

Stephan Allgeier*, Klaus-Martin Reichert, Ekaterina Sorokovaia, Sebastian Bohn, Ralf Mikut, Oliver Stachs, and Karsten Sperlich

Evaluation strategy for image acquisition protocols in confocal microscopy mosaic imaging of the cornea

<https://doi.org/10.1515/cdbme-2025-0142>

Abstract: In vivo confocal laser-scanning microscopy (CLSM) of the cornea provides image data from the corneal tissues with cellular resolution. The ability to resolve the nerve fiber bundles of the sub-basal nerve plexus (SNP), in particular, opens the opportunity to use the SNP morphometry as a readily accessible, image-based quantitative biomarker for various neuropathic diseases. Several techniques have been proposed to overcome the small field of view of CLSM systems by capturing a dataset of many partially overlapping images covering an extended SNP area and creating a mosaic image from this dataset. So far, these mosaicking techniques have only been assessed qualitatively, by subjective visual inspection of the results. The present contribution describes an objective, quantitative evaluation strategy for these CLSM image montaging approaches by comparing the mosaicking results against ground truth data. Both the ground truth data and experimental datasets are acquired in a controlled experimental setup. Based on datasets from different specimens and different motion patterns, we benchmark our own mosaicking algorithm and report an "average of geometric distance" error value of ≤ 4.81 pixels. We provide the experimental data so that other researchers can evaluate and compare their own algorithms using the same base data.

Keywords: confocal laser-scanning microscopy, image registration, mosaicking, algorithm evaluation, open access data

1 Introduction

The transparent nature of all elements along the pathway that incident light has to traverse from entering the human eye until it is detected at the retina—i.e. the cornea, anterior

chamber, lens, and vitreous body—is crucial for our visual perception. This fact not only allows us to see our environment, but also makes these biological structures accessible for light-based imaging techniques. Confocal laser-scanning microscopy (CLSM) has been used in medical research and clinical practice for in vivo imaging of the human cornea for more than two decades. Even though newer modalities, such as optical coherence tomography (OCT), have since then received increasing attention, the lateral resolution of corneal CLSM of $\approx 1 \mu\text{m}$ has not been achieved by other techniques.

The corneal sub-basal nerve plexus (SNP) has drawn particular interest in clinical research, as this densely interconnected network of peripheral nerve fibers is readily accessible for in vivo CLSM imaging. In contrast to the broad range of widely used nerve function tests, CLSM opens the opportunity for quantitative assessment of nerve fiber morphology. SNP morphometry assessed by in vivo CLSM has been studied and proposed as an early biomarker for the diagnosis of neuropathic diseases such as diabetic neuropathy, chemotherapy-induced neuropathy, or Parkinson's disease [1]. It may also play a role as a sensitive surrogate marker to assess the efficacy of therapeutic approaches.

A key challenge of in vivo CLSM of the SNP is the small field of view (FOV); for the devices usable for corneal imaging, the FOV size is usually $400 \times 400 \mu\text{m}^2$ or smaller. It has been established that this area is insufficient for reliable morphometry results. To address this challenge, several methods have been developed to collect datasets of partially overlapping CSLM images of the SNP and to create large-area SNP montages [2, 3]. These mosaicking algorithms, including our own [4], are usually only evaluated qualitatively by visual inspection of the resulting montages.

The aim of this contribution is to bridge this gap. We describe a methodical framework for the quantitative evaluation of mosaicking techniques for datasets, apply it to our CLSM mosaicking algorithm, and present the results. We also describe and provide the datasets that we use to offer other research groups in the application domain the opportunity to evaluate their own techniques on the same basis and, in this way, establish the foundation for objective comparability among competing approaches.

*Corresponding author: **Stephan Allgeier**, Karlsruhe Institute of Technology (KIT), Institute for Automation and Applied Informatics, Karlsruhe, Germany, e-mail: stephan.allgeier@kit.edu

Klaus-Martin Reichert, Ekaterina Sorokovaia, Ralf Mikut, Karlsruhe Institute of Technology (KIT), Institute for Automation and Applied Informatics, Karlsruhe, Germany

Sebastian Bohn, Oliver Stachs, Karsten Sperlich, Rostock University Medical Center, Department of Ophthalmology, Rostock, Germany

2 Materials and Methods

2.1 Data Description

The experimental data contains datasets from four specimens:

- SiliconeTiO2: a transparent silicone volume with differently-sized TiO₂ particles embedded,
- PorcineEye1: a dead porcine eye,
- PorcineEye2: a second dead porcine eye,
- PorcineCornea2: the cornea trephined from PorcineEye2.

All four specimens were fixated in specific, custom-made holding fixtures, which were mounted on a motorized x-y-translation stage (8MTL20XY-S, Standa, Vilnius, Lithuania) for image acquisition. An HRTII (Heidelberg Engineering, Heidelberg, Germany) with an RCM2.0 (research prototype, Rostock University Medical Center) [5] was used for CLSM imaging. This setup has a FOV of $350 \times 350 \mu\text{m}^2$. According to the imaging protocol for human eyes, a protective single-use cap covered the RCM2.0 objective lens. A self-developed C# software program controlled the image acquisition process and the movement pattern of the translation stage.

The experiments included three different movement patterns, one for the ground truth data acquisition and two for the evaluation datasets. The ground truth data of each specimen encompasses an area of approximately $2.38 \times 2.38 \text{ mm}^2$ that was imaged by a rectilinearly arranged grid of 8×8 single CLSM images, spaced $290 \mu\text{m}$ so that adjacent images overlap by $60 \mu\text{m}$. The movement between grid positions happened in between the recording of the single images, while no movement was exerted during the actual acquisition of the images.

Two predefined movement patterns were used to record the experimental data: An extending spiral pattern and a simulated fixational eye movements pattern. The spiral pattern simulates our in vivo imaging technique, which uses a moving fixation target on a display device in front of the contralateral eye of the examined person to guide their eye movements. The spiraling movement starts at a central position and then expands outwards until the spiral radius reaches 1 mm. The simulated fixational eye movements pattern with a length of 30 s was generated using the open-source FixEM simulation software [6]. The movement patterns were used on the four specimens with different numbers of repetitions (see Tab. 1).

2.2 Evaluation Method

First, the ground truth mosaic images are created by placing the static images at their intended grid positions and then correcting the alignment based on the results of phase correla-

Tab. 1: Repetition counts per specimen and movement pattern

Specimen	ground truth	spiral	fixation
SiliconeTiO2	2	6	1
PorcineEye1	1	3	0
PorcineEye2	1	1	10
PorcineCornea2	1	1	10

tion registration of the overlapping areas. This correction is required because the translation stage axes and the microscope image axes are not perfectly aligned. With their final positioning known, the image frames are stitched together with a weighted averaging in the overlapping areas. The test datasets from the spiral and fixational movement pattern experiments are all processed with our mosaicking software, which performs motion artifact correction and creates the test mosaic images. As a further preliminary step, we align each test mosaic with the corresponding ground truth mosaic, using the phase correlation function to estimate a translation between both images. Nothing else but a translation is used for alignment, as the specimens were not moved in between the ground truth data acquisition and the test dataset acquisition.

Finally, we use a metric called “average of geometric distance” [7] for the quantitative comparative evaluation of the mosaicking results. If I_{test} is a test mosaic and I_{gt} is the corresponding ground truth mosaic, we define a set of control points $\vec{x}_i$ arranged in a regular grid over the overlapping area of I_{test} and I_{gt} . The control points are then tracked in I_{test} using a Kanade-Lucas-Tomasi feature tracker. If $\delta_i = \|\vec{x}'_i - \vec{x}_i\|_2$ is the geometric distance between $\vec{x}_i$ and the tracked points $\vec{x}'_i$, the error metric is defined as

$$\epsilon_{\text{agd}} = \frac{1}{L} \sum_i \delta_i, \quad (1)$$

where L is the number of (successfully) tracked control points. In addition to ϵ_{agd} , we also examine the local distribution of geometric errors. Except for the mosaicking algorithm under test, the evaluation procedure has been implemented in Matlab.

3 Results

When processing the test datasets, it turned out that the mosaicking algorithm could not successfully handle the strong motion artifacts in the single SiliconeTiO2 fixation dataset. This dataset therefore had to be excluded from further analysis. Furthermore, analyzing the mosaicking results, we found that during the acquisition of the PorcineEye datasets (for both eyes), the in-focus-tissue level continuously changed over the course of minutes. The reason for this effect is not certain, but the most plausible explanation is a continuous decrease

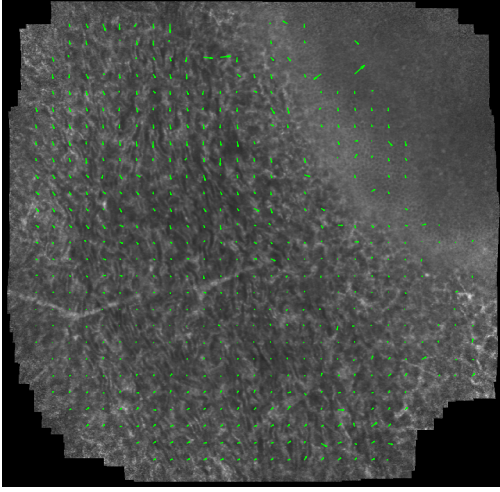


Fig. 1: Geometric error measured by control point displacements in a mosaic reconstructed from a test dataset (PorcineCornea2, fixation, run 3). The control point displacement ($\epsilon_{\text{agd}} = 4.78$) is measured with respect to a ground truth mosaic (not shown).

Tab. 2: Control point displacement, separated by coordinate

Specimen (pattern)	ϵ_{agd}	x -coordinate	y -coordinate
SiliconeTiO2 (spiral)	3.18	-0.4 ± 2.0	-1.2 ± 2.8
PorcineCornea2 (spiral)	4.20	0.1 ± 4.5	-0.6 ± 5.7
PorcineCornea2 (fixation)	4.81	1.5 ± 3.5	0.1 ± 4.5

of the intraocular pressure. As a consequence, the PorcineEye test datasets could not be aligned with their respective ground truth data and had to be excluded as well. In conclusion, the spiral datasets of the SiliconeTiO2 specimen and the datasets of the PorcineCornea2 specimen could be further evaluated.

Figure 1 shows the control point displacements measured in an exemplary fixation movement pattern dataset from the PorcineCornea2 specimen. An average of geometric distance metric of $\epsilon_{\text{agd}} = 4.78$ pixels was calculated for this specific dataset (PorcineCornea2, fixation, run 3). Mean error values for all evaluated repetitions are given in Tab. 2.

Figure 2 shows a histogram of the control point displacements measured in an exemplary spiral movement dataset from the SiliconeTiO2 specimen, separated by their x - and y -components. For this dataset (SiliconeTiO2, spiral, run 5), the control point displacements have a mean $\pm$ standard deviation of -0.7 ± 2.0 pixels for their x -components and of -1.2 ± 3.4 pixels for their y -components. Table 2 lists the respective average displacement components for all evaluated datasets.

Further examination of the distribution of displacement errors reveals that there is a strong local dependency. Figure 3 plots the distribution of error x - and y -components with respect to the control point position x - and y -coordinates. The plot reveals a strong linear dependency of the x -error compo-

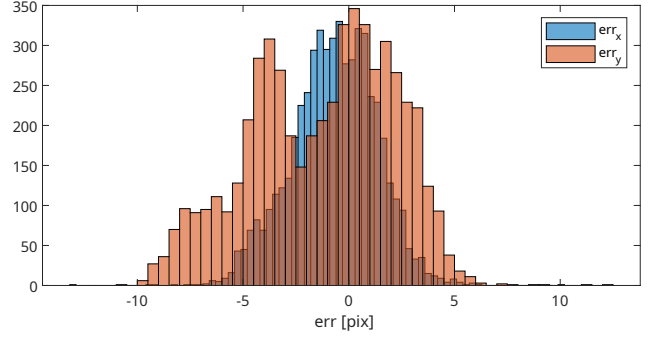


Fig. 2: Distribution of displacement errors from a test dataset (SiliconeTiO2, spiral, run 5), separated by x - and y -component.

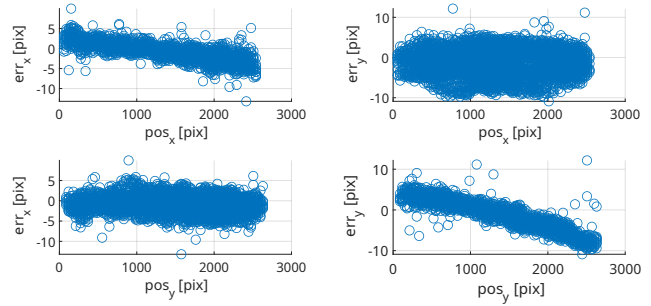


Fig. 3: Local dependency of displacement error components of a test dataset (SiliconeTiO2, spiral, run 5).

nent with the x -coordinate of the corresponding control point and a similar linear dependency of the y -error component with the y -coordinate of the corresponding control point. The other combinations show no discernible pattern. The shown figure is representative of all spiral datasets. The fixation datasets exhibit the same general characteristics.

4 Discussion

When reviewing the literature on quantitative evaluation of image mosaicking techniques, the majority uses pixel intensity-based metrics such as the mean square error (MSE) [7] or the structural similarity (SSIM) [8] to measure the differences between the mosaic image and the reference ground truth data. However, such measures convey little about the geometrical accuracy of the underlying image registration algorithms or of the resulting mosaic image. Since morphometric SNP features that are frequently measured in CLSM images (nerve fiber length and tortuosity) and also features such as the nerve migration rate (measured from aligned mosaic images from examinations at different time points) [9] all strongly depend on the geometric representation of the SNP, we use a feature point-based comparison metric in the present contribution.

The measured average geometric displacement of up to 4.81 pixels needs to be appraised with the downstream application of morphometric assessment of the SNP in mind. The decision of whether this error level is tolerable or not needs to remain with the researchers actually using the method. Irrespective of this yes-or-no question, the availability of this quantitative measure can finally put this decision on a firm objective foundation and in perspective also allows the assignment of uncertainty measures to morphometric SNP parameters extracted from the created mosaic images.

Furthermore, the local dependency of the displacement error from the control point position (see Fig. 3) is an interesting result of these examinations. The correlation between the displacement error and control point y -components can be explained plausibly by taking a closer look at the characteristics of the motion artifacts in CLSM images. Vertical movement (with respect to the image coordinate system) during the imaging process introduces vertical scaling effects (vertical stretching or compression of the image data). Every attempt to correct these artifacts needs to determine the correct amount of vertical scaling correction. The analysis of image pairs can only ever find the *relative* vertical scaling of one image with respect to another; the *absolute* global reference is inherently unknown. In effect, this missing global reference is a degree of freedom needs to be resolved by a regularization term. The observed correlation in the y -components might therefore be an indication of a non-perfect regularization. Further research is needed into the details of the regularization step in our mosaicking algorithm. However, the observed error of approx. 4 pixels at the upper parts and of approx. -8 pixels at the lower parts of a mosaic image also needs to be regarded in the context of the vertical extent of the mosaic of 2500 pixels.

The correlation between the displacement error and control point x -components, on the other hand, is not so easily explained. There is also an inherent degree of freedom with respect to global reference for horizontal motion artifacts. In analogy to the argument above for the y -component, the correlation of the x -components would point to a global horizontal scaling effect, but this is not what horizontal motion artifacts look like. Instead, horizontal motion during the CLSM image acquisition process introduces horizontal shearing effects, and such a residual motion artifact component in the mosaic image would manifest as a linear dependency between the x -component of the error displacements and the y -coordinate of the respective control point. However, such a correlation is not observed. For the moment, the nature of the correlation between the displacement error and control point x -components needs to remain unsolved and requires further research.

Overall, we have presented an experimental quantitative evaluation of our mosaicking algorithm for CLSM datasets of the human cornea. To our knowledge, this is the first time that

the mosaicking error has been reported in a quantitative way in this application domain. We make the experimental data and the evaluation methods freely available to the scientific community for others to use with their own mosaicking approaches [10]. This paves the way for objective and quantitative comparisons among mosaicking algorithms in this domain. As an even further step, we also plan to release our mosaicking algorithm as open source software in the near future.

Author Statement

Research funding: This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – project number 469107515. Conflict of interest: Authors state no conflict of interest. Informed consent: Not applicable. Ethical approval: Not applicable.

References

- [1] Cruzat A, Qazi Y, Hamrah P. In vivo confocal microscopy of corneal nerves in health and disease. *Ocul Surf* 2017;15:15–47.
- [2] Edwards K, Pritchard N, Gosschalk K, Sampson GP, Russell A, Malik RA, et al. Wide-field assessment of the human corneal subbasal nerve plexus in diabetic neuropathy using a novel mapping technique. *Cornea* 2012;31:1078–1082.
- [3] Qiao Y, Li G, Li F, Zhang C. NerveStitcher2.0: Evolution of stitching algorithm for corneal confocal microscope images with optical flow. In: *Proceedings of the 2023 8th International Conference on Biomedical Imaging, Signal Processing*. ACM; 2023:46–54.
- [4] Allgeier S, Bartschat A, Bohn S, Guthoff RF, Hagenmeyer V, Kornelius L, et al. Real-time large-area imaging of the corneal subbasal nerve plexus. *Sci Rep* 2022;12:2481.
- [5] Bohn S, Sperlich K, Allgeier S, Bartschat A, Prakasam R, Reichert KM, et al. Cellular in vivo 3D imaging of the cornea by confocal laser scanning microscopy. *Biomed Opt Express* 2018;9:2511–2525.
- [6] Allgeier S, Anzlinger F, Bohn S, Mikut R, Neumann O, Reichert KM, et al. An open-source software for the simulation of fixational eye movements. *Curr Dir Biomed Eng* 2024;10:25–28.
- [7] Azzari P, Di Stefano L, Mattoccia S. An evaluation methodology for image mosaicing algorithms. In: *Blanc-Talon J, Bourennane S, Philips W, Popescu D, Scheunders P, editors. Advanced Concepts for Intelligent Vision Systems*. Berlin, Heidelberg: Springer; 2008:89–100.
- [8] Wang Z, Bovik AC, Sheikh HR, Simoncelli EP. Image quality assessment: From error visibility to structural similarity. *IEEE Trans Image Process* 2004;13:600–612.
- [9] Edwards K, Pritchard N, Poole C, Dehghani C, Al Rashah K, Russell A, et al. Development of a novel technique to measure corneal nerve migration rate. *Cornea* 2016;35:700–705.
- [10] Allgeier S, Reichert KM, Sorokovaia E, Bohn S, Sperlich K. Evaluation strategy for image acquisition protocols in confocal microscopy mosaic imaging of the cornea [Data set]. Zenodo 2025. doi:10.5281/zenodo.15116181.