imprecise) positions of matched keypoints.

Normality of stage shift errors was tested using Shapiro-Wilk normality test [33] and visually by comparing the empirical distribution of stage shifts to the normal distribution by using Q-Q plots (Quantile-Quantile plots). Randomness of stage shift errors was tested using Kolmogorov-Smirnov test [34], by comparing the empirical distribution function to distribution function of normal distribution with

$\mu = 0$. The significance level for both Shapiro-Wilk and Komogorov-Smirnov tests was set to 5%.

The source codes for loading the data, optimization of the parameters and stitching the panoramas were implemented in C++ programming language and are provided in the electronic attachment of the thesis. The most of image processing tasks were implemented by using standard OpenCV library [15], including ORB keypoint detection. The implementation of SIFT was based on the code by Robb Hess [35], which was rewritten to feature modern OpenCV function calls instead of old C language interface, and to avoid memory reallocation of internal structures in consecutive calls on images of the same size. NLOpt [27] library was used for the local optimization.

Randomness and normality tests were done in post-processing by using statistical software R. Stitched panoramas in full resolution are also a part of the electronic attachment.

7.1 Suitability of SIFT and ORB keypoints

SIFT and ORB were selected as potentially appropriate methods for keypoint detection. Table 7.2 lists average numbers of matched keypoints per overlap as results of the proposed keypoint matching pipeline. The computation time was measured in milliseconds on a computer with Intel i3 2.53GHz CPU and 3 GB RAM.

Dataset	SIFT		SIFT-rank		ORB	
	Avg. matches per overlap	Avg. comp. time	Avg. matches per overlap	Avg. comp. time	Avg. matches per overlap	Avg. comp. time
1	13.79	343.81	13.35	381.95	3	10.83
2	43.23	458.26	38.45	514.516	29.87	10.48
3	395.97	1542.46	397.219	1763.38	133.71	36.06
4	38.19	222.61	36.21	249.71	4.99	3.72
5	13.38	103.8	12.45	115.05	0	1.9
6	11.05	87.85	10.05	96.15	0	1.4

Table 7.2: Comparison of keypoint detection methods.

The computation of ORB took only a small fraction of SIFT computation time. However, ORB failed to compute any keypoints in datasets with small overlaps (dataset 5, 6), and only a small number of corre-

spondences were obtained in datasets 1 and 4. Generally, matching of ORB features provided sufficient number of matches in datasets featuring larger overlaps and featuring higher contrast objects (datasets 2, 3). On the other hand, results show the robustness of SIFT keypoints – the counts of matches were high enough in all the datasets, regardless the nature of image data or overlap size. Surprisingly, results of Toews and Wells [16] were not confirmed in the given datasets. Ranked SIFT descriptor resulted in better matching only in dataset 3. As a consequence, the next experiments was performed by using SIFT keypoint detection with standard descriptor, except for the dataset 3, where ORB was used. In practice, it is recommended to use ORB where possible because of the computation time.

7.2 Stitching by simple tile alignment

Each panorama was stitched together without applying any correction to the tiles (further referred as no-correction panoramas) , and with application of shifts between the overlapping regions found by the phase correlation method (further referred as shifts-only panoramas). Both are intended to serve for the comparison of the panoramas obtained by applying the proposed method. No-correction panoramas display the initial state and its qualities that are aimed to be improved. The panoramas obtained by correcting the misalignments only by using the shifts have very high visual quality, due to the fact that the tiles were captured on a flat surface, and that the shift vectors between the matched keypoints are uniform (similar if not equal distance and direction). The panoramas obtained by the proposed distortion correction method aim to reach at least similar visual qualities as the panoramas obtained only by considering shifts.

Tables 7.3 and 7.4 provides the quality measures of no-correction panoramas and shifts-only panoramas for each of the datasets.

Dataset	SSIM	MSSIM	Cost per correspondence	Median keypoint distance [μm]
1	0.6061	0.9803	1755.67	142.39
2	0.6596	0.9289	156.37	4.63
3	0.9254	0.9881	84.038	6.18
4	0.7504	0.9461	148.81	64.59
5	0.8140	0.9780	18.48	8.58
6	0.7844	0.9783	20.64	14.37

Table 7.3: Quality of no-correction panoramas.

Dataset	SSIM	MSSIM	Cost per correspondence	Median keypoint distance [μm]
1	0.8386	0.9945	190.06	12.20
2	0.9291	0.9852	10.22	0.41
3	0.9741	0.9965	3.07	0.96
4	0.8626	0.9720	43.85	6.20
5	0.9003	0.9873	7.51	2.14
6	0.8612	0.9871	5.77	3.93

Table 7.4: Quality of shifts-only panoramas.

7.3 The optimization results

7.3.1 Results obtained by enforcing the expected bounds

The purpose of this experiment was to check whether the expected intervals for each of the distortion parameters were sufficient to find a visually satisfying result. The opposite was found true, as can be seen in Table 7.5.

Table 7.5 shows that the visual quality of all panoramas is worse compared to shifts-only panoramas. It can be most notably seen in dataset 1, where the SSIM value is close to SSIM in case of its no-correction panorama. The main reason of this result is the fact that dataset 1 was captured at small magnification, therefore having large pixel size. $5\mu m$ stage shift bound corresponds to 1.47 pixels, which is not enough to correctly align the tiles. On the other hand, datasets 2 and dataset 3 can be considered of relatively good quality, even

7. EXPERIMENTS AND RESULTS

Dataset	Model	SSIM	MSSIM	Cost per correspondence	Median keypoint distance [μm]
1	T_2^{-1}	0.6583	0.9828	1605.02	136.94
2	T_1^{-1}	0.9248	0.9839	9.64	0.44
3	T_1^{-1}	0.9624	0.9942	13.53	1.53
4	T_1^{-1}	0.8378	0.9599	74.59	17.03
5	T_2^{-1}	0.8779	0.9707	9.42	3.06
6	T_2^{-1}	0.8190	0.9811	7.47	6.16

Table 7.5: Quality of solutions obtained by enforcing the expected bounds.

though in case of dataset 3, large uniform areas compensate for small misaligned blobs, so there is still room for improvement. Regarding distances between keypoints, they can be improved as well if larger bounds for the parameters are allowed.

7.3.2 Best results by using alternative bounds

The results presented in Table 7.6 were obtained by setting the bounds so that: $Err_{scale} = 5\%$, $Err_{rotation} = 1.2^\circ$, $Err_{nonortho} = 1.2^\circ$. The shift errors Err_{shift} were set individually for each dataset with respect the initial distances of matched keypoints. Table 7.7 presents the parameters of the best solutions.

Dataset	Model	SSIM	MSSIM	Cost per correspondence	Median keypoint distance [μm]
1	T_1^{-1}	0.8235	0.9888	20.25	10.82
2	T_1^{-1}	0.9218	0.9833	9.32	0.44
3	T_1^{-1}	0.9640	0.9944	10.24	2.16
4	T_1^{-1}	0.8584	0.9614	37.07	8.87
5	T_2^{-1}	0.8779	0.9707	9.42	3.06
6	T_2^{-1}	0.8366	0.9825	6.50	4.52

Table 7.6: Quality of solutions obtained by alternative bounds.

Dataset	Model	Global transform parameters					Medians of stage shift errors [μm]			
		s_x	s_y	φ	ψ_1	ψ_2	δ_{i,j_x}	δ_{i,j_y}	ε_{i,j_x}	ε_{i,j_y}
1	T_1^{-1}	0.950	0.952	-0.282°	0.761°	-0.618°	144.37	0.587	-1.46	124.58
2	T_1^{-1}	0.994	0.993	0.891°	0.240°	0.373°	-1.23	-0.01	0	-2.23
3	T_1^{-1}	0.992	0.990	0.839°	-1.145°	0.704°	6.02	0	0	5.62
4	T_1^{-1}	1.003	1.003	0.694°	0.015°	-0.181°	-25.09	0.21	1.95	-16.12
5	T_2^{-1}	0.998	0.997	-0.068°	-0.162°	0.054°	3.72	0	0	5
6	T_2^{-1}	0.996	0.997	-0.345°	0.059°	-0.207°	6.39	3.96	0.60	5.17

Table 7.7: Parameters of the best solutions.

According to the results, the expected intervals for the errors were sufficient in case of rotation for all datasets and in scaling and non-orthogonality for the majority of datasets. The scaling error in dataset 1 was up to 5%. Even though s_x value lies at the border of allowed interval, no better solution in terms of SSIM and MSSIM was found by allowing larger scale error. Non-orthogonality expectations were broken in case of dataset 3.

It is interesting to see that the best solution for dataset 5 is the intermediate solution in two-step optimization, enforcing all δ_{i,j_y} and ε_{i,j_x} to 0. However, because of the small size of the overlaps, keypoint detection and matching did not perform very well and it affect the stability of the solutions.

7.3.3 Comparison of model T_1^{-1} vs. T_2^{-1}

Two models were defined due to the issue of ordering non-orthogonality and scaling transform in the global distortion. The optimization results showed both models performed very well, with only marginal differences in terms of resulting SSIM and MSSIM. However, the parameters of the distortions are different. Two examples are provided for demonstration: Table 7.8 compares the best T_1^{-1} and T_2^{-1} solutions found for dataset 1, Table 7.9 compares the two best solutions found for dataset 2.

7. EXPERIMENTS AND RESULTS

Dataset	Global transform parameters					Medians of stage shift errors [μm]			
	s_x	s_y	φ	ψ_1	ψ_2	δ_{i,j_x}	δ_{i,j_y}	ε_{i,j_x}	ε_{i,j_y}
T_1^{-1}	0.950	0.952	-0.282°	0.760°	-0.618°	144.37	0.587	-1.46	124.58
T_2^{-1}	0.950	0.958	0.15°	0.327°	-0.187°	144.42	0.413	-1.46	126.78

Table 7.8: Dataset 1: Parameters of best solutions.

Derived global transform matrices:

$$T_1^{-1} = \begin{pmatrix} 0.950 & -0.006 & -1.811 \\ 0.008 & 0.952 & 1.781 \\ 0 & 0 & 1 \end{pmatrix},$$

$$T_2^{-1} = \begin{pmatrix} 0.950 & -0.006 & 0.960 \\ 0.008 & 0.958 & -0.957 \\ 0 & 0 & 1 \end{pmatrix}.$$

Solution	Global distortion parameters					Shift errors [μm]			
	s_x	s_y	φ	ψ_1	ψ_2	minimum	median	mean	maximum
T_1^{-1}	0.994	0.993	0.891°	0.240°	0.373°	-1.23	-0.01	0	-2.23
T_2^{-1}	0.999	0.998	0.216°	0.914°	-0.293°	-1.26	-0.02	0.01	-2.22

Table 7.9: Parameters of best solutions.

Derived global transform matrices:

$$T_1^{-1} = \begin{pmatrix} 0.994 & -0.009 & 5.944 \\ 0.020 & 0.993 & -5.860 \\ 0 & 0 & 1 \end{pmatrix},$$

$$T_2^{-1} = \begin{pmatrix} 0.998 & -0.009 & 1.457 \\ 0.020 & 0.998 & -1.421 \\ 0 & 0 & 1 \end{pmatrix}.$$

In both presented cases, the global distortion matrices are the same except for the translation terms. This result was quite expected – there is only one reality and both models try to adapt to it. Still, the majority of the best solutions (in terms of SSIM and MSSIM) per dataset were found by optimizing model T_1^{-1} .

7.3.4 One-step vs. two-step optimization

Two-step optimization was suggested in order to increase the impact of the global distortion correction transform and to reduce the influence of shifts. It was found that the results given by the two-step optimization process exhibited better visual qualities according to SSIM and MSSIM in general.

The difference between one-step and two-step solutions for dataset 2 is demonstrated in Table 7.10. Direct optimization solution has SSIM= 0.9017 and MSSIM= 0.97918, whereas two-step optimization solution has SSIM= 0.9218 and MSSIM= 0.9834.

Solution	Global distortion parameters					Shift errors [μm]			
	s_x	s_y	φ	ψ_1	ψ_2	minimum	median	mean	maximum
One-step	0.973	1.01	1.200°	-0.123°	-1.200°	-1.11	-0.22	-7.84	-2.23
Two-step	0.994	0.993	0.891°	0.240°	0.373°	-1.23	-0.01	0	-2.23

Table 7.10: Comparison of T_1^{-1} solutions obtained by one-step and two-step optimization.

Derived global transform matrices:

$$T_{1_{one-step}}^{-1} = \begin{pmatrix} 0.973 & -0.04 & 8.078 \\ 0.019 & 1.010 & -8.057 \\ 0 & 0 & 1 \end{pmatrix},$$

$$T_{1_{two-step}}^{-1} = \begin{pmatrix} 0.994 & -0.009 & 5.944 \\ 0.020 & 0.993 & -5.860 \\ 0 & 0 & 1 \end{pmatrix}.$$

It can be seen that the parameters and bounds are different and so are the global distortion correction matrices. One-step optimization result almost seems like if the global distortion parameters were chosen arbitrarily and stage shift errors were assigned to do the alignments. Note the median value for ε_{i,j_x} . Not only it is out of the expected error interval, but also it is much larger than the corresponding median value in two-step solution. Out of the best solutions presented in previous sections, only the solution for dataset 6 was found by using one-step optimization strategy. As a consequence, it is recommended to always use two-step strategy.

7.3.5 Normality of stage shift errors

The stage shift errors were expected to be random and small. The results were separately evaluated for each type of stage shifts (δ_{i,j_x} , δ_{i,j_y} , ε_{i,j_x} , ε_{i,j_y}). According to the Shapiro-Wilk normality test, the normality was rejected in most of the cases on the significance level 5%. Such results are marked by \times in Table 7.11. The shifts there whose normality was not rejected are marked by \checkmark . Visual comparison of the empirical and normal distribution is provided in Q-Q plots, which are available in the appendix.

	δ_{i,j_x}	δ_{i,j_y}	ε_{i,j_x}	ε_{i,j_y}	All together
1	\times	\checkmark	\checkmark	\times	\times
2	\times	\checkmark	\times	\times	\times
3	\times	\checkmark	\times	\times	\times
4	\times	\checkmark	\times	\times	\times
5	\checkmark	–	–	\times	\times
6	\times	\checkmark	\checkmark	\times	\times

Table 7.11: Normality of stage shift errors.

Randomness was performed by using Kolmogorov-Smirnov test and it was rejected for all the test-cases. It can be concluded that the stage shift errors are not random, nor normal.

7.4 Stability of solutions

The aim of this test was to verify whether the global distortion correction matrix estimated for data captured at the same device with the same settings are equal. The test was conducted on 2 largest datasets (datasets 3 and dataset 4). The datasets were divided into four quarters and the distortion correction method was run for each quarter separately. The result are illustrated in Table 7.12 for dataset 3 and in Table 7.13 for dataset 4.

The resulting parameters for the subsets of datasets are not equal. Nonetheless, the result is pretty interesting as in both cases there are

	s_x	s_y	φ	ψ_1	ψ_2
1	0.987	0.998	-0.365°	-0.247°	-0.180°
2	0.989	0.989	0.520°	-0.796°	0.361°
3	0.987	0.987	0.543°	-0.811°	0.396°
4	0.990	0.988	0.644°	-0.899°	0.489°

Table 7.12: Global distortion estimated for 4 distinct non-overlapping subsets of the dataset 3.

	s_x	s_y	φ	ψ_1	ψ_2
1	0.998	0.998	1.1°	-0.377°	0.233°
2	0.998	1.005	1.1°	-0.404°	0.219°
3	0.995	0.995	-0.15°	0.852°	-1.1°
4	0.999	0.995	-0.14°	0.853°	-1.1°

Table 7.13: Global distortion estimated for 4 distinct non-overlapping subsets of the dataset 4.

presented two similar results. In case of dataset 3, the first quarter converged to a different minima than the other three quarters and similarly, in case of dataset 4 there were 2 distinct solutions for two pairs of dataset quarters. The results show that there are more local minima and given the initial estimate of the minima obtained by the simulated annealing method, the local methods can succeed or fail to find the best solution. By comparing the quality of the partial solutions, for example in case of dataset 3, the solution cost per correspondence is 3.19 for the first solution and 3.04 for the fourth solution, so there is only a small difference between the two minima. Also, SSIMs and MSSIMs are close for both cases (0.977, resp. 0.996 in case of the first and 0.978, resp. 0.996 in case of the fourth solution for dataset 3).

7.5 Optical distortion estimation from a subset of tiles

It is interesting to see whether the global distortion correction transform can be correctly estimated only by using a small subset of tiles.

The experiment was performed over dataset 2. Even though already the first run for 2×2 subset of tiles delivered almost exactly the same global distortion parameters as in the previously demonstrated solution in Table 7.9, the consequent runs for subsets of tiles up to 3×3 tiles delivered different solutions. Rather than concluding the original question it can be claimed that the optimization cost function is probably very flat and has many minima of the same cost. The best solution does not have to be always returned from the method, thanks to the probabilistic global optimization algorithm. Further research can focus on trying different cost functions, that would for example penalize the distortion correction by shift parameters in favour of increasing the effect of the optical system distortion correction. Ideally, the resulting cost functions would not require two-step optimization and would be more steep.

8 Discussion

8.1 Distortion parameters

The results confirm the expert's constraints on global distortion parameters. The rotation was always found within the expected interval, non-orthogonality and scaling errors were larger only in one case each. The estimate on the distortion generated by specimen stage shifts was found to be wrong. Even though some dataset could be seamlessly assembled by using small-enough shift correction parameters, the datasets captured in smaller magnification could not. The reason is that there is always found some misalignment in pixels, but the size of the pixel in micrometres is larger in case of small magnification. It seems more reasonable to express the shift error bounds in the percentage of the real stage shift of the specimen stage. Moreover, the stage shifts were shown to be systematic, the test for randomness rejected the assumption and furthermore, the normality of shifts were also rejected in the majority of cases. Further research concerning the stage shifts is left to experts.

8.2 Precision of stitched panoramas

It was claimed that panoramas stitched by using the proposed method should have higher metrological precision; however, the experimental confirmation is missing in the Results section. Proper evaluation can only be done by experimenting with datasets of image tiles containing objects of known shape and size. Because no such datasets were available, metrological precision remains an open question to be answered in the following work.

The only thing that can be commented regarding the metrological precision is that the keypoint distances in all solutions were about 1 – 3 pixels. These are most probably the results of imprecise keypoint localisations, but if these were correctly localised and matched, they would generate metrological imprecisions. In terms of the absolute error in micrometres, the measurements in highly magnified datasets would be more precise compared to measurements in panoramas captured at low magnification.

8.3 Note on quality evaluation

Quality of stitched panoramas was evaluated by SSIM and MSSIM. SSIM is designed so that it evaluates luminance, contrast and structure of the overlaps together. Because no attention was paid for example to brightness compensation in the overlaps, a better evaluation technique would be based on considering only the details in the images, such as edges. A work presented by Qureshi et al. [36] distinguishes between the geometric and photometric qualities of stitched panoramas. For geometric quality evaluation, the authors propose computing SSIM between the high-frequency components of a panorama and a tile in 32×32 non-overlapping windows. The evaluation of the proposed method by using Qureshi's approach can be done in the consecutive research.

9 Conclusion

The thesis focused on the problem of estimating imaging distortion generated in the optical system of Scanning Electron Microscopes, and by imprecise specimen stage positioning. The global distortion correction is dominated by rotation, scaling, and non-orthogonality. The ambiguity concerning the order of scaling and non-orthogonality led to definition of two global distortion models. The global distortion parameters together with specimen stage positioning errors were estimated by solving a constrained optimization problem. The optimization cost function was defined as a sum of squared distances between the pairs of corresponding keypoints, which were detected by using SIFT or ORB feature detection methods in the overlapping regions of the image tiles. Keypoint matching was performed by a brute-force algorithm, with several filtering steps to avoid false matches.

The optimization was performed by Simulated Annealing, followed by a committee of local non-linear programming methods to ensure convergence to minima. Because the optimization of all variables together reduces the impact of the global correction, a two-step optimization was proposed to maximize the effect of the global transform on the alignment.

The optimization constraints were preliminarily given by an expert so that they reflected the physical limits of the acquisition device. However, the assumptions regarding maximal size of the errors needed to be checked.

The experiments were performed on six datasets with different properties. The quality of the solution was evaluated by using structural similarity index. It was found that the distortion model considering non-orthogonality prior to scaling (and rotation) performs better on the given datasets, even though the difference is minor. The error bounds on rotation error (max 1°), non-orthogonality (max 1°) and scaling error (max 1%) were confirmed in majority of the datasets, although both non-orthogonality and scaling constraints were broken in one case. It was found that the stage shift errors are much larger than expected in datasets captured in low magnification. Moreover, the normality and randomness assumptions were statistically rejected.

9. CONCLUSION

There are plenty of open questions for further research. Firstly, stage shift errors need to be investigated to explain why they are not normal and random. Also, they need a better quantification, for example respecting the zoom level at which the datasets were captured. Secondly, the optimization cost function is not ideal, as it contains possibly many local minima of similar cost. Other optimization cost functions, and other optimization algorithms can be used in order to improve the stability of the results. Finally, the metrological precision should be tested and evaluated on appropriate datasets.

Bibliography

- [1] Verena Kaynig et al. "Fully automatic stitching and distortion correction of transmission electron microscope images". In: *Journal of structural biology* 171.2 (2010), pp. 163–173.
- [2] Joseph Goldstein et al. *Scanning electron microscopy and x-ray microanalysis*. 3rd ed. New York: Springer, 2003. ISBN: 978-0-306-47292-3.
- [3] Max Knoll. "Aufladepotential und sekundäremission elektrodenbestrahlter körper". In: *Z. tech. Phys* 16 (1935), pp. 467–475.
- [4] Agnès Bogner et al. "A history of scanning electron microscopy developments: towards "wet-STEM" imaging". In: *Micron* 38.4 (2007), pp. 390–401.
- [5] Australian Microscopy & Microanalysis Research Facility. *MyScope: training for advanced research*. Feb. 13, 2017. URL: <http://www.ammrf.org.au/myscope/sem/practice/principles/layout.php>.
- [6] Ray F Egerton. *Physical principles of electron microscopy*. Springer, 2005.
- [7] Australian Microscopy & Microanalysis Research Facility. *MyScope: training for advanced research*. Apr. 13, 2017. URL: <http://www.ammrf.org.au/myscope/sem/practice/principles/lenses.php>.
- [8] Junichi Nakamura. *Image sensors and signal processing for digital still cameras*. CRC press, 2016.
- [9] Paul van Walree. *Distortion*. Apr. 17, 2017. URL: <http://toothwalker.org/optics/distortion.html>.
- [10] Richard Szeliski. "Image alignment and stitching: A tutorial". In: *Foundations and Trends® in Computer Graphics and Vision* 2.1 (2006), pp. 1–104.
- [11] David G Lowe. "Distinctive image features from scale-invariant keypoints". In: *International journal of computer vision* 60.2 (2004), pp. 91–110.
- [12] Herbert Bay, Tinne Tuytelaars, and Luc Van Gool. "Surf: Speeded up robust features". In: *Computer vision—ECCV 2006* (2006), pp. 404–417.

BIBLIOGRAPHY

- [13] Motilal Agrawal, Kurt Konolige, and Morten Rufus Blas. “Censure: Center surround extremas for realtime feature detection and matching”. In: *European Conference on Computer Vision*. Springer. 2008, pp. 102–115.
- [14] Ethan Rublee et al. “ORB: An efficient alternative to SIFT or SURF”. In: *Computer Vision (ICCV), 2011 IEEE International Conference on*. IEEE. 2011, pp. 2564–2571.
- [15] Gary Bradski et al. “The opencv library”. In: *Doctor Dobbs Journal* 25.11 (2000), pp. 120–126.
- [16] Matthew Toews and William Wells. “Sift-rank: Ordinal description for invariant feature correspondence”. In: *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*. IEEE. 2009, pp. 172–177.
- [17] Edward Rosten and Tom Drummond. “Machine learning for high-speed corner detection”. In: *Computer vision–ECCV 2006 (2006)*, pp. 430–443.
- [18] Michael Calonder et al. “Brief: Binary robust independent elementary features”. In: *Computer Vision–ECCV 2010 (2010)*, pp. 778–792.
- [19] Martin A Fischler and Robert C Bolles. “Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography”. In: *Communications of the ACM* 24.6 (1981), pp. 381–395.
- [20] Monique Laurent. “Semidefinite programming in combinatorial and polynomial optimization”. In: *Nieuw Arch. Wiskd.(5)* 9.4 (2008), pp. 256–262.
- [21] Pablo A Parrilo. “Structured semidefinite programs and semi-algebraic geometry methods in robustness and optimization”. PhD thesis. California Institute of Technology, 2000.
- [22] Jean B Lasserre. “Global optimization with polynomials and the problem of moments”. In: *SIAM Journal on Optimization* 11.3 (2001), pp. 796–817.
- [23] Peter Wittek. “Algorithm 950: Ncpol2Sdpa—sparse semidefinite programming relaxations for polynomial optimization problems of noncommuting variables”. In: *ACM Transactions on Mathematical Software (TOMS)* 41.3 (2015), p. 21.

- [24] Katsuki Fujisawa et al. "SDPA (SemiDefinite Programming Algorithm) User's Manual–Version 6.2. 0". In: *Department of Mathematical and Computing Sciences, Tokyo Institute of Technology. Research Reports on Mathematical and Computing Sciences Series B: Operations Research* (2002).
- [25] Pablo A Parrilo. "Semidefinite programming relaxations for semialgebraic problems". In: *Mathematical programming* 96.2 (2003), pp. 293–320.
- [26] Scott Kirkpatrick, C Daniel Gelatt, Mario P Vecchi, et al. "Optimization by simulated annealing". In: *science* 220.4598 (1983), pp. 671–680.
- [27] SG Johnson. *The NLOpt nonlinear-optimization package [Software]*. 2017. URL: <http://ab-initio.mit.edu/wiki/index.php/NLOpt>.
- [28] Michael JD Powell. "A direct search optimization method that models the objective and constraint functions by linear interpolation". In: *Advances in optimization and numerical analysis*. Springer, 1994, pp. 51–67.
- [29] Dieter Kraft. "A software package for sequential quadratic programming". In: *Forschungsbericht- Deutsche Forschungs- und Versuchsanstalt fur Luft- und Raumfahrt* (1988).
- [30] Krister Svanberg. "A class of globally convergent optimization methods based on conservative convex separable approximations". In: *SIAM journal on optimization* 12.2 (2002), pp. 555–573.
- [31] Ernesto G Birgin and Jose Mario Martinez. "Improving ultimate convergence of an augmented Lagrangian method". In: *Optimization Methods and Software* 23.2 (2008), pp. 177–195.
- [32] Zhou Wang et al. "Image quality assessment: from error visibility to structural similarity". In: *IEEE transactions on image processing* 13.4 (2004), pp. 600–612.
- [33] S. S. Shapiro and M. B. Wilk. "An analysis of variance test for normality (complete samples)†". In: *Biometrika* 52.3-4 (1965), p. 591. doi: 10.1093/biomet/52.3-4.591.
- [34] Frank J Massey Jr. "The Kolmogorov-Smirnov test for goodness of fit". In: *Journal of the American statistical Association* 46.253 (1951), pp. 68–78.

BIBLIOGRAPHY

- [35] Rob Hess. "An open-source SIFTLibrary". In: *Proceedings of the 18th ACM international conference on Multimedia*. ACM. 2010, pp. 1493–1496.
- [36] HS Qureshi et al. "Quantitative quality assessment of stitched panoramic images". In: *IET Image Processing* 6.9 (2012), pp. 1348–1358.

A Stitched panoramas



Figure A.1: Dataset 1

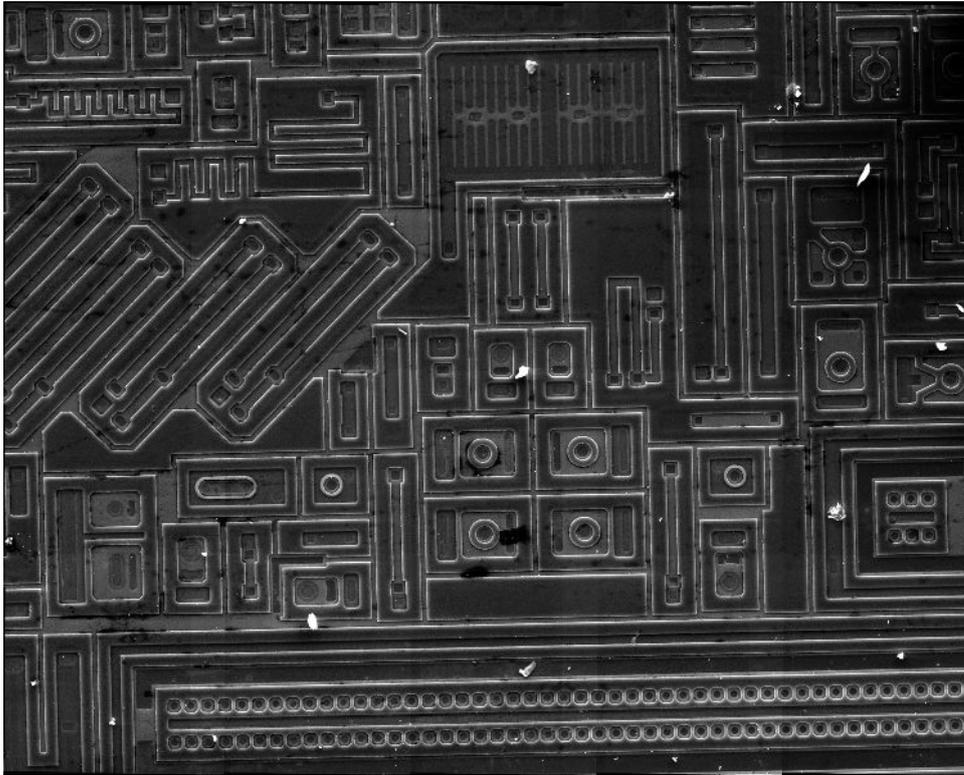


Figure A.2: Dataset 2

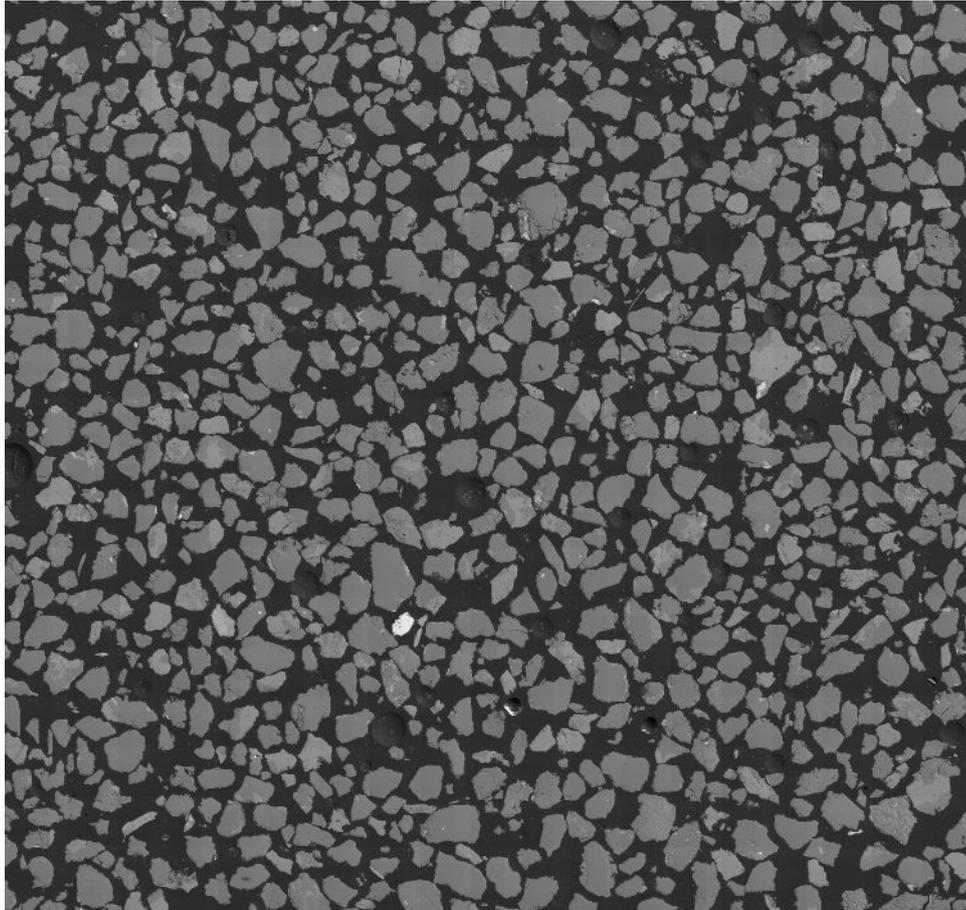


Figure A.3: Dataset 3

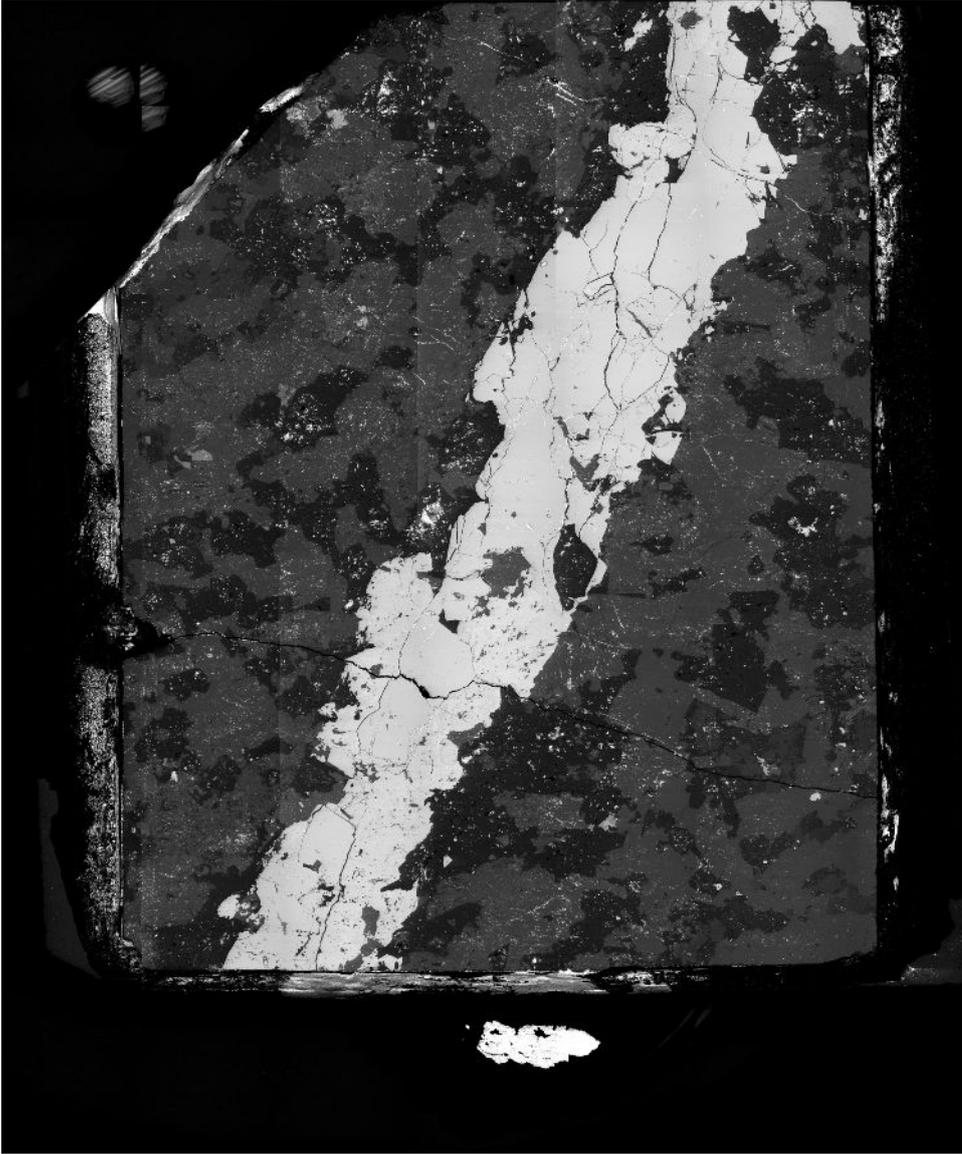


Figure A.4: Dataset 4

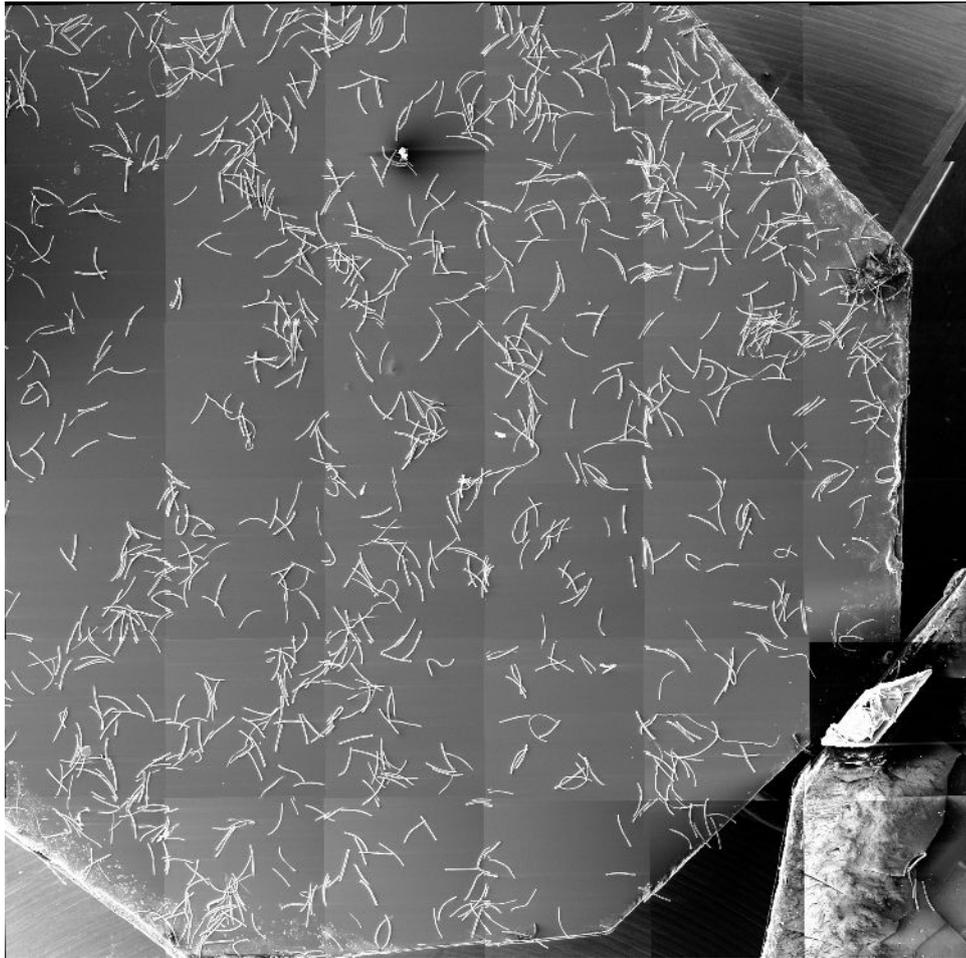


Figure A.5: Dataset 5

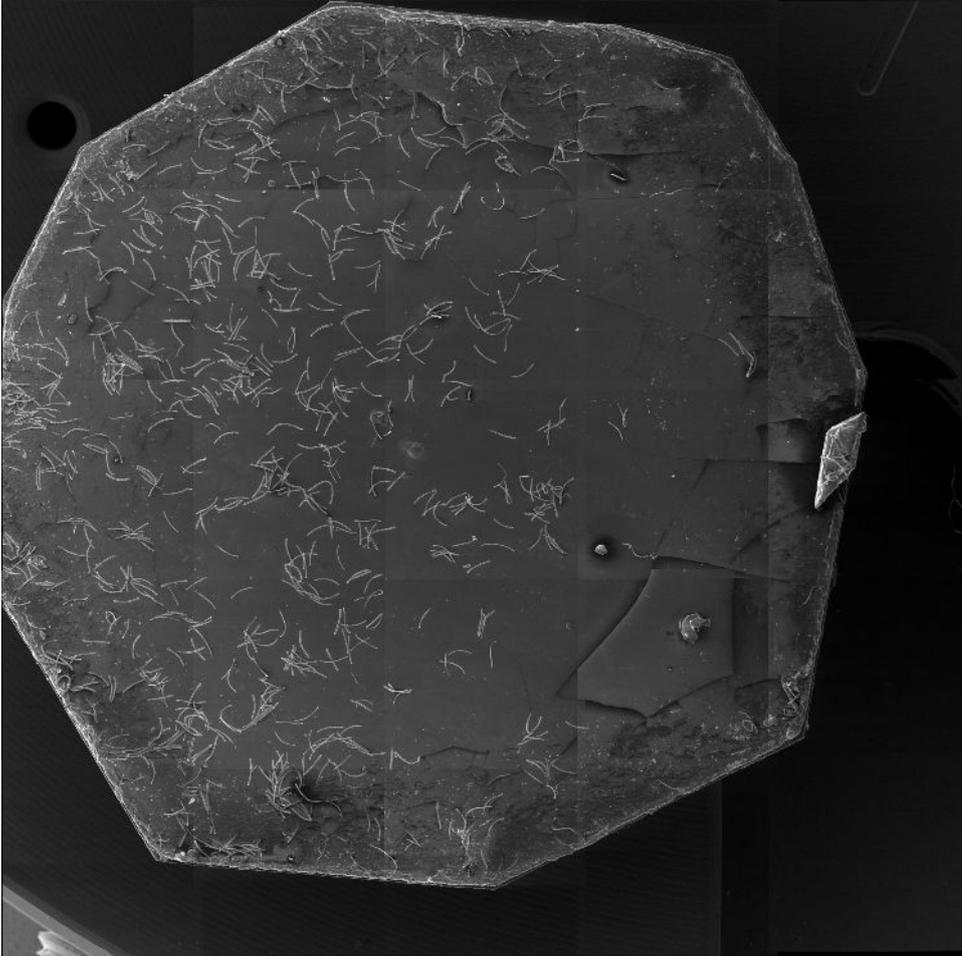
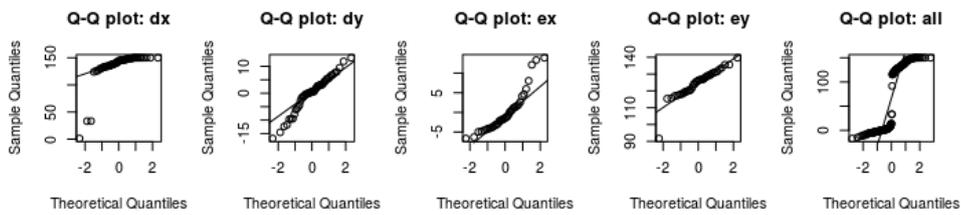


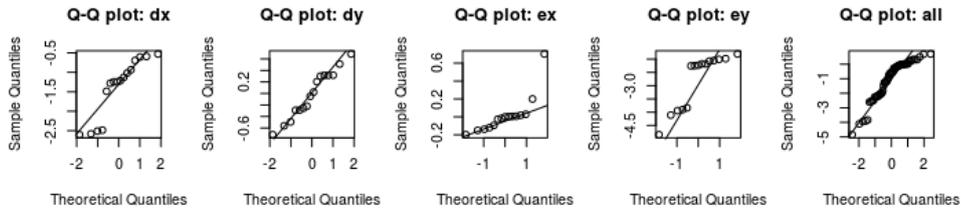
Figure A.6: Dataset 6

B Q-Q plots for stage shift errors

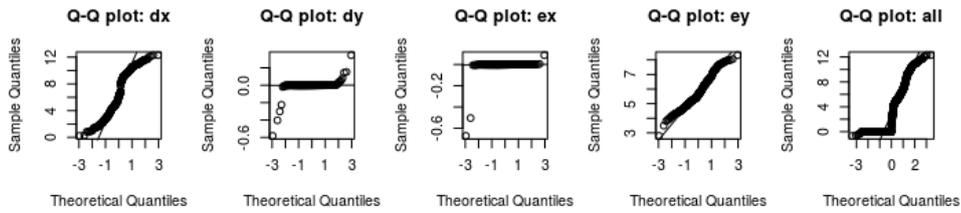
Figure B.1 presents the Q-Q plots for comparison between the stage shift errors estimation in the best solution for each overlap and theoretical normal distribution. The closer the dots follow the diagonal of each graph, the higher chance they have the are normal. A separate Q-Q plot is available for each of δ_{i,j_x} , δ_{i,j_y} , ε_{i,j_x} and ε_{i,j_y} .



(a) Q-Q plot for the best solution for dataset 1.

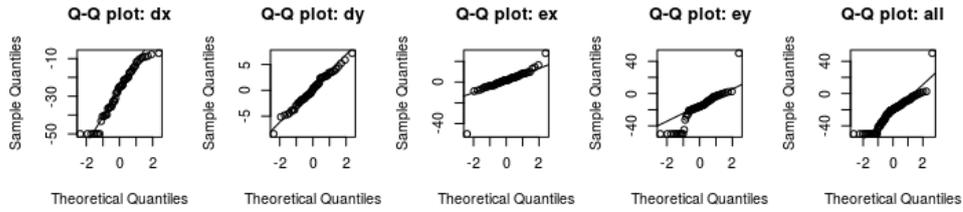


(b) Q-Q plot for the best solution for dataset 2.

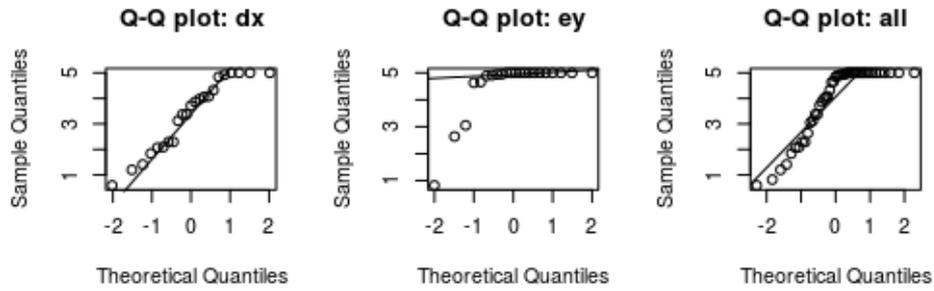


(c) Q-Q plot for the best solution for dataset 3.

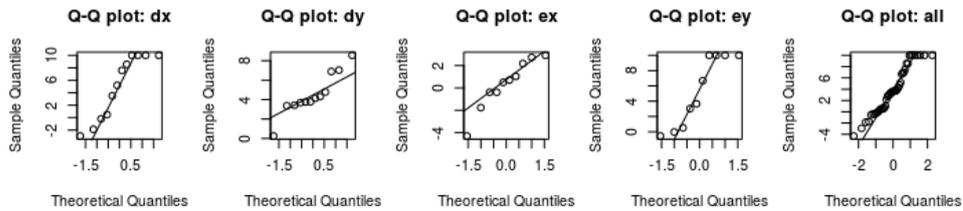
B. Q-Q PLOTS FOR STAGE SHIFT ERRORS



(d) Q-Q plot for the best solution for dataset 4.



(e) Q-Q plot for the best solution for dataset 5.



(f) Q-Q plot for the best solution for dataset 6.

Figure B.1: Comparison of empirical distribution of stage shift positioning errors and normal distribution in Q-Q plots.

C Brief description of the application framework

The implementation is decomposed into several classes:

- **DataSet**: the main class for manipulation with datasets,
- **FileNameFormat**: the class for parsing and composition of tiles' filenames,
- **Keypoint**: a structure encapsulating the position of a keypoint in a tile and in an overlap,
- **Matcher**: static class providing the methods for keypoint matching,
- **NLoptOptimization**: the class encapsulating NLopt library calls, suited for optimization of **StitchingOptimalityProblem**,
- **OptimizationProblem**: an interface for optimization problems solvable for example by **SimulatedAnnealing** class,
- **Solution**: a solution to the stitching problem,
- **Stitcher**: a static class proving the methods for panorama stitching and drawing of the correspondences on top of panoramas,
- **StitchingOptimalityProblem**: implements the optimization problem, including evaluation of the cost function and generation of new solutions,
- **TileData**: image data and metadata of tiles.

Separate applications can be build on top of this framework. Please find more information in the electronic attachment.