



Master *Color in Science and Industry* (COSI)



MEASURING AND MITIGATING SPECKLE NOISE IN DUAL-AXIS CONFOCAL MICROSCOPY IMAGES

Master Thesis Report

Presented by

Davit Gigilashvili

and defended at the

University Jean Monnet, Saint Etienne

7th of September 2017

Academic Supervisor(s): Prof. Dr. Jon Yngve Hardeberg, Assoc. Prof. Dr. Marius Pedersen

Host Co-supervisor (if applicable): Assoc. Prof. Dr. Jonathan T.C. Liu

Jury Committee:

Prof. Dr. Markku-Hauta Kasari

Dr. Ville Heikkinen

Norwegian Colour and Visual Computing Laboratory (Colourlab)
Norwegian University of Science and Technology

Faculty of Mechanical Engineering
University of Washington

Measuring and Mitigating Speckle Noise in Dual-Axis Confocal Microscopy Images

Master's Thesis

Author:	Davit Gigilashvili
Academic Supervisor(s):	Prof. Dr. Jon Yngve Hardeberg Assoc. Prof. Dr. Marius Pedersen
Host Co-supervisor:	Assoc. Prof. Dr. Jonathan T.C. Liu
Jury Committee:	Prof. Dr. Markku-Hauta Kasari Dr. Ville Heikkinen
Submission Date:	21 August, 2017

Abstract

Speckle noise is a form of multiplicative noise that corrupts the quality of medical images. It is well described and studied in medical ultrasound imaging, but less attention has been paid to its presence in reflectance microscopy images. Presence of the speckle noise not only limits the application of further post-processing and computer vision techniques, like edge detection, it also makes diagnosing more difficult and less reliable for the physicians.

Many speckle mitigation techniques have been studied by various researchers, but the vast majority of them limit itself just to single image of the target tissue. While averaging of uncorrelated images of the same tissue taken from different spatial conditions is mentioned among the possibilities of speckle mitigation, it is considered too complicated due to the need of several image acquisitions.

In my study, I decided to use the videos recorded by dual-axis confocal microscopes that contain lots of redundant data and can be considered a substitute for several independent image acquisitions. The separate frames extracted from the video sequence contain the overlapping regions of the same tissue – providing uncorrelated data from different spatial positions. The study was dedicated to explore the potential usage of this redundant data for speckle mitigation purposes. On the one hand, optimal ways for detection and registration of the overlapping regions among several frames were studied, whilst on the other hand, different ways of utilization of the redundant data was explored and I tried to go beyond simple averaging and apply more sophisticated approaches. Finally, our approach was compared against conventional methodologies under different circumstances.

The Molecular Biophotonics Laboratory at the University of Washington, led by Prof. Jonathan T.C. Liu is focused on designing dual-axis confocal microscopes. As their microscopes use reflectance based technology (in contrast to fluorescence-based microscopes that are free from speckle noise), mitigation of the speckle noise is named among the primary concerns of the team at the University of Washington.

The project carried out by Davit Gigilashvili was a collaboration between Norwegian University of Science and Technology and the University of Washington – supervised by Prof. Jon Yngve Hardeberg and Prof. Marius Pedersen on Norwegian side and by Prof. Jonathan T.C. Liu on American side.

Acknowledgements

The author acknowledges the help of the Molecular Biophotonics Laboratory of the Faculty of Mechanical Engineering at the University of Washington, Seattle, WA, USA. I would like to thank the head of the laboratory, the host supervisor Prof. Jonathan T.C. Liu and the researchers of the laboratory Chengbo Yin and Linpeng Peter Wei for hosting me in their laboratory, and providing access to their datasets and their medical equipment.

Also, special thanks to Dr. Kivanc Kose of Memorial Sloan Kettering Cancer Center and Dr. Zhen Qiu of Stanford University for collaboration and providing access to their proprietary automatic image mosaicing software and its source code.

Table of Contents

Chapter 1: Introduction.....	1
1.1. Problem and Motivation.....	1
1.2. Aim of the Work.....	4
1.3. Report Outline.....	5
1.4. Chapter 1 Bibliography.....	6
Chapter 2: Background	7
2.1. Dual-axis Confocal Microscopy – Architecture Overview.....	7
2.2. Speckle Quantification and Quality Evaluation Metrics	9
2.2.1. Speckle Index	10
2.2.2. Structural Similarity	11
2.3. Conventional “Single-channel” approaches.....	13
2.3.1. Median Filtering	13
2.3.2. Gaussian Filtering	14
2.3.3. Bilateral Filtering	15
2.4. Other Approaches.....	16
2.5. Psychometric Scaling Experiments.....	17
2.6. Chapter 2 Bibliography	25
Chapter 3: Approach used within the project.....	29
3.1. Overview of the approach	29
3.2. Registration Problem	30
3.3. Approach Based on the Smaller Patches	32
3.4. Full Image Processing (automatic registration) Approach.....	39
3.5. Chapter 3 Bibliography	41

Chapter 4: Results..... 42

 4.1. Smaller Patches 42

 4.2. Full Image Approach..... 53

 4.3. Psychometric Scaling Experiments..... 55

Chapter 5: Summary 64

Chapter 6: Further Work..... 67

 6.1. Chapter Bibliography..... 68

Chapter 1: Introduction

1.1. Problem and Motivation

Speckle noise is a multiplicative noise of granular shape often present in medical images. The project this report is dedicated to is based on the images captured with dual-axis confocal-microscope (details discussed in subsequent chapters). Even though the vast majority of the literature regarding speckle noise is related to either ultrasound medical imaging, or sometimes SAR – Synthetic Aperture Radar, the similar nature of images (grayscale images, suffering from speckle noise – in case of medical imaging, the structures of the tissue (e.g. cellular structures) are essential to diagnosing) makes it applicable to microscopy images as well. It is worth explicitly mentioning that all the images discussed later in this project are grayscale images and any type of processing of the color images was beyond the scope of this project.

Even though some sources do not consider speckle as true noise, justifying this with the fact that it might be carrying useful information about the image and is not truly random in its nature [1], from image processing point of view, it is still considered a noise that deteriorates the quality of the images, obscures key structures and makes diagnosing difficult for the physicians [1][2].

As mentioned above, speckle is a high frequency granular multiplicative noise and its causes should be searched in the interference phenomenon. Speckle was described and characterized by various researchers even back in 1970s, Goodman [2], Wagner [3] and Burckhardt [4] being among the worth mentioning examples.

The cause of granularity was studied as early as in 60s of the past century, by Rigden and Gordon [5], and Oliver [6]. The causes and formation of speckle noise is well summarized by Goodman [2]:

“The vast majority of surfaces, synthetic or natural, are extremely rough on the scale of an optical wavelength. Under illumination by coherent light, the wave reflected from such a surface consists of contributions from many independent scattering areas. Propagation of this reflected light to a distant observation point results in the addition of these various scattered components with relative delays which may vary from several to many wavelengths, depending on the microscopic surface and the geometry, Interference of these dephased but coherent wavelets results in the granular pattern we know as speckle, Note that if the observation point is moved, the path lengths traveled by the scattered components change, and a new and independent value of intensity may result from the interference

process. Thus the speckle pattern consists of a multitude of bright spots where the interference has been highly constructive, dark spots where the interference has been highly destructive, and irradiance levels in between these extremes. “ [2]

The typical speckle pattern can be seen on *Figure 1.1.1.* below.

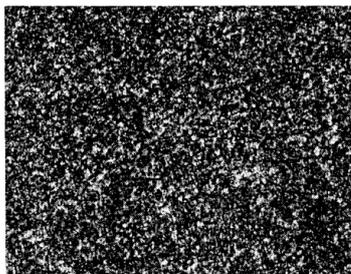


Figure 1.1.1.¹ – A typical speckle pattern

As the thorough review of the nature of the speckle noise is beyond the scope of this project, further details can be found in the references [7][8][9][10].

Considering again above cited paragraph from Goodman’s work [2], the speckle is not truly and completely random in its nature, but it’s dependent on various factors, like microscopic structure of the object viewed and the point of observation. The details of the system will be provided later the following chapters, but to explain it shortly:

The device in question is a miniature, handheld dual-axis confocal microscope used for in vivo examination of the tissues recording the video sequences as the operator manually moves the microscope across the tissue [11]. As the living tissue is constantly deforming and as the microscope is moving (even if it is held static, some tilting is caused by natural movement of a human hand holding the device), we can conclude that different frames of the video sequence that are overlapping at some extent depicting the same parts of the tissue (as the microscope is moving smoothly) are uncorrelated images with independent from each other speckle pattern. Theoretically, the speckle pattern can be reproducible, but considering the importance and impact of the microscopic units, in practice, it can be considered impossible and the speckle pattern can be considered random for a given frame.

If we get back to the study of Goodman [2],

“How is it possible to reduce the fluctuations present in a detected speckle pattern? The answer to this

¹ Credit for the figure: Goodman [2]

question follows from the fundamental result of probability theory that the sum of M identically distributed, real-valued, uncorrelated random variables has a mean value which is M times the mean of any one component, and a standard deviation which is \sqrt{M} times the standard deviation of one component. Thus, if we add M uncorrelated speckle patterns on an irradiance basis, the contrast of the resultant speckle pattern is reduced...
Uncorrelated speckle patterns can be obtained from a given object by means of time, space, frequency, or polarization diversity.”[2]

This opinion is explicitly substantiated by Suri et al. [1], as they name “averaging of uncorrelated images of the same tissue recorded under different spatial positions” among speckle mitigation techniques. Although they mention the approach as effective but too complicated, we are going to challenge this opinion within this project.

Thus, to summarize the idea, thesis or a key statement of the project - averaging uncorrelated images of the same tissue can lead to speckle noise mitigation and thus, improved quality of the image in terms of diagnosing difficulty.

But there are other challenges related to this issue. We should understand that while speckle noise is too apparent in some cases (e.g. *Figure 1.1.2.*), sometimes it is too challenging to identify what is speckle and what is inherent part of the tissue (e.g. *Figure 1.1.3.*)– thus, what has to be removed and what should be kept? This is one of the reasons objective metrics for quality and speckle quantification are needed and will be addressed later.



Figure 1.1.2.² – the noise is too apparent here

² Credit for the figure:
https://www.researchgate.net/publication/274202418_A_Comparative_Study_on_Approaches_to_Speckle_Noise_Reduction_in_Images [20 August 2017]

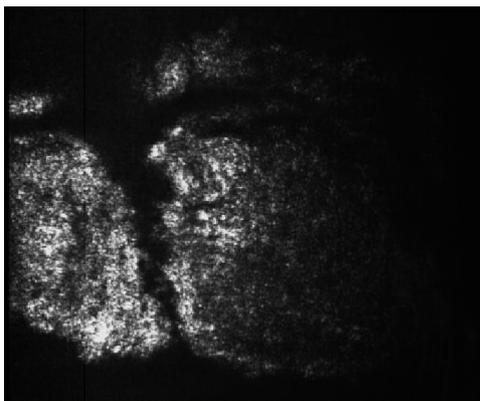


Figure 1.1.3. – it is too challenging to decide what is noise and what is not.

1.2. Aim of the Work

The aim of the project is to study speckle mitigation opportunities using uncorrelated images of the single tissue. The exact details of the approach will be provided in *Chapter 3*, but to summarize it shortly, the aim of the project is to utilize the redundant data of the overlapping regions present in the consecutive frames of the video sequence recorded by the handheld miniature dual-axis confocal microscope, as this particular application is less studied in terms of speckle removal techniques, in contrast to extensively studied ultrasound medical imaging applications. Using video footage solves the complication problems related to several image acquisition procedures stated as one of the advantages of the approach in the survey of Suri et al. [1]

Secondary aim of the project is to go beyond the simple averaging of the uncorrelated images and to try more sophisticated ways of processing of this redundant data to either improve the speckle mitigation quality, or to overcome the artifacts present due to improper registration of the consecutive frames (this point will be addressed thoroughly in subchapter 3.2.

Additional point of interest is to examine ways of image registration and detection of overlapping regions that as we will see later in the report (subchapter 3.2.) is a pivotal part of this approach.

I realize that this problem is not possible to be fit within a single master's thesis research project, but considering the follow-up opportunities, the global goal of the work can be defined as creation of a robust fully automatic framework that will take a video footage recorded with a dual-axis confocal microscope designed by the Molecular Biophotonics Laboratory team, will utilize the redundant data present in the consecutive frames (the same region is present in several frames of the video) to reduce the speckle noise in the video, or at least in individual frames to visualize the key structures (e.g. cellular borders and shapes, blobs) and fine details better and to make diagnosing easier for the physician.

As the project is a product of a collaboration between ColourLab at the Norwegian University of Science and Technology in Gjøvik, Norway and the Molecular Biophotonics Laboratory at the University of Washington in Seattle, Washington, United States – an additional aim can be defined as building further, long-term collaborations and follow-up research project opportunities between the two institutions.

1.3. Report Outline

The report is composed of 6 chapters. Initially, the first chapter is dedicated to the introduction into the problem and the aim of the work. The bibliography of the respective chapter is provided as the last sub-chapter at the end of each chapter.

The 2nd chapter provides detailed overview of the background knowledge used as a basis for this project: the principles of dual-axis confocal microscopy – the architecture of the devices in question are overviewed, followed by the description of the objective metrics used for evaluation purposes, then description of speckle mitigation metrics examined within the project follow and finally, the chapter is concluded with subjective evaluation methodologies and bibliography, as already mentioned.

The 3rd chapter provides all the details about the approaches and techniques implemented within the framework of this project, as well as the difficulties faced during this implementation and their solutions and workarounds found.

The next, 4th chapter, puts all the results together after evaluation of the approaches discussed in the previous chapter. The results of objective evaluation methodologies are presented separately in two sub-chapters for the two different approaches, while the chapter ends with the third sub-chapter discussing the results of psychometric scaling experiments and the bibliography is given in the fourth one.

The 5th chapter summarizes the whole project, implementations and approaches completed within the project, their results, and finally, findings and conclusions drawn from them.

The last, 6th chapter puts together all the open questions left for the future work – defining the possible directions for the follow-up projects.

Finally, one of the algorithms used within the project can be found in the appendix.

1.4. Chapter 1 Bibliography

- [1] Suri, JS, Chang, R, & Kathuria, C (eds) 2008, Advances in Diagnostic and Therapeutic Ultrasound Imaging, Artech House, Norwood. pp.37-63 Available from: ProQuest Ebook Central. [8 August 2017].
- [2] Goodman, J.W., 1976. Some fundamental properties of speckle. *JOSA*, 66(11), pp.1145-1150.
- [3] Wagner, R.F., Smith, S.W., Sandrik, J.M. and Lopez, H., 1983. Statistics of speckle in ultrasound B-scans. *IEEE Transactions on sonics and ultrasonics*, 30(3), pp.156-163.
- [4] Burckhardt, C.B., 1978. Speckle in ultrasound B-mode scans. *IEEE Transactions on sonics and ultrasonics*, 25(1), pp.1-6.
- [5] Rigden, J.D. and Gordon, E.I., 1997. The granularity of scattered optical maser light. *SPIE MILESTONE SERIES MS*, 133, pp.213-213.
- [6] Oliver, B.M., 1963. Sparkling spots and random diffraction. *Proceedings of the IEEE*, 51(1), pp.220-221.
- [7] Goodman, J.W., 1975. Statistical properties of laser speckle patterns. *Laser speckle and related phenomena*, 9, pp.9-75.
- [8] Beckmann, P. and Spizzichino, A., 1987. The scattering of electromagnetic waves from rough surfaces. Norwood, MA, Artech House, Inc., 1987, 511 p.
- [9] Dainty, J.C., 1970. Some statistical properties of random speckle patterns in coherent and partially coherent illumination. *Journal of Modern Optics*, 17(10), pp.761-772.
- [10] Duncan, D.D., Kirkpatrick, S.J. and Wang, R.K., 2008. Statistics of local speckle contrast. *JOSA A*, 25(1), pp.9-15.
- [11] Yin, C., Glaser, A.K., Leigh, S.Y., Chen, Y., Wei, L., Pillai, P.C.S., Rosenberg, M.C., Abeytunge, S., Peterson, G., Glazowski, C. and Sanai, N., 2016. Miniature in vivo MEMS-based line-scanned dual-axis confocal microscope for point-of-care pathology. *Biomedical optics express*, 7(2), pp.251-263

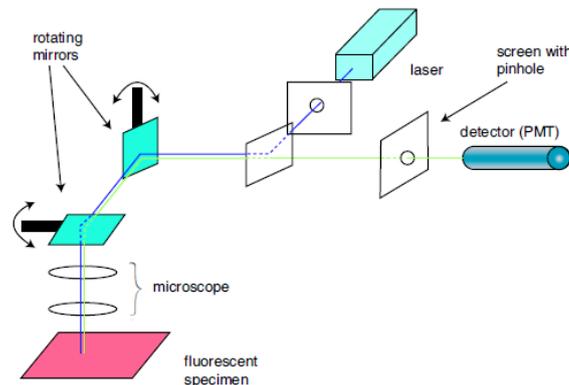
Chapter 2: Background

2.1. Dual-axis Confocal Microscopy – Architecture Overview

Dual-axis and confocal microscopy concepts are developed separately and neither are new. Confocal design of the microscopes was studied by Shaw [1] back in 1995 and was well summarized in the works of Pawley as well [2].

The key idea of confocality of the microscope is the following: a pinhole or a slit is installed in front of the detector in order to detect light only from the focal plane of the system and eliminate out-of-focus light. Detection of out of focus light creates haze and blurs the image of the specimen. Elimination of out-of-focus light and isolated collection of in-focus light leads to haze-free, sharper images – thin optical sections of it, to be more precise [1][3][4][5], - and can achieve sub-cellular resolution and clear images of optically “thick” tissues according to Piyawattanametha and Wang [6]. On the other hand, diffraction and high signal-to-noise-ratio limit the size of the pinhole [3].

The basic setup of the confocal microscope is illustrated on *Figure 2.1.1.* [3]



*Figure 2.1.1.*³ - Basic setup of a confocal microscope. Light from the laser is scanned across the specimen by the scanning mirrors. Optical sectioning occurs as the light passes through a pinhole on its way to the detector. [3]

The idea of dual-axis architecture is the separation of illumination and collection axes that ensures that no or negligible amount of backscattered light from the tissue enters and biases the collection beam. This enables achieving longer working distance, higher dynamic range of acquisition and scalability of the device without sacrificing the resolution [6][7][8] – and hence, it is especially helpful while miniaturizing the device [7].

³ Credit for the figure: Semwogerere and Weeks [3]

The illustration of basic dual-axis architecture can be seen on *Figure 2.1.2*.

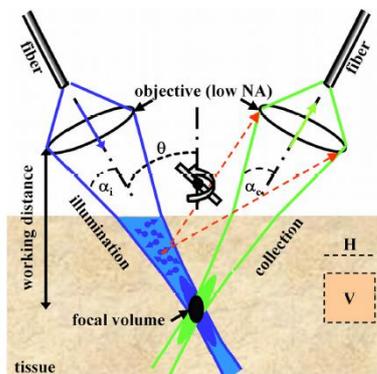


Figure 2.1.2.4 – illustration separate illumination and collection axes of the microscope

The actual microscope used within the project was designed by the team of Molecular Biophotonics Laboratory of the University of Washington and combines the advantages of confocal and dual-axis microscopy and relies on reflectance-based technology, as it is intended for in vivo usage [5].

Reflectance-based confocal microscopes have been used earlier for in vivo diagnosing of malignant melanoma [9], as well as directly within intra-operative process [10]. The microscope is intended for in vivo noninvasive point-of-care pathology, free from chemical preprocessing of ex vivo histopathology.

The architecture of the microscope in question is illustrated below on *Figure 2.1.3*. Two tilting MEMS (microelectromechanical systems – electro mechanical technology of microscopic devices, mostly with moving or tilting parts [11]) mirrors (M1 and M2) are used for scanning and two separate axis – blue from 660 laser source for illumination and green for collection – are used, which intersect at the back focal plane of the objective lens. The miniature size of the device is illustrated on *Figure 2.1.4*. The further details of the device are beyond the scope of the report and can be found the cited paper [5].

All further work described in the thesis is dedicated to speckle mitigation in the images captured with the device illustrated below. It is worth mentioning that the primary task to be fulfilled with this device is in vivo detection of oral cancer, although the device is extensively used for skin examination as well [5].

⁴ Credit for the figure: Piyawattanametha and Wang [6]

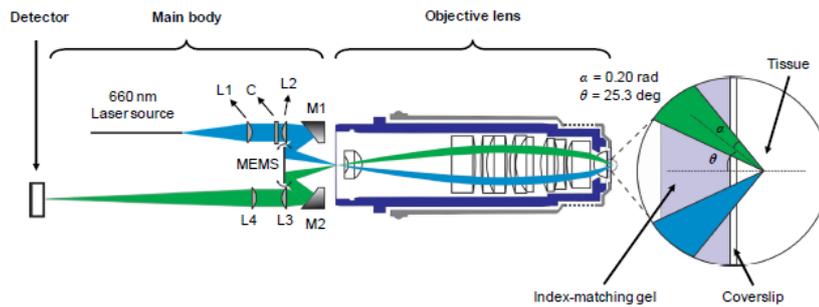


Figure 2.1.3⁵ – the dual-axis architecture of the device in question

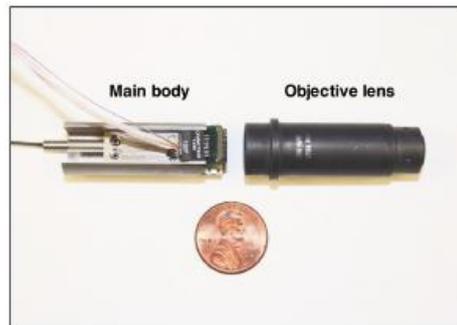


Figure 2.1.4⁶ – the size of the device in relation to US 1-cent coin

2.2. Speckle Quantification and Quality Evaluation Metrics

The summaries of despeckling techniques [12][13] offer various ways of speckle and despeckling quality quantification, from the approaches as simple as mean-square-error and peak-signal-to-noise-ratio, to sophisticated image quality metrics and classification tasks.

The example of the latter was initially found very interesting. Christodoulou et al. [14] suggest running a classification task on a set of symptomatic and asymptomatic images using kNN classifier (k-nearest-neighbor – a statistical classifier using Euclidean distance for neighborhood determination. k was equal to 7 in the case discussed [12][14]) classifying images as symptomatic or asymptomatic before and after speckle mitigation processing. If the application of speckle mitigation technique increases the classification accuracy, the technique is considered acceptable. Unfortunately, this way of evaluation was impossible to implement due to lack of symptomatic images in the available datasets. But the approach was still worth mentioning, as it can be used in further, follow-up works.

⁵ Credit for the figure: Yin et al. [5]

⁶ Credit for the figure: Yin et al. [5]

Two simple quality evaluation metrics – MSE (mean square error) and PSNR (peak signal to noise ratio) were initially applied but were found inappropriate due to their inconsistency and inability to reflect structural and spatial changes.

Finally, two metrics were used for evaluating the success of particular speckle mitigation algorithms: widely cited and used Speckle Index and conventional image quality metric SSIM – structural similarity. Both of them are shortly summarized below.

2.2.1. Speckle Index

Speckle index was suggested by Crimmins [15] back in 1986. Crimmins suggested that finding a ratio of the deviation of the intensities and the mean intensity values could be a reasonable metric to quantify the speckle noise, due to its multiplicative nature. The ratios are initially found locally, within 3-by-3 windows. The deviation is defined, as the difference between maximal and minimal values (*Equation 2.2.1.1*). The mean is defined as the mean value within the 3-by-3 window (*Equation 2.2.1.2*). Finally, one speckle index value is defined as an average of all local (all possible 3-by-3 windows) values (*Equation 2.2.1.3*, *Figure 2.2.1.1*).

For $1 < m \leq M$ and $1 < n \leq N$,

$$\sigma(m,n) = \max f(m+a, n+b) - \min f(m+a, n+b) , \text{ where } -1 \leq a, b \leq 1$$

Equation 2.2.1.1.

$$\mu(m,n) = (1/9) \sum_{a,b=-1}^1 f(m+a,n+b)$$

Equation 2.2.1.2.

$$\text{speckle index} = (1/MN) \sum_{m=1}^M \sum_{n=1}^N \sigma(m,n) / \mu(m,n)$$

Equation 2.2.1.3.

where,

f is a function representing the image,

M and N are the dimensions of the image.

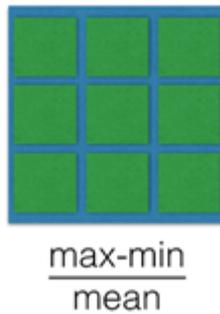


Figure 2.2.1.1. – the local ratios are found within 3-by-3 window, dividing the deviation by the mean

Initially, I thought that the deviation on the edges could bias the result. But empirical and practical examination showed that it had insignificant influence on the final value of the metric.

Finally, it is important to mention that the value of the speckle index varies between 0 and 1 and the lower value means lower level of speckle noise.

2.2.2. Structural Similarity

Decreasing the speckle, thus decreasing the ratio between deviation and mean, logically leads to blurrier images. As keeping the cellular structures sharp enough is important for proper diagnosing, and as we are working with the grayscale images, SSIM – Structural Similarity [16] was selected as another metric for quality evaluation, particularly, for evaluating how well the key structures were preserved after speckle mitigation.

It is intuitive that decreasing speckle and blurring the image blurs and weakens the key structures as well. Therefore, simultaneous usage of the Speckle Index and Structural Similarity was defined as a way to find the golden mean and optimal extent of speckle mitigation.

The work by Wang et al. [16], describes the approach into details. While MSE, PSNR and even Speckle Index measure only the absolute difference, SSIM takes features of human visual system into consideration – namely, luminance masking and contrast masking phenomena – and measures the similarity between the structural elements separately from luminance and contrast within the image (refer to *Figure 2.2.2.1.* below).

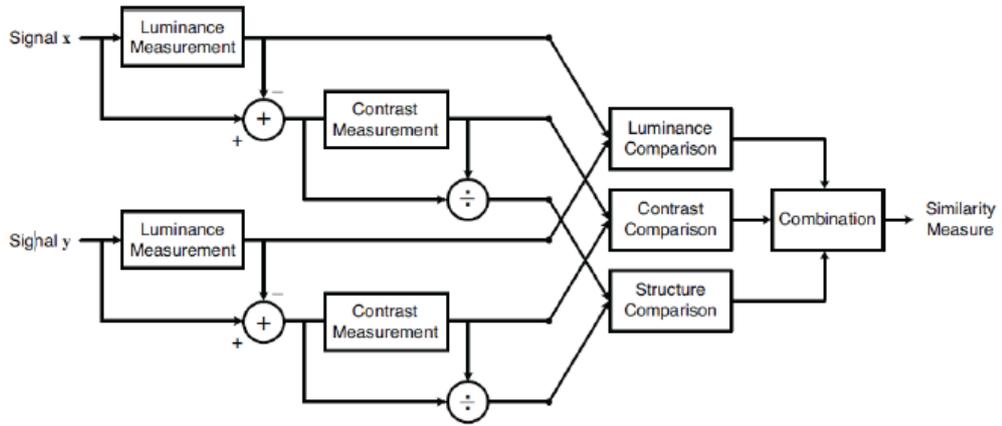


Figure 2.2.2.1.7 – the workflow of SSIM algorithm [16]

For more detailed understanding of the approach, refer to the Equation 2.2.2.1. below:

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

Equation 2.2.2.1[16][17]

where,

μ_x is the average of x,

μ_y is the average of y,

σ_x^2 is the variance of x,

σ_y^2 is the variance of y,

σ_{xy} is the covariance of x and y,

$C_1 = (k_1L)^2$, $C_2 = (k_2L)^2$ are two variables to stabilize the division with weak denominator

L is the dynamic range of pixel-values (typically $2^{\text{\#bits per pixel}} - 1$),

$k_1=0.01$ and $k_2=0.03$ by default. [16][17]

Finally, it is important to mention that the value of the structural similarity also varies between 0 and 1 as that of speckle index, but in contrast to speckle index, lower value means lower structural similarity and is considered less acceptable.

⁷ Credit for the image: Wang et al. [16]

2.3. Conventional “Single-channel” approaches

As mentioned earlier, the vast majority of speckle mitigation techniques use just a single image for despeckling this very image, disregarding other frames depicting the same part of the tissue.

One commonly used technique for speckle noise removal is usage of wavelet domains [18][19][20]. The idea that the speckle noise is high-frequency noise in its nature, inspired difference research works that transform the images into wavelet domains and try to remove speckle through thresholding or discarding HH, HL and LH high frequency domains. But the biggest disadvantage of this technique is the fact that sometimes key structures and useful information is also completely discarded from the image.

Initially, I tried to explore this approach, but it dramatically biased the final mosaic after automatic registration of several frames (will be addressed later) apparently through diminishing the key structures (*Figure 2.3.1.*). As the conventional methodologies were not the point of primary focus for this research work and the time and resources were limited, I left this question open for further exploration.

Finally, I came up with several simple conventional speckle removal approaches cited throughout the literature [11][19] to compare my approach against.

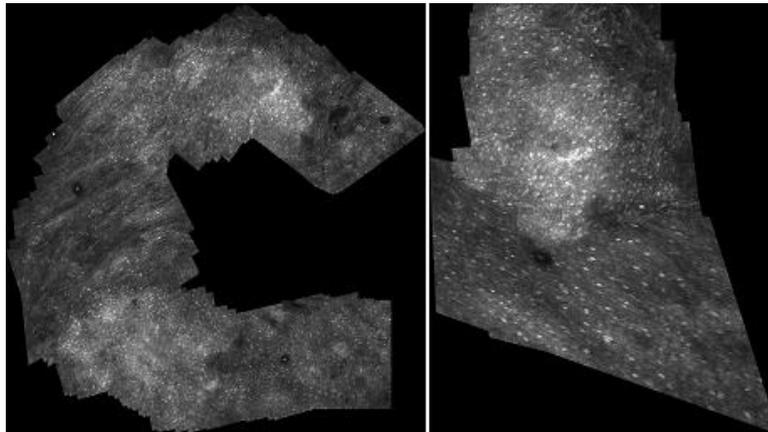


Figure 2.3.1. – left: ground truth – video sequence frames registered without any pre-processing (3829x3961px original resolution) right: video sequence frames registered after wavelet-domain pre-processing (1663x2304px resolution)

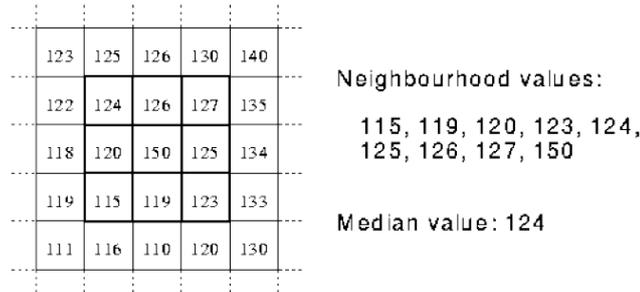
2.3.1. Median Filtering

Median filtering is one of the simplest and oldest techniques used for image blurring. The main idea of the approach is that the intensity value of a pixel is replaced by the median value of its neighbors. The neighborhood is defined as m-by-m window with the given pixel in the center. m can be any odd number, with 3-by-3 smallest possible

window size [21][22]. The blurriness increases as the window size increases. Several options were examined within the project and it will be addressed later.

The biggest advantage of median filtering is its simplicity, low computational cost and the fact that biased values (e.g. saturated pixel) cannot affect the median value.

For a better illustration, how median values are found, refer to the *Figure 2.3.1.1.* below :



Figure⁸ 2.3.1.1.- the illustration how median values are found

2.3.2. Gaussian Filtering

Gaussian filter is rotationally symmetric lowpass filter that can have not only square, but rectangular kernel window as well. Instead of taking the median value among the neighbors, Gaussian filter puts weight on the value of each neighboring pixel depending on the distance from the central pixel and manually defined standard deviation. Higher the standard deviation is, higher the weights of the neighboring pixels are. And on the contrary, lower the standard deviation - lower is the FWHF of the Gaussian function and the weight is more “concentrated” at the center. For illustration, refer to *Figure 2.3.2.1.*

6.9625e-08	2.8089e-05	2.0755e-04	2.8089e-05	6.9625e-08
2.8089e-05	0.0113	0.0837	0.0113	2.8089e-05
2.0755e-04	0.0837	0.6187	0.0837	2.0755e-04
2.8089e-05	0.0113	0.0837	0.0113	2.8089e-05
6.9625e-08	2.8089e-05	2.0755e-04	2.8089e-05	6.9625e-08

Figure 2.3.2.1. (a) – 5-by-5 lowpass filter with default standard deviation (0.5)

⁸ Credit for the figure: <https://jiteshgupta1192.files.wordpress.com/2013/11/image-denoising.pdf> [20 August 2017] [23]

0.0030	0.0133	0.0219	0.0133	0.0030
0.0133	0.0596	0.0983	0.0596	0.0133
0.0219	0.0983	0.1621	0.0983	0.0219
0.0133	0.0596	0.0983	0.0596	0.0133
0.0030	0.0133	0.0219	0.0133	0.0030

Figure 2.3.2.1. (b) – 5-by-5 lowpass filter with standard deviation equal to 1.0

0	0	1.2647e-14	0	0
0	1.1238e-07	3.3501e-04	1.1238e-07	0
1.2647e-14	3.3501e-04	0.9987	3.3501e-04	1.2647e-14
0	1.1238e-07	3.3501e-04	1.1238e-07	0
0	0	1.2647e-14	0	0

Figure 2.3.2.1. (c) – 5-by-5 lowpass filter with standard deviation equal to 0.25

2.3.3. Bilateral Filtering

Previously described methodologies are “edge-blind”, smoothing any area, regardless it is a homogenous region or an edge. As mentioned earlier, it is crucial to keep structures as good as possible. One of the solutions for this problem is bilateral filtering – blurring the image, while keeping the edges. Similarly to Gaussian filtering, it is based on the filters composed of the weights. Bilateral filter is a weighted-average filter, where value of each pixel is calculated as the average of its neighboring pixels, weighted by the Gaussian of not only spatial distance, but intensity distance as well. Considering intensity distance within the weighting process, enables the approach to preserve edges and blur different areas of the image at different extent depending on the content and presence of the edges [24][25]. The method can successfully used for denoising tasks as demonstrated in the literature [26].

For a better illustration, refer to *Figure 2.3.3.1*.

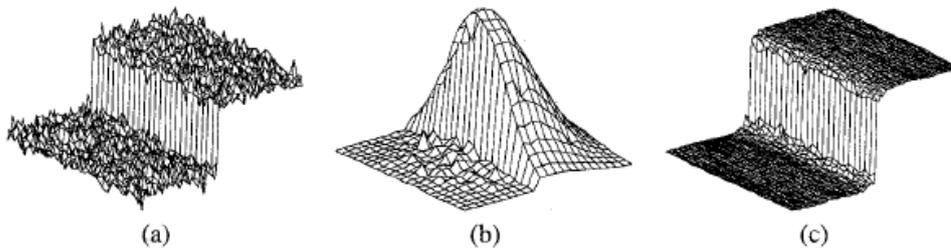


Figure 2.3.3.1.⁹ – (a) A 100-gray-level step perturbed by Gaussian noise with std= 10 gray levels. (b) Combined similarity weights for a 23 x 23 neighborhood centered two pixels to the right of

⁹ Credit for the figure: Tomasi and Manduchi [24]

the step in (a). The range component effectively suppresses the pixels on the dark side. (c) The step in (a) after bilateral filtering with $\text{std}_{\text{intensity}} = 50$ gray levels and $\text{std}_{\text{geometric distance}} = 5$ pixels. [24]

2.4. Other Approaches

Further sophistication of edge-preserving filtering and the state-of-the-art for speckle mitigation is guided image filtering [25][27]. As it is intuitive from its name, guided image filtering uses an image as a guide for filtering - this can be the same image that has to be filtered, thus filtering guided by itself, or this can be filtering guided by another image.

While applying this technique, we do have predefined guidance image I , predefined input image p and output image q . The idea is based on the assumption that q is a linear transform of I in a window w_k centered on a pixel k . In order to determine the linear coefficients, constraints are taken from input image p , as output is modeled as input image after subtraction of unwanted noise components.

$$q = p - n,$$

Equation 2.4.1.

where,

q is output image,

p is input image,

n is unwanted noise.

The solution is found through minimization of a cost function. The algorithm of the guided filter can be found in appendix.

The difference between bilateral filtering and spatial filtering is well illustrated on the *Figure 2.4.1.* below. The guide I is the identical image in case of bilateral filter, while guide I for guided image filtering can be an image. While bilateral filter takes spatial kernel and range kernel from itself and simply applies those ones combined - the bilateral kernel, guided image filtering takes two guide and filtering input images, finding the optimal output q through cost minimization function. As the output is a linear transform of the guide I , the empirical study has demonstrated an interesting feature that in case of dissimilarity between the structures of I and p , the structures from I are prioritized in the output in the majority of the cases.

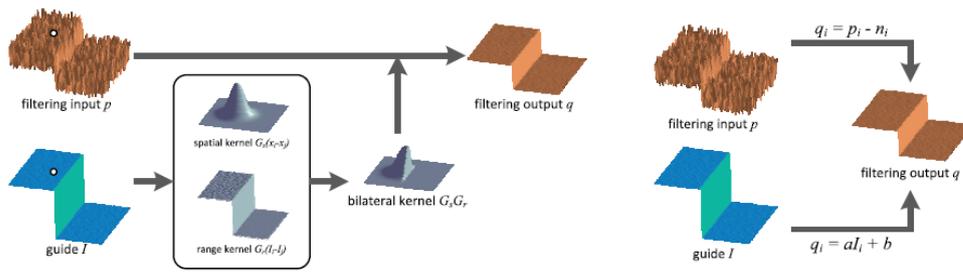


Figure 2.4.1.¹⁰ – (left) bilateral filtering; (right) guided image filtering

Finally, we come up to a regularization parameter ϵ . If the variance within w_k is above ϵ , the window is considered high variance region and is not smoothed a lot, while if the variance is below ϵ , the window is considered a “flat patch” and higher degree of smoothing is applied. Therefore, there are two parameters that we can control manually : the windows size and regularization parameter ϵ . Tuning those parameters, we can control the “agressiveness” of the technique and find the optimal setup for a given dataset. This point is explored further in the next chapters.

2.5. Psychometric Scaling Experiments

Even though the objective metrics provide very interesting information for understanding the speckle noise amount and general image quality, as already mentioned above, the final purpose of the speckle mitigation techniques is still to make diagnosing easier and more reliable for the doctors. Therefore, ideally, the most reliable and effective way of evaluation is evaluating the resulting images by the experts, i.e. the physicians, who should actually use the images for diagnosing purposes.

Unfortunately, there were no opportunities to collaborate directly with the clinicians and evaluation by the experts was left open for the future development of the research. Approval by the experts is one of the key factors impacting the implementability of the developed technique.

Regardless of the lack of expert observers, the task in question: “decreasing the speckle noise level, while keeping the key structures” was still intelligible for the observers without medical background. Hence, I considered organizing a psychometric scaling experiment and analyzing its results could still provide very interesting insight into the performance comparison of several different techniques.

Psychometric scaling methods are used for image evaluation purposes, for instance, to determine compression quality and its limit, color tolerances, watermarking quality

¹⁰ Credit for the figure: He et al. [25]

and many more. Three most common and conventional ways for conduction psychometric scaling experiments are: [28]

- a) Category Judgment
- b) Paired Comparison [28][29]
- c) Rank Order

In category judgment [28], several categories (usually, 5 or 7, from worst to best) are defined and the observer should assign each stimulus to a particular category. It is usually faster, as lower number of judgments/comparisons is needed and gives valuable information about the distance between the stimuli (i.e. the images, in our case). On the other hand, it can become too complicated for the observer [28] and can suffer from extreme aversion bias [29], when observer demonstrates the tendency of avoiding extreme categories (too bad or excellent) and tries to select the intermediate values [30].

The example of category judgment is visualized on the *Figure 2.5.1.* below, where the patches (30,40,50,60,70), are assigned categories 1-to-7 based from their difference extent from the reference patch.

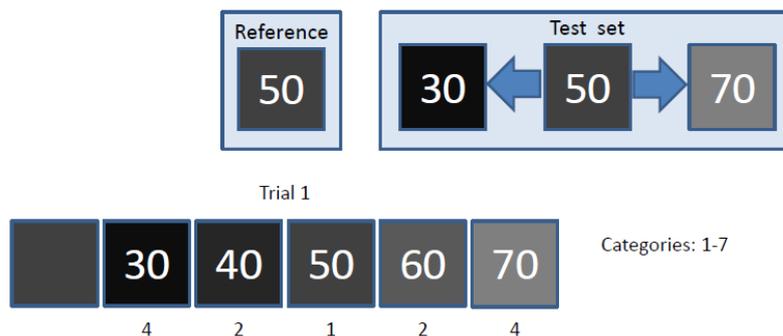


Figure 2.5.1. ¹¹- the patches (30,40,50,60,70), are assigned categories 1-to-7 based from their difference extent from reference patch.

In case of paired comparison [28][29], the observer is shown the pairs of the stimuli and he/she should judge which one is preferable according to a predefined criterion. The original may or may not be present (e.g. questions like: “which of the two is more similar to the original?”, or “which of the two has a better quality?”). In contrast to category judgment, it does not record distance information, and number of comparisons can be high, if the image set is large, but it is still very popular approach

¹¹ Credit for the figure: “IMT 4172 Color Image Quality and Processing in an Imaging Workflow” course materials. Autumn 2016. Credit belongs to: Assoc. Prof. Marius Pedersen, The Norwegian Colour and Visual Computing Laboratory, Norwegian University of Science and Technology.

due to its simplicity, as it is very easy to understand and judge, and little knowledge is required from the observer [28][29]. It is the most popular method to evaluate gamut mapping [31].

The example of paired comparison is visualized on the *Figure 2.5.2.* below, where the two patches are judged according to their similarity to the reference patch (50).

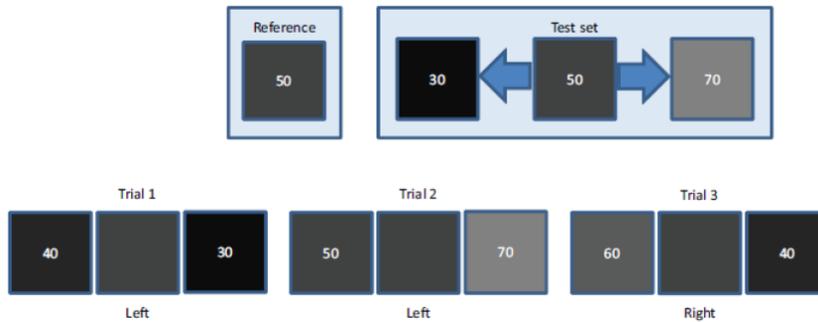


Figure 2.5.2.¹² - the two patches are judged according to their similarity to the reference patch (50).

The rank order [28] is similar to paired comparison, but in contrast to it, several images are judged simultaneously and ranked from “best” to “worst” according to a given criterion. Even though it can be very fast (small number of comparisons), it can become too complicated and confusing for the observer, when the number of the stimuli/images is high.

The example of rank order is visualized on the *Figure 2.5.3.* below, where five patches (30,40,50,60,70) are ranked according to their similarity to the reference patch (50) from closed to most different one.

¹² Credit for the figure: “IMT 4172 Color Image Quality and Processing in an Imaging Workflow” course materials. Autumn 2016. Credit belongs to: Assoc. Prof. Marius Pedersen, The Norwegian Colour and Visual Computing Laboratory, Norwegian University of Science and Technology.

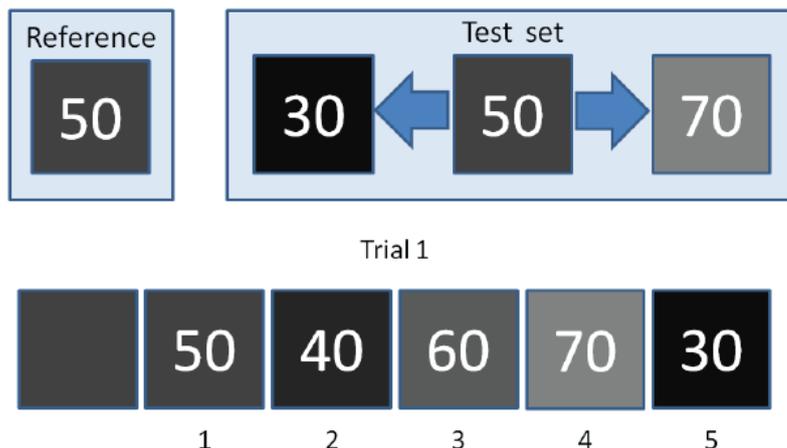


Figure 2.5.3.¹³ - the five patches (30,40,50,60,70) are ranked according to their similarity to the reference patch (50) from closed to most different one.

Considering the importance of the simplicity of the task, Paired Comparison approach was selected for psychometric scaling experiments. The primary reason, why the simplicity was important was the lack of expertise and little knowledge about the stimuli among the observers. On the other hand, small image sets allowed shorter and relatively faster experiments.

It is important for successful psychometric scaling experiment that best practices were taken into consideration, therefore, the work by Engeldrum [32], was used as a key reference while determining the practical issues for the experiment.

The reference [32] states that the number of observers is primarily dependent on availability of the observers and increasing the number of the latter, increases the precision. The recommended number of observers is from ten to thirty. In our case, I selected the mean of the two – twenty observers. The group of the observers was diverse in terms of age, gender and nationality. Observers of both genders participated with the age ranging from 12 to 61, with 25 years being a median age and 26.75 the mean age of them. The observers were representing seven different nationalities. The national, age or gender groups were not studied separately within this project, but the information is still mentioned for possible future enquiries.

The expert and average observers are differentiated in the literature [32], considering their skills and knowledge related to the stimuli. As mentioned earlier, no observers with medical expertise were available, but they were still differentiated for balance's sake – 10 observers had no technical/IT/engineering/image processing background,

¹³ Credit for the figure: “IMT 4172 Color Image Quality and Processing in an Imaging Workflow” course materials. Autumn 2016. Credit belongs to: Assoc. Prof. Marius Pedersen, The Norwegian Colour and Visual Computing Laboratory, Norwegian University of Science and Technology.

while the remaining 10 had it. To mention again, for more credibility, the participation of the observers with medical background were left open for future studies.

The reference [32] suggests that the number of stimuli (images in our case) should be dependent on the method of scaling experiment used. For paired comparison, the number of comparisons is defined as:

$$(n*(n-1)*m)/2$$

Equation 2.5.1.

where,

n is the number of reference images (different frames or pairs of the frames, in our case)

m is the number of reproductions (frames processed with different speckle mitigation algorithms, in our case).

In order to avoid exhaustingly long experiments, five 64-by-64 patches and 5 large scale images were selected, therefore, 10 “reference images” in total. It is worth mentioning that in this case, the samples were not compared against the original and no original was shown explicitly. But the original was secretly put among the “reproductions”, in order to study, whether any speckle mitigation algorithm is preferable or needed at all. More than ten speckle mitigation approaches were addressed in this project, but considering them all could make the experiments extremely long and unrealizable. Thus, I selected three of them that demonstrated the most interesting and promising results (discussion in Chapter 4): bilateral filtering, the point-wise average and guided image filtering when the point-wise average of the two consecutive frames is filtered with the guidance of the first frame. Adding the unprocessed original among them, I ended up with 4 different “reproductions”. This means 60 comparisons.

But to avoid the possible bias - some observers opting mostly either for the left or the right image, the same pair of the stimuli were presented twice – in a flipped order. That increased the number of comparisons up-to 120 per experiment in total. The practice has shown that average time needed for completion of each experiment was 15 minutes. The observers could control the time through the explicit clock present on the screen and no complaints from the observers about the duration of the experiment were recorded. Besides, the pairs were shown in a random order to avoid any significant bias by previous pairs.

Another important factor considered was forced choice versus possible tie option – whether to force the observer to choose between the two, or not. As the majority of the pairs had subtle difference, I considered allowing tie too tempting and opted for the forced choice option.

In order to simplify the observer recruitment process (no enough observers were present in the lab), the experiment was held online with the help of QuickEval [33] platform developed at Norwegian University of Science and Technology with the very purpose of organizing psychometric scaling experiments by its researchers. All the experiments were held on the personal computers and no smartphones were allowed due to incompatibility with the platform. I realize that the difference among the displays of the observers, as well as uncontrolled ambient surround could lead to some bias, but as we were not dealing with colors (but with structures), nor having access to actual calibrated devices used in practice for diagnosing by the physicians and as small number of the “reproductions” in question and clear and obvious difference between them were assumed to have little impact on decision making, the experiment was considered still worth organizing. In order to claim more credibility, in the future study, the experiment should be repeated under controlled conditions, on calibrated displays actually used in the clinical field and most importantly, with the observers having expertise in medical field.

The background color was set to neutral gray, as mentioned in the literature [32]. The interface with the actual images used in the experiment can be seen on the *Figures 2.5.4* and *2.5.5* below:



Figure 2.5.4. – Interface of the platform used for the experiment: comparing 64-by-64 px patches

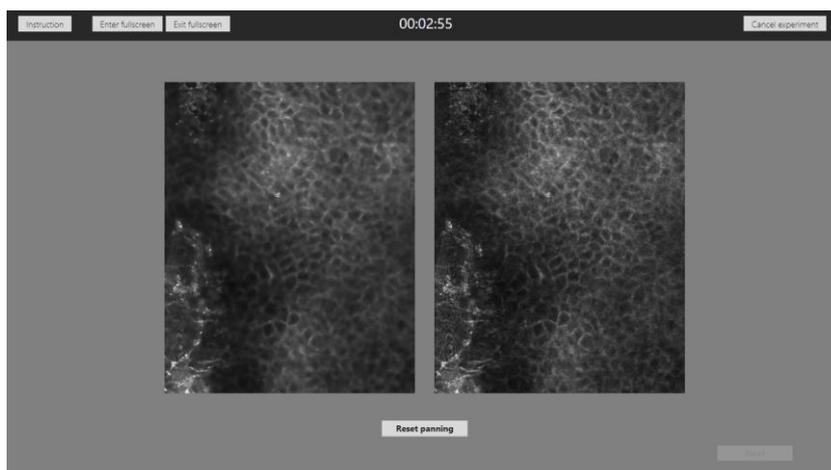


Figure 2.5.5. – Interface of the platform used for the experiment: comparing large scale images (512-by-512 px on this particular one)

Another crucially important aspect needs detailed review is the instruction given to the observers. Refer to the practices again summarized by Engeldrum [32]:

“Next to the sample image set, observer instructions are the most significant item that controls the context of the observers' judgments in a scaling study.” [32]

The instructions define how the observers behave and how do they define their preferences. As the observers did not have medical expertise, they were given explicit and detailed instruction to select the image that seems to them having lower level of noise and **at the same time(!)** better visualization of key structures (like edges, corners, cellular shapes). The idea was to consider both parameters and find the image balancing the two better. For the observers without technical background, *Figure 1.1.2.* was shown while explaining what does the speckle noise and noise generally mean. In addition to this, the platform displayed the instruction before starting the experiment, stating:

“Judge which image (either left or right) keeps the better balance between low level of noisiness on the one hand – and visibility of the structures and fine details, on the other hand.”

Finally, the data was analyzed using Thurstone’s Law of Comparative Judgment [34] – the data collected through pair comparisons was converted into interval scale data. The final results, z-scores, from this data represent the distance of a given image from the mean of the whole set examined. This is a kind of standardization of the data setting mean to zero and standard deviation to 1, while keeping skewness and kurtosis of the original data set.

The details of the approach can be found in the reference [34]. The key idea is to build the frequency matrix, how many times each image is preferred over another one. Then this frequency is converted to percentage matrix. E.g. if A is considered more preferable over B in 9 cases out of 10, it is considered more preferable in 90% of the cases. *Table 2.5.1.* and *2.5.2.* represent sample frequency and percentage matrices respectively. In the first table, we see that stimulus A is preferred over B 2 times, while B is preferred over A 18 times. A is preferred over C 6 times, while the latter is preferred over A 14 times by observers. B is preferred over C 4 times out of 20. Those numbers are converted into percentages in *Table 2.5.2.*

	A	B	C
A		2	6
B	18		4
C	14	16	

Table 2.5.1. – sample summed frequency matrix representing preferences among 3 samples: A,B and C

	A	B	C
A		0.1	0.3
B	0.9		0.2
C	0.7	0.8	

Table 2.5.2. – sample summed percentage matrix representing preferences among 3 samples: A,B and C

Afterwards, logistic Function Matrix is built with the equation proposed by Bartleson [35].

$$LFM = \ln ((f+x) / (N - f + c))$$

Equation 2.5.2.

where,

f is the value from the frequency matrix,

N is number of observations,

and c is arbitrary constant, usually equal to 0.5. as proposed by Bartleson [35].

Finally, LFM is transformed to z-scores by simple scaling coefficient that is found through linear regression considering the relationship between the standard normal cumulative distribution for percentage matrix and LFM.

Z-scores are represented usually by error bar plots, illustrating their 95% confidence interval, which defines the range that includes the unknown parameter with probability of 95% [36]. CI (confidence interval) is calculated with the equation below [36]:

$$CI = 1.96 * \sigma / \sqrt{N}$$

Equation 2.5.3.

where,

σ is standard deviation,
and N is the number of observations.

2.6. Chapter 2 Bibliography

- [1] Shaw, P.J., 1995. Comparison of wide-field/deconvolution and confocal microscopy for 3D imaging. In *Handbook of biological confocal microscopy*(pp. 373-387). Springer US.
- [2] Pawley, J.B., 2006. Fundamental limits in confocal microscopy. In *Handbook of biological confocal microscopy* (pp. 20-42). Springer US.
- [3] Semwogerere, D. and Weeks, E.R., 2005. Confocal microscopy. *Encyclopedia of Biomaterials and Biomedical Engineering*, 23, pp.1-10.
- [4] Rai, V. and Dey, N., 2011. The basics of confocal microscopy. In *Laser Scanning, Theory and Applications*. InTech.
- [5] Yin, C., Glaser, A.K., Leigh, S.Y., Chen, Y., Wei, L., Pillai, P.C.S., Rosenberg, M.C., Abeytunge, S., Peterson, G., Glazowski, C. and Sanai, N., 2016. Miniature in vivo MEMS-based line-scanned dual-axis confocal microscope for point-of-care pathology. *Biomedical optics express*, 7(2), pp.251-263.
- [6] Piyawattanametha, W. and Wang, T.D., 2010. MEMS-based dual-axes confocal microendoscopy. *IEEE Journal of Selected Topics in Quantum Electronics*, 16(4), pp.804-814.
- [7] Wang, T.D., Mandella, M.J., Contag, C.H. and Kino, G.S., 2003. Dual-axis confocal microscope for high-resolution in vivo imaging. *Optics letters*, 28(6), pp.414-416.

- [8] Wang, T.D., Contag, C.H., Mandella, M.J., Chan, N.Y. and Kino, G.S., 2004. Confocal fluorescence microscope with dual-axis architecture and biaxial postobjective scanning. *Journal of biomedical optics*, 9(4), pp.735-742.
- [9] Curiel-Lewandrowski, C., Williams, C.M., Swindells, K.J., Tahan, S.R., Astner, S., Frankenthaler, R.A. and González, S., 2004. Use of in vivo confocal microscopy in malignant melanoma: an aid in diagnosis and assessment of surgical and nonsurgical therapeutic approaches. *Archives of dermatology*, 140(9), pp.1127-1132.
- [10] Flores, E.S., Cordova, M., Kose, K., Phillips, W., Rossi, A., Nehal, K. and Rajadhyaksha, M., 2015. Intraoperative imaging during Mohs surgery with reflectance confocal microscopy: initial clinical experience. *Journal of biomedical optics*, 20(6), pp.061103-061103.
- [11] Gad-el-Hak, M. ed., 2001. The MEMS handbook. CRC press.
- [12] Suri, JS, Chang, R, & Kathuria, C (eds) 2008, Advances in Diagnostic and Therapeutic Ultrasound Imaging, Artech House, Norwood. pp.37-63 Available from: ProQuest Ebook Central. [8 August 2017].
- [13] Loizou, C.P., Pattichis, C.S., Christodoulou, C.I., Istepanian, R.S., Pantziaris, M. and Nicolaides, A., 2005. Comparative evaluation of despeckle filtering in ultrasound imaging of the carotid artery. *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, 52(10), pp.1653-1669.
- [14] Christodoulou, C.I., Pattichis, C.S., Pantziaris, M. and Nicolaides, A., 2003. Texture-based classification of atherosclerotic carotid plaques. *IEEE transactions on medical imaging*, 22(7), pp.902-912.
- [15] Crimmins, T.R., 1986. Geometric filter for reducing speckle. *Optical Engineering*, 25(5), pp.651-654.
- [16] Wang, Z., Bovik, A.C., Sheikh, H.R. and Simoncelli, E.P., 2004. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4), pp.600-612. [8 August 2017]
- [17] Available from : https://en.wikipedia.org/wiki/Structural_similarity [8 August 2017]
- [18] Attlas, N. and Gupta, D.S., 2014. Wavelet Based Techniques for Speckle Noise Reduction in Ultrasound Images. *Nishtha Atlas et al Int. Journal of Engineering Research and Applications*, ISSN, pp.2248-9622.

- [19] Mateo, J.L. and Fernández-Caballero, A., 2009. Finding out general tendencies in speckle noise reduction in ultrasound images. *Expert systems with applications*, 36(4), pp.7786-7797.
- [20] Kaur, A. and Singh, K., 2010, March. Speckle noise reduction by using wavelets. In *National Conference on Computational Instrumentation CSIO NCCI* (pp. 198-203).
- [21] Huang, T., Yang, G.J.T.G.Y. and Tang, G., 1979. A fast two-dimensional median filtering algorithm. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 27(1), pp.13-18.
- [22] Lim, J.S., 1990. Two-dimensional signal and image processing. *Englewood Cliffs, NJ, Prentice Hall, 1990, 710 p*
- [23] Available from : <https://jiteshgupta1192.files.wordpress.com/2013/11/image-denoising.pdf> [2 August 2017]
- [24] Tomasi, C. and Manduchi, R., 1998, January. Bilateral filtering for gray and color images. In *Computer Vision, 1998. Sixth International Conference on*(pp. 839-846). IEEE.
- [25] He, K., Sun, J. and Tang, X., 2013. Guided image filtering. *IEEE transactions on pattern analysis and machine intelligence*, 35(6), pp.1397-1409.
- [26] Liu, C., Freeman, W.T., Szeliski, R. and Kang, S.B., 2006, June. Noise estimation from a single image. In *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on* (Vol. 1, pp. 901-908). IEEE.
- [27] Shi, H., Chen, L., Mai, Z.H., Chen, H. and Bi, F.K., 2015. The guided image filtering for speckle reduction in SAR images.
- [28] Engeldrum, P.G., 2000. *Psychometric scaling: a toolkit for imaging systems development*. Imcotek.
- [29] David, H.A., 1963. *The method of paired comparisons* (Vol. 12). London.
- [30] Available from : <https://www.interaction-design.org/literature/article/extreme-aversion-bias-sometimes-the-risk-is-worth-the-reward> [20 August 2017]
- [31] CIE. Guidelines for the evaluation of gamut mapping algorithms. Technical Report ISBN : 3-901-906-26-6, CIE TC8-03, 156 :2004.
- [32] Engeldrum, P.G., 2001, April. Psychometric scaling: avoiding the pitfalls and hazards. In *PICS* (pp. 101-107).

[33] Available at: <http://www.ansatt.hig.no/mariusp/quick>

[34] Thurstone, L.L., 1927. A law of comparative judgment. *Psychological review*, 34(4), p.273.

[35] Bartleson, C.J. and Grum, F. eds., 1984. *Optical Radiation Measurements: A Treatise. Vol. 5, Visual Measurements*. Academic Press.

[36] Easton, V.J. and McColl, J.H., 1997. *Statistics Glossary* v1. 1.

Chapter 3: Approach used within the project

3.1. Overview of the approach

“Early attempts to suppress speckle noise were implemented by averaging of uncorrelated images of the same tissue under different spatial positions. Although these methods are effective for speckle reduction, they require multiple images of the same object to be obtained.” [1]

As mentioned earlier, the idea of my approach is to make use of the redundant data present in the consecutive frames of the video sequence recorded by the dual-axis confocal microscope. While capturing several images might make the process too complicated, the overlapping areas are inherently present in the consecutive frames of the video sequence. For illustration, refer to *Figure 3.1.1.* below :

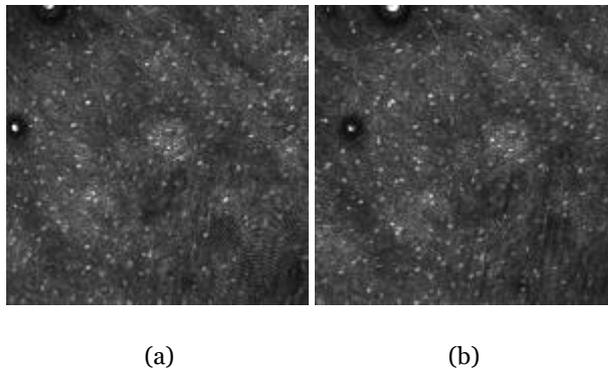


Figure 3.1.1. (a) and (b) examples of the two consecutive frames extracted from the video sequence

Figure 3.1.2. below represents the mosaic of the two images mentioned in the figure above (Figure 3.1.1. (a) and (b)), automatically registered through the software provided by the MSKCC (Memorial Sloan Kettering Cancer Center) by the partner laboratory. The details about the software cannot be revealed due to confidentiality agreement, but the general information could be found in the references below [2][3]. The overlapping utilizable area is marked with the red frame on *Figure 3.1.2.* Indeed the general overlap is bigger, but there were several limitations that will be named below.

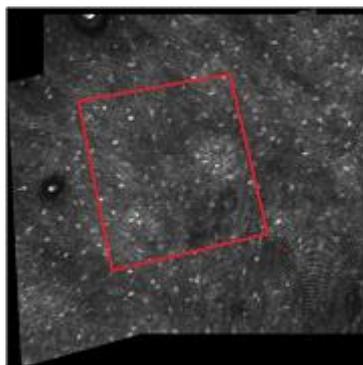


Figure 3.1.2. – overlapping and utilizable region marked with red frame

Several challenges and limitations are worth mentioning :

First of all, as mentioned in [4], due to curved nature of the objective lens, the edges of the image are out of focus and are useless for our task. Therefore, just the central parts of the image could be utilized.

Secondly, due to the fact mentioned above and the fast movement of the handheld microscope, sometimes overlaps between consecutive frames are too small to be used for making some significant improvements in terms of speckle noise.

And thirdly, automatic registration of the consecutive frames are sometimes extremely challenging and even impossible. The first reason is that the microscope is constantly moving and tilting as it is manually held and moved across the tissue by a human-operator. Furthermore, the tissue as a living object, is never completely static during the in vivo scanning and is deforming from time to time. Therefore, even if the overlap is big between the two consecutive frames, the transform needed for registration can differ dramatically between different parts of the same image. The situation is even worsened with the fact that near-to-perfect registration is needed, as one of the fundamental techniques within this approach of the speckle mitigation is pixelwise averaging of the overlapping regions. And in case the alignment is poor, we will definitely end up with apparent artifacts biasing the result even further and making the outcome even more complicated for the physician. Finally, another constraint was induced due to the fact that the tissue images were too complex, without apparent characteristic structures, and detection of the matching points was not possible by the system (discussed later in this chapter).

The exact details for the approach used and the workaround found for above mentioned problems are discussed below in this chapter.

3.2. Registration Problem

The pivotal part of the approach is to extract the redundant data, the overlapping regions depicting the same part of the tissue present in both frames. The idea sounds

simple: average the overlapping regions in the consecutive frames of the video sequence. There are two key challenges here: firstly, how to extract the redundant data, how to find the overlap and corresponding pixels between the two frames? It is important to consider, as mentioned earlier, that near-perfect registration is needed for pixel-wise averaging of the parts of the images. And second: after best possible alignment is reached and redundant data is extracted, how should this data be utilized? Is pixel-wise average the best approach, or can more sophisticated ways of processing improve the result, or can they even mitigate the artifacts created due to some extent of misalignment?

The second question will be addressed later. As for the registration problem, that was caused irregular movements of the handheld microscope and constant deformations of the tissue during in vivo scanning.

The primary aim of the project was finding the ways of speckle mitigation through utilization of the redundant data in the overlapping regions. As the primary goal of the project was not studying the ways of accessing those redundant data, the first idea was simply to use the code of the software of our partner laboratory at the Memorial Sloan Kettering Cancer Center [2][3], which was thought that was detecting the overlapping areas, registering the two images and discarding some redundant data that was the primary point of our interest. But during the examination of the code, it turned out that the software extensively uses stitching technique, where optimal seam is detected and two consecutive frames are simply stitched across this seam. In this case, the software preserves some data from either frame, and no redundant data and overlapping information is accessible.

It became obvious that I had to tackle the registration problem myself. Initially, conventional ways of feature detection – SIFT [5][6] and SURF [7] were tried, but due to the limitations and complexities described above, the registration of the images went sometimes terribly wrong, apparent even to the naked eye, or was not possible at all (*Figure 3.2.1.- as we see on the figure, the overlapping area is blurry, inconsistent with the true movement of the microscope and creating the artifacts with unnatural seams arisen*), or due to insufficient number of detected control points (*Figure 3.2.2.*).

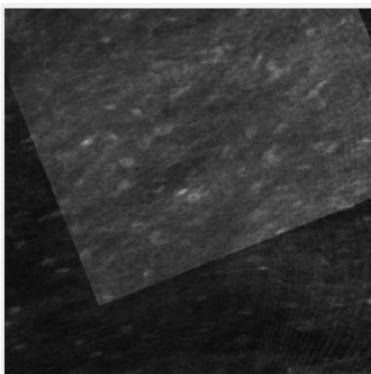
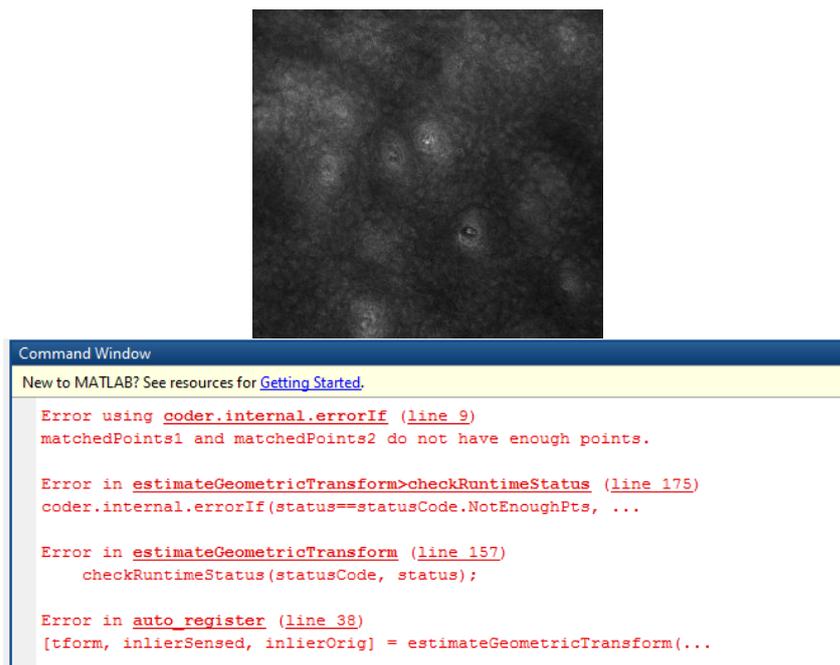


Figure 3.2.1. – The automatic registration went terribly wrong



(a) (b)

Figure 3.2.2. – Sometimes the registration was not possible at all, as the frame was too complicated to find enough matching points. (a) the frame. (b) screenshot from the MATLAB software documenting the registration failure.

The workaround found for this problem, was the following : the dataset was split into two types of images :

1. The frames, where no single transform can lead to satisfactory registration of the images. To solve this problem, the large images were split into smaller square patches, translation was assumed between those patches for simplicity’s sake and the registration was carried out manually.
2. The frames, what are recorded with very smooth and slow movement of the microscope that enables us to automatically and reliably register the two consecutive frames with nearly-perfect registration that enables us find the pixelwise average without creating substantial artifacts.

The both approaches are discussed in the sub-chapters below.

3.3. Approach Based on the Smaller Patches

As mentioned in the sub-chapter above, the larger images were split into smaller patches where automatic registration of the whole images was not possible. Different sizes, like 32-by-32 pixels or 128-by-128 pixels were examined for the patches, but the empirical study has shown that the optimal size was 64-by-64 pixels. The larger patches were difficult to register well enough, while smaller patches did not cover

enough key points necessary for the registration. The out of focus part of the images was discarded and the remaining part of the frame was split into 64-by-64 pixel patches. Due to limitedness of time and resources, simple translation was assumed between those locally selected sub-regions of the two frames. Afterwards, manually selected corresponding sub-regions were manually registered with the help of Adobe Photoshop and GIMP softwares. In this case, the study of the speckle mitigation was prioritized and automatization of the registration with more sophisticated tools was left open for future projects.

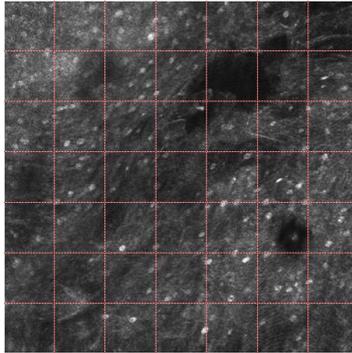


Figure 3.3.1. – In-focus part of the frame was split into 64-by-64 smaller patches

The idea of manual registration was that the patch from the first frame was registered manually against the second frame and the corresponding patch was cropped out, as the patch from the first frame and the second, full frame were overlaid one over another. The opacity of the patch was decreased and the the key structures were matched manually, as good as possible. The coordinates of the overlapping regions was detected and later processed in MATLAB software.

The principle is demonstrated on the *Figure 3.3.2.* below (for illustration's sake, larger patches are shown on the figure, but in practice, it was further split into 64-by-64 pixel sub-regions).

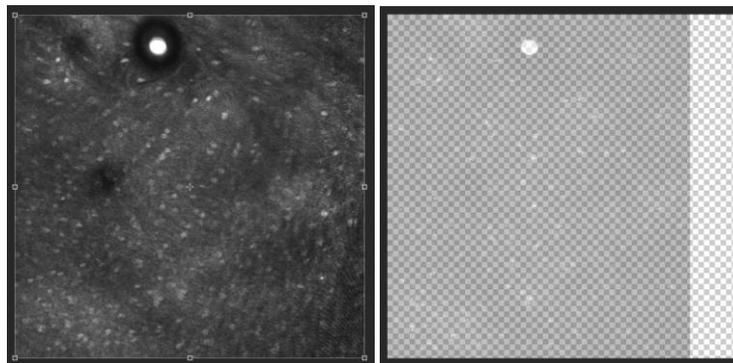


Figure 3.3.2. – Consecutive frames were placed on top of one another and after decreasing the opacity of the top one, the registration was carried out manually.

In some particular cases, even manual registration was not perfect that led to the substantial artifacts above I was talking about. Refer to the figures below:

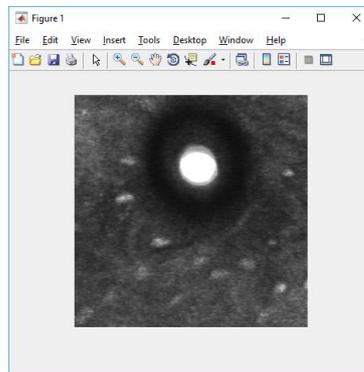


Figure 3.3.3. – The pixelwise average of two manually best possibly aligned patches

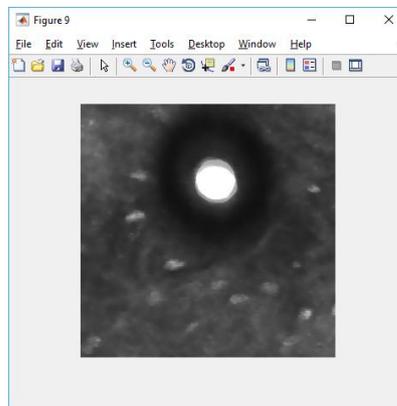


Figure 3.3.4. - The bilateral filtered version the pixelwise average illustrated in Figure 3.3.3. above

Finally, in total 10 pairs of the corresponding patches were selected as a focus set for the examination. The patches were selected from the frames of different types of tissues with various mean luminance, speckle noise level and structural layout to make the generalization of the results more reliable.

Another point of interest was to study, how much does misalignment influence the final result and what extent of misalignment can be considered satisfactory. In order to study that, the patches were intentionally misaligned after best possible manual alignment by 3, 7 and 12 pixels. The direction of the misalignment was randomly selected. The example set can be seen on the *Figure 3.3.5.* below, where

the first column is the patch cropped from the first frame.

The second column is the corresponding part from the second frame after best possible alignment.

The third, fourth and fifth columns depict misaligned “corresponding” areas from the second frame after 3-, 7- and 12 pixels of misalignment respectively. It is worth mentioning that those numbers, 3 pixel, 7 pixel and 12 pixel were selected empirically, as misalignment beyond 12 pixel led to extremely biased results, while misalignment of 1 pixel had little effect on the final outcome.

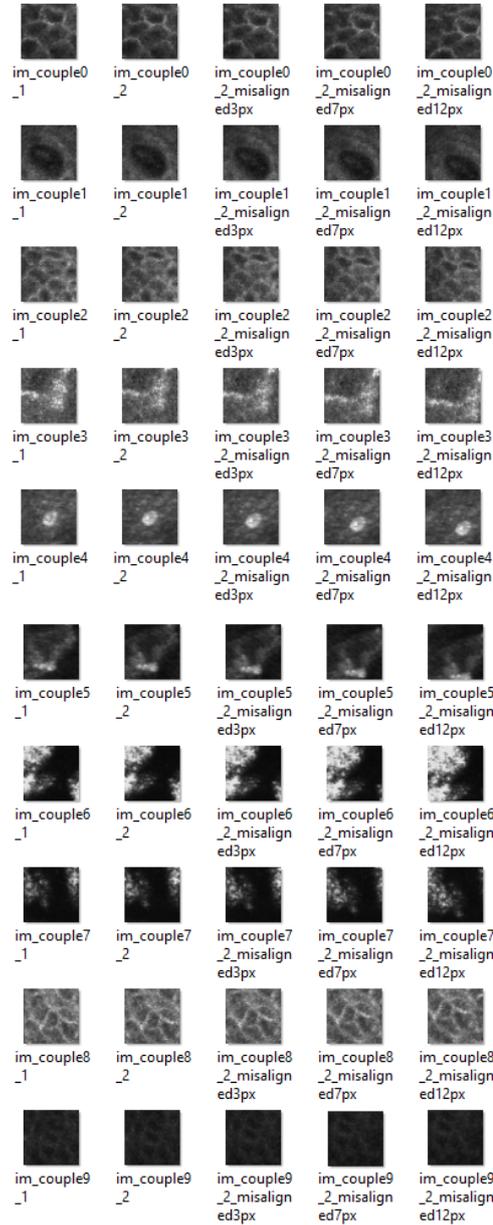


Figure 3.3.5. – 1st column – patch from frame 1 ; 2nd column – corresponding patch from frame 2 ; 3rd column – corresponding patch from frame 2 misaligned by 3 px ; 4th column – corresponding patch from frame 2 misaligned by 7 px ; 5th column – corresponding patch from frame 2 misaligned by 12 px ;

So, in total, we come up with 4 different pairs of images from the two frames: one with best-possible alignment, and three with varying degree of misalignment.

Various methods of utilization of the overlapping data were studied. They can be divided into two groups: the ones, where only single frame is used, and in this case, this is the frame present in column 1 in *Figure 3.3.5*. - or the approaches, where information from both frames are used. 3D median filter is one example of those, where in contrast to the 2D median described in *sub-chapter 2.3.1*, the values from both overlapping regions are used and the median is found among 18 values, instead of 9.

Special point of interest was optimal selection of the input and guidance images in guided image filtering. The approaches or the combinations of the approaches are summarized in the *Table 3.3.1*.

<i>Using single frame only</i>
<i>2D median filtering</i>
<i>Gaussian Filtering</i>
<i>Average Filtering (average instead of median)</i>
<i>Bilateral Filtering</i>
<i>Guided Filter – 1st frame guided by itself</i>
<i>Guided Filter – 1st frame guided by median of itself</i>
<i>Guided Filter – 1st frame guided by average of itself</i>
<i>Using double-frame approach</i>
<i>3D median filtering</i>
<i>Point-wise Average</i>
<i>Guided Filter – 2nd frame guided by the 1st frame</i>
<i>Guided Filter – 2nd frame guided by the median of the 1st frame</i>
<i>Guided Filter – Point-wise Average guided by the 1st frame</i>

Table 3.3.1. – summarizing the ways of speckle mitigation

As mentioned above, the parameters can be tuned for some of the functions that can itself change the performance dramatically. The performance and influence under several different parameters was studied empirically and the best-performing ones were brought into the report for comparison.

For median filtering, the default 3-by-3 neighborhood was found optimal, while for Gaussian filtering, the neighborhood was defined as 5-by-5, while standard deviation was left as a default value, equal to 0.5.

For bilateral filter, spatial distance was defined as 3-pixels, while the isotropic (equal in all directions) standard deviation was set to 0.1. For guided image filtering, the default values are 5-by-5 pixel window, and a default regularization parameter is defined as:

$$0.01 * \text{diff}(\text{getrangefromclass}(G)).^2$$

Equation 3.3.1.

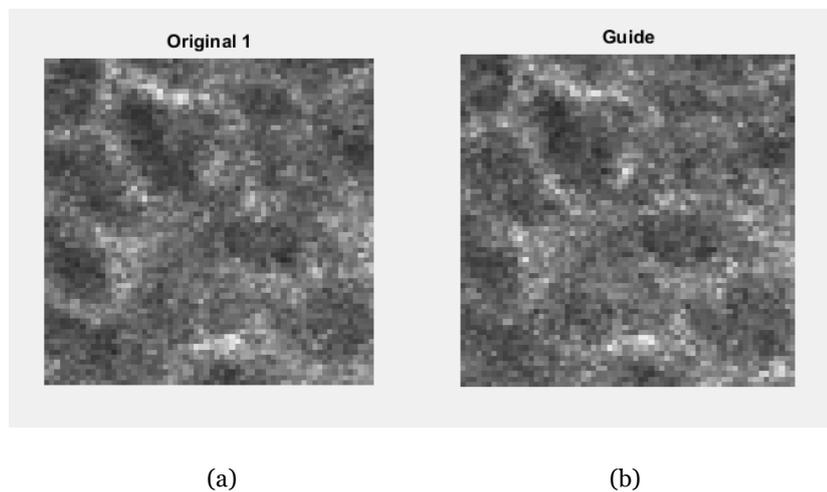
where,

$\text{diff}(X)$ – is a function which calculates differences between adjacent elements of X along the first array dimension whose size does not equal to 1.

$\text{getrangefromclass}(I)$ – is a function which returns the default display range of the image I , based on its class type.

The default parameters turned out to be too blurring. The visual evaluation was held to compare the combination of two possible window sizes 3-by-3 and 5-by-5 and three possible regularization parameters: the default value, 10% of the default value, and 1% of the default value.

The visual examination has shown that the optimal parameters were 3-by-3 window and 1% of the default regularization parameters, as they kept the key structures necessary for medical evaluation and hence, those values were applied for further study. The examples can be seen on the *Figure 3.3.6.* below:



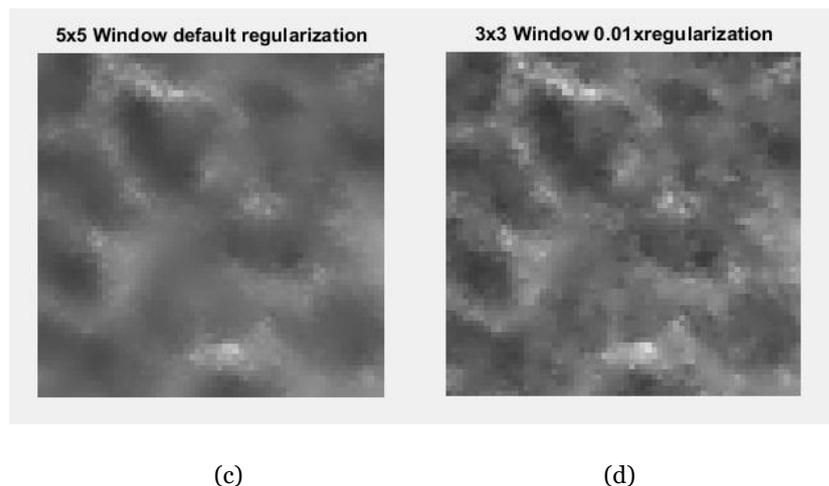


Figure 3.3.6. – (a) original image, (b) guidance image (c) default parameters for guided image filtering (d) selected parameters as optimal

While splitting the images into smaller patches, it was thought that they could be stitched back together seamlessly, assuming near-perfect alignment. But another problem arose, as the practice has shown that sometimes blocking artifacts are visible to a naked eye, when stitching the 64-by-64 patches after separate processing. On the illustration below (Figure 3.3.7), the seams, the blocking artifacts are visible more in case of pixelwise averaging approach, while it is less salient for guided image filtering.

In any case, this question remains open and makes us conclude that if possible, the automatic registration and processing of the whole image should be applied, while splitting into blocks should be implemented only in the extreme cases, when no automatic alignment is possible at all. But the question about techniques and improvements in automatic registration of the larger images remains open for follow-up work.

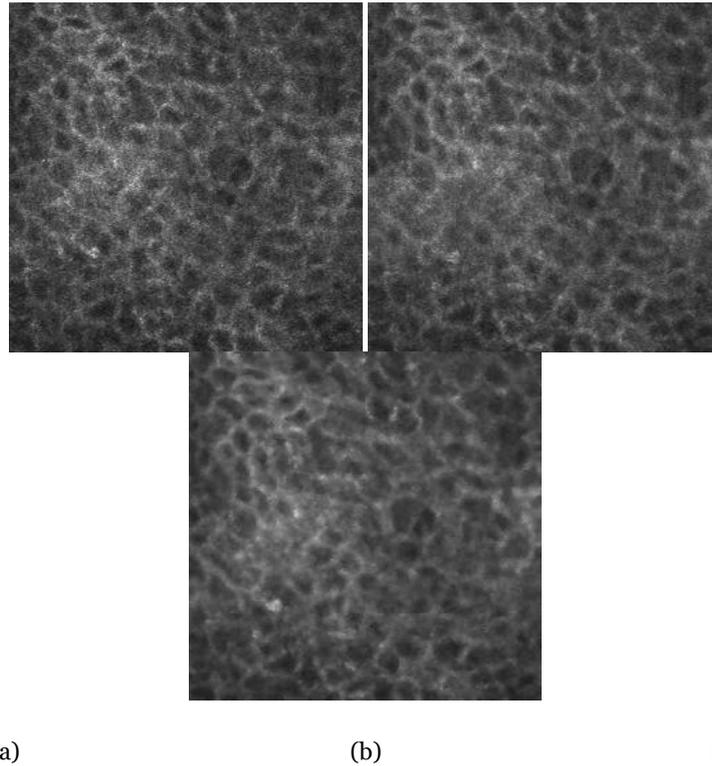


Figure 3.3.7. – (a) original 256-by-256 pixel image (b) the image in (a) reconstructed by stitching of 16 64-by-64 pixel patches that are separate point-wise average of the 64-by-64 pixel the two frames. (c) the image in (a) reconstructed by stitching of 16 64-by-64 pixel patches that are the result of filtering the second frame guided by the first one.

3.4. Full Image Processing (automatic registration) Approach

In some cases, that was unfortunately the minority of the cases in the dataset available during the project, we have insignificant deformations of the tissue and very smooth and slow motion of the microscope. In this case, very slight difference is present between two consecutive frames and they can be automatically registered and the overlapping regions can be automatically found, without further simplification of the task (splitting into smaller patches and assuming translation).

In those cases, similarity transform was found sufficient and the SURF features were used for automatic alignment of the larger images. Even though the texture of some images is still quite complicated, the system successfully found enough matching points (*Figure 3.4.1.*).

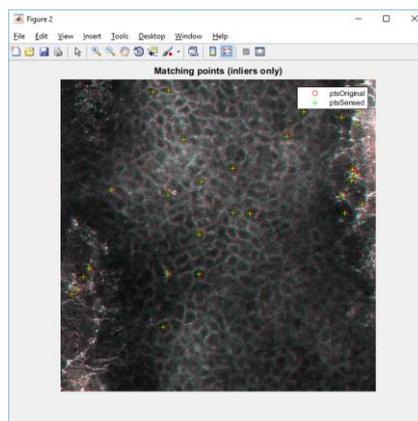


Figure 3.4.1.- the system successfully found enough matching points

Finally, the same processing techniques were applied to 256-by-256 images, that were used in case of smaller patches, but with one significant difference – no back-stitching was needed in this case and the results are seamless and free from blocking artifacts, even if we compare them against single-image solutions. One of the illustrations of this, can be found on the *Figure 3.4.2.* below:

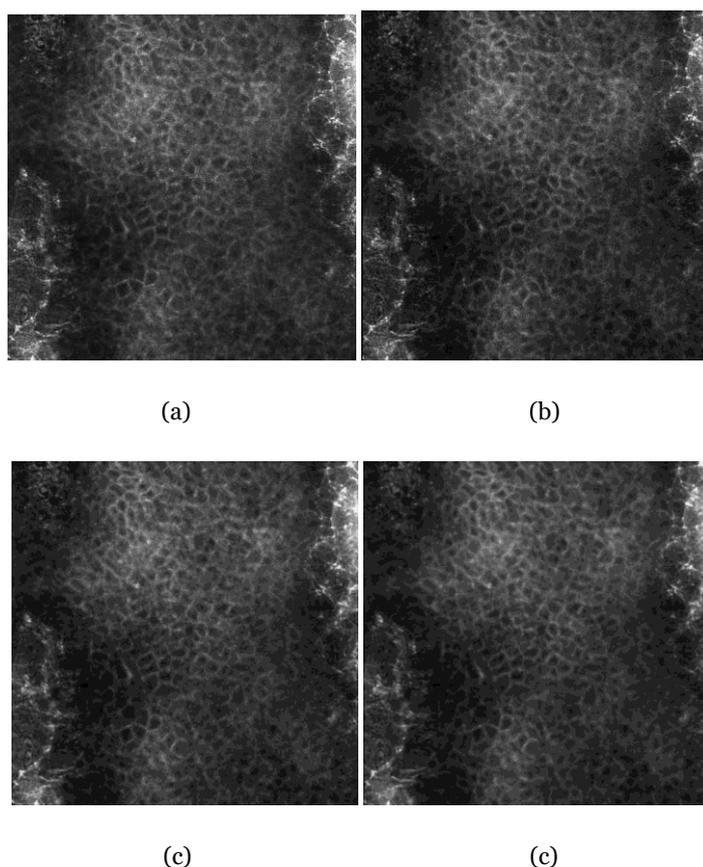


Figure 3.4.2 – (a) Original frame 1 (b) Original frame 2 (c) 2D-median filtered version (using frame 1 (a) only) (d)3D-median filtered version, using both ((a) and (b)) frames

3.5. Chapter 3 Bibliography

- [1] Suri, JS, Chang, R, & Kathuria, C (eds) 2008, *Advances in Diagnostic and Therapeutic Ultrasound Imaging*, Artech House, Norwood. pp.37-38 Available from: ProQuest Ebook Central. [8 August 2017].
- [2] Kose, K., Cordova, M., Duffy, M., Flores, E.S., Brooks, D.H. and Rajadhyaksha, M., 2014. Video-mosaicing of reflectance confocal images for rapid examination of large areas of skin in vivo. *The British journal of dermatology*, 171(5), p.1239.
- [3] Floresa, E., Yelamosa, O., Cordovaa, M., Kosea, K., Phillipsa, W., Rossia, A., Nehala, K. and Rajadhyakshaa, M., 2017, February. Peri-operative imaging of cancer margins with reflectance confocal microscopy during Mohs micrographic surgery: feasibility of a video-mosaicing algorithm. In *Proc. of SPIE Vol (Vol. 10037, pp. 100370B-1)*.
- [4] Yin, C., Glaser, A.K., Leigh, S.Y., Chen, Y., Wei, L., Pillai, P.C.S., Rosenberg, M.C., Abeytunge, S., Peterson, G., Glazowski, C. and Sanai, N., 2016. Miniature in vivo MEMS-based line-scanned dual-axis confocal microscope for point-of-care pathology. *Biomedical optics express*, 7(2), pp.251-263.
- [5] Lowe, D.G., 2004. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2), pp.91-110.
- [6] Lowe, D.G., 1999. Object recognition from local scale-invariant features. In *Computer vision, 1999. The proceedings of the seventh IEEE international conference on (Vol. 2, pp. 1150-1157)*. Ieee.
- [7] Bay, H., Ess, A., Tuytelaars, T. and Van Gool, L., 2008. Speeded-up robust features (SURF). *Computer vision and image understanding*, 110(3), pp.346-359.

Chapter 4: Results

4.1. Smaller Patches

The ten sample patches were processed by twelve speckle mitigation techniques, and their speckle index together with that of original were found. Besides, the performance was evaluated using Structural Similarity. Furthermore, each case was reviewed for well aligned cases, as well as different degrees of gradual misalignment.

As we remember from *Chapter 2*, lower value of speckle index, means less amount of speckle noise. But we should not be tempted with too low values, as it might mean too blurry images with key structures and fine details lost. Therefore, the most interesting picture will be comparing speckle index against structural similarity. But before that, it will be still interesting to study the behavior of each metric under various conditions.

Let's start by reviewing the statistics given in the figures. Each line represents each pair of the images – 10 in total. Let's refer to the table below to clear up how each speckle mitigation technique is mentioned on the figures.

Speckle Mitigation Approach	How it is abbreviated on the figures
<i>Median Filtering using 1 image – 2D median filtering</i>	<i>Med-1</i>
<i>Median Filtering using 2 images – 3D median filtering</i>	<i>Med-2</i>
<i>Bilateral filtering</i>	<i>Bilateral</i>
<i>Gaussian filtering</i>	<i>Gauss</i>
<i>Average filtering (average using single image)</i>	<i>Avg-1</i>
<i>Poinwise average (average using two images)</i>	<i>Avg-2</i>
<i>Guided Image Filtering – 1st image filtered guided by itself</i>	<i>Guide-1-1</i>
<i>Guided Image Filtering – 1st image filtered guided by the median of itself</i>	<i>Guide-1-med</i>
<i>Guided Image Filtering – 1st image filtered guided by the average of itself</i>	<i>Guide-1-Avg</i>
<i>Guided Image Filtering – 2nd image filtered guided by the first image</i>	<i>Guide-2-1</i>
<i>Guided Image Filtering – 2nd image filtered guided by the median of the first image</i>	<i>Guide-2-1med</i>

<i>Guided Image Filtering – 2nd image filtered guided by the average of the first image</i>	<i>Guide-Avg1</i>
<i>Original Image (first frame)</i>	<i>Original</i>

Table 4.1.1.-abbreviations used for speckle mitigation algorithms

Let's start by reviewing *Figure 4.1.1.*

The first thing that can be concluded from this figure is that even though there are a pair of exceptions, the shapes of the plots for all then pairs of the images is similar – that means that the performance of each particular speckle mitigation technique is not dependent on the content of the medical image.

It is quite logical that highest speckle is present in the original images. This means that all the techniques blur it and remove speckle at some extent.

The most effective ones in terms of speckle index decrease are bilateral filter and guided filter when the image is filtered by its own average. On the other hand, the worst performing one is the case, when the image is filtered and identical image is used for guidance. In most of the cases, it is close to the original speckle level.

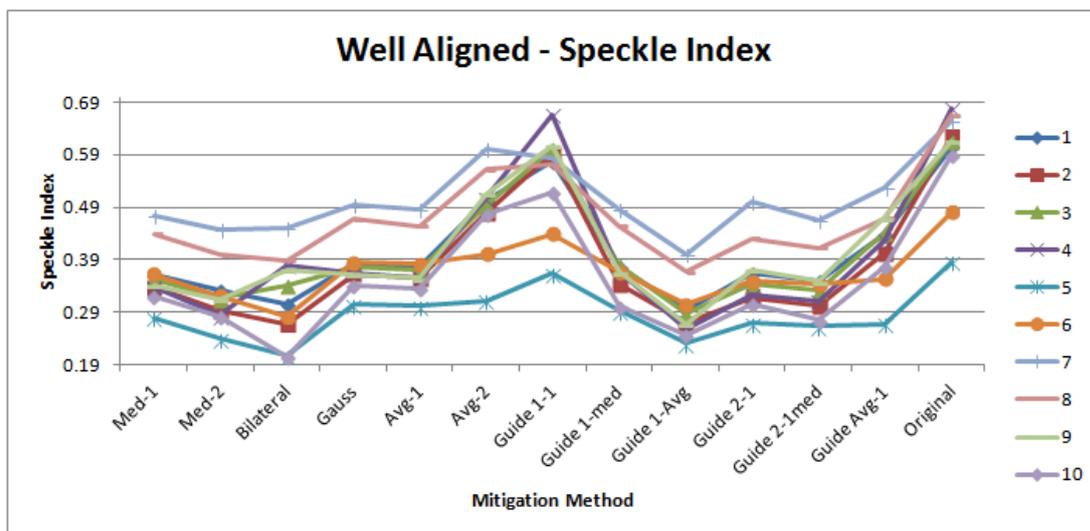


Figure 4.1.1. – Speckle Index for Well Aligned Samples. Each line represents separate pair of the sample images (patches)

It is quite logical that misalignment does not really affect the speckle mitigation extent. The similar traits are demonstrated by 3-,7- and 12-pixel misaligned pairs of the images. This can be observed on the figures 4.1.2 through 4.1.4.

However, all of them have slightly lower level of speckle in comparison of that of the original image. This can be explained with the fact that misalignment creates additional artifacts that in case of complex and high frequency structures can lead to higher degree of blur. But this conclusion needs further examination for more credibility.

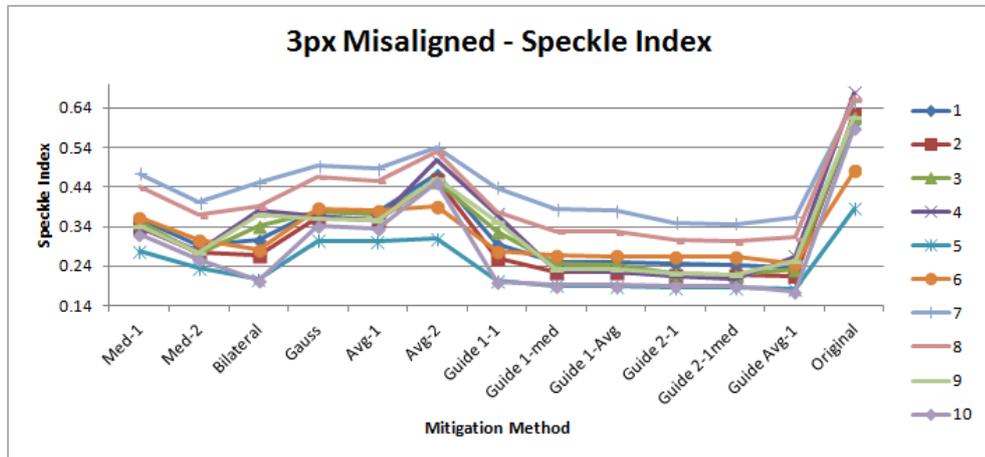


Figure 4.1.2. – Speckle Index for the samples with 3-pixel misalignment. Each line represents separate pair of the sample images (patches)

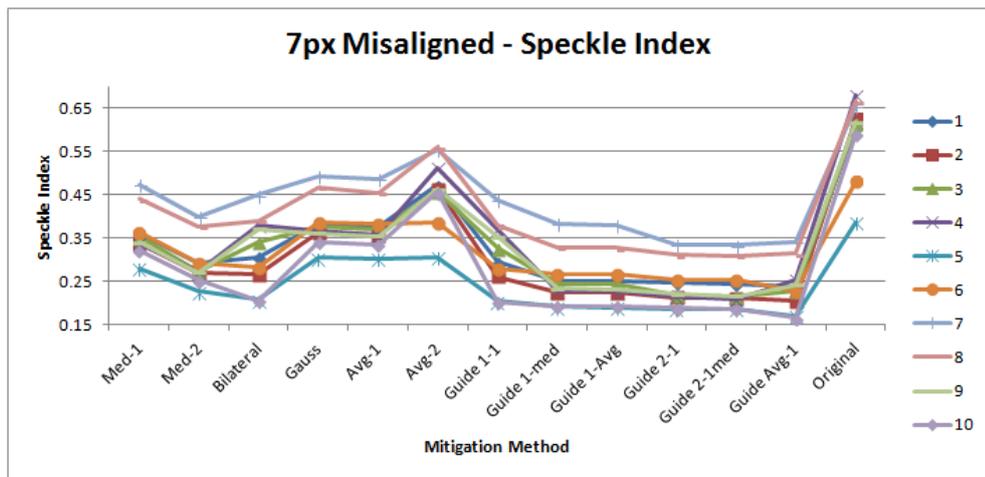


Figure 4.1.3. – Speckle Index for the samples with 7-pixel misalignment. Each line represents separate pair of the sample images (patches)

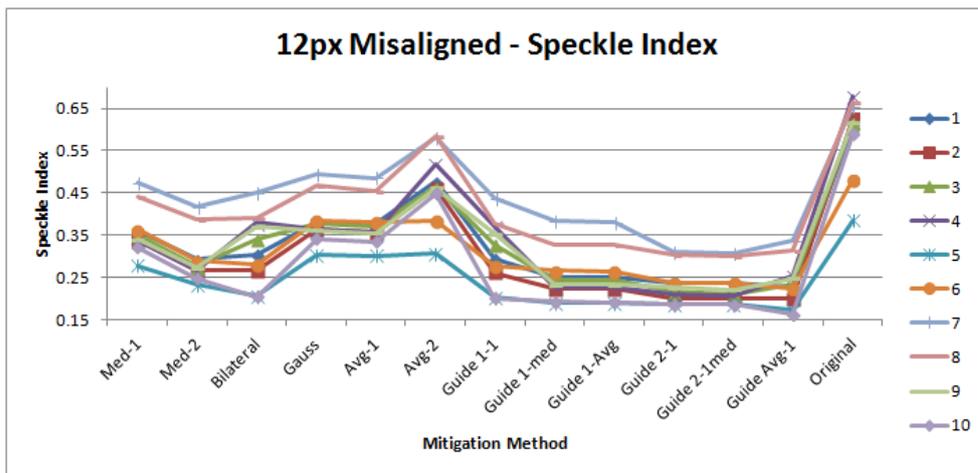


Figure 4.1.4. – Speckle Index for the samples with 12-pixel misalignment. Each line represents separate pair of the sample images (patches)

A way more appropriate method of evaluating, how misalignment influences the final result, is examining the structural similarity values. As mentioned in *Chapter 2*, higher value means higher structural similarity with 1 for the perfect match, while 0 is worst possible case in terms of structural similarity.

Figure 4.1.5. represents the values for all speckle mitigation techniques. This makes the picture too complicated, because it is logical that the approaches that use only single image, should have very high structural similarity, close to 1. The original is not represented on those plots, because it is indeed equal to 1 for all cases, as we compare and calculate the structural similarity in relation to the original frame. And therefore, its similarity with itself is always equal to 1.

If we draw parallels with the speckle index statistics, we see that in this case guided image filtering, when the image is filtered guided by itself, is still the extreme case and the most similar one to the original. Thus, we can already start concluding that this could be the worst performing approach. But we will return to this topic later in this subchapter.

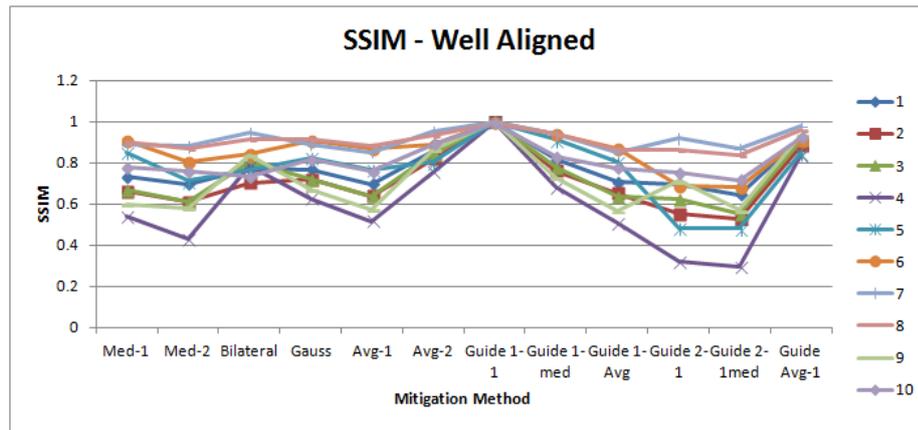


Figure 4.1.5. – Structural Similarity for well aligned pairs given for all speckle mitigation techniques.

In order to make the plot more visually understandable and clear, I left only double-image approaches on the plot. Refer to Figure 4.1.6. The shape among the approaches is still similar that makes us conclude that the performance of the processing techniques in terms of SSIM is neither too dependent on the content of the image.

It is interesting to note that the pixel-wise average option, as well as its filtered version the one using the first frame as a guidance is the best performing in terms of structural similarity. But it is quite logical, as this is cited in the literature discussed in the previous chapters as a “default” way of data processing in case of near-perfect alignment.

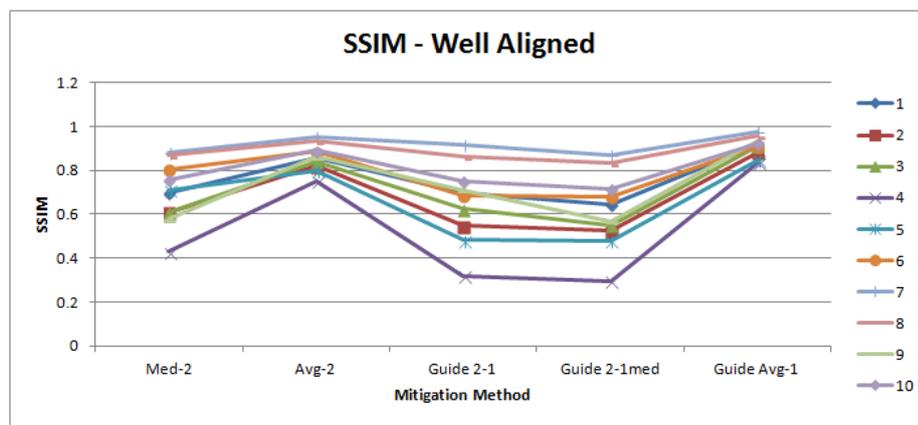


Figure 4.1.6. – Structural Similarity for well aligned pairs given for double-image-based speckle mitigation techniques.

But let’s have a look, how they perform when we induce some degree of misalignment. Refer to the figures 4.1.7. through 4.1.9.

It is quite understandable that generally SSIM values have decreased, but it is now a bit surprising that a point-wise average and its filtered version guided by the first frame are among the best. We have seen earlier in the report that in case of misalignment, point-wise average creates very apparent and substantial artifacts.

But the explanation for this is in nature of the guided image filtering. Even though in majority of the cases the structures from the guidance image is prioritized in the output image, the visual examination of those images in question has demonstrated that when the misalignment is substantial, then the structures are kept from the input image. As we have frame number 2 as an input image in those cases, we compare it against frame 1 and as the two are clearly misaligned, structural similarity is decreased, while point-wise average creates artifacts and intermediary version of the two, being closer with either in structure.

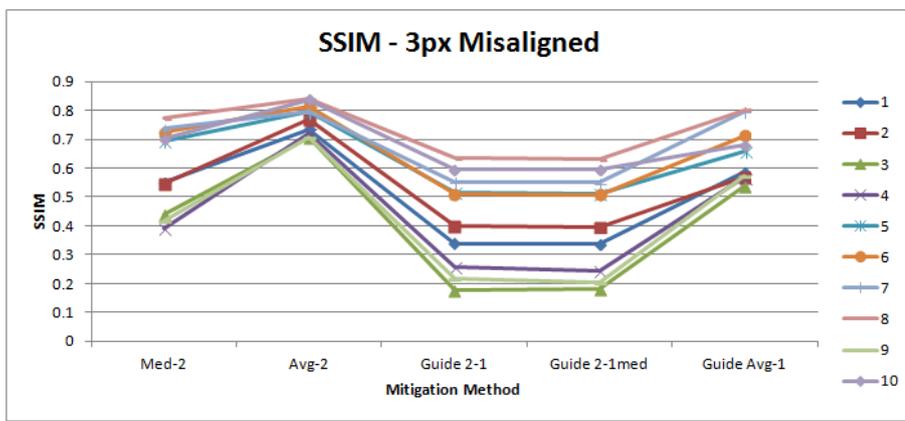


Figure 4.1.7. – Structural Similarity for 3-pixel misaligned pairs given for double-image-based speckle mitigation techniques.

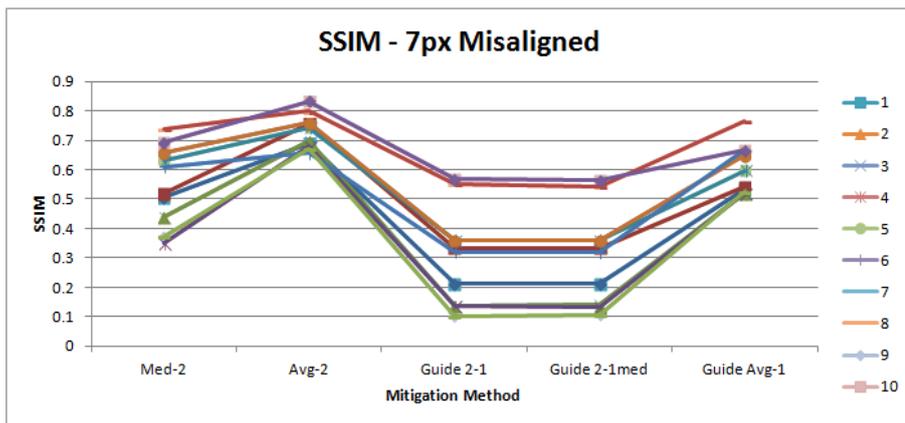


Figure 4.1.8. – Structural Similarity for 7-pixel misaligned pairs given for double-image-based speckle mitigation techniques.

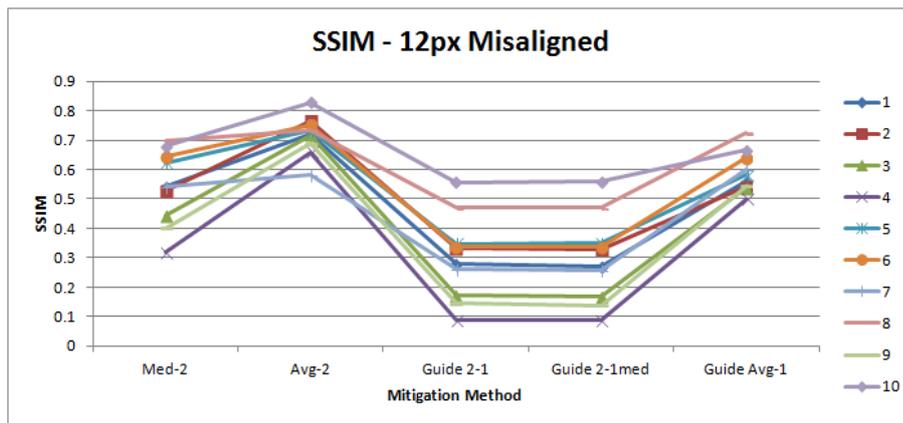


Figure 4.1.9. – Structural Similarity for 12-pixel misaligned pairs given for double-image-based speckle mitigation techniques.

The separate plots for each image enables us observe the trends and their dependence on the content of the image, but it is still difficult to perceive the global picture. Therefore, it is interesting to have a look at the average values.

Let’s refer to the figures 4.1.10. and 4.1.1. In those plots, as single line represents the single alignment-misalignment degree, while speckle index and SSIM values are the average of the 10 separate pairs of the images.

As we obviously see from Figure 4.1.10. the alignment issue does not have much influence on the speckle noise level and it is quite logical. The only interesting exception is guided image filtering in case of “perfect” alignment. In this case, slightly more speckle is remaining in the images. While median, Gaussian and bilateral filtering are unambiguously blurring the image, the structures and borders are sharply preserved when they match in input and guidance images. This happens due to the fact that we decreased the regularization parameter 100 times, thus blurring only low variance areas and keeping high variance areas intact and thus, keeping their speckle noise as well. The solution in this case could be gradual increasing of the regularization parameter, but in this case we have to control the performance in terms of SSIM as well.

The similar plot for structural similarity can be found on Figure 4.1.11. The shape is the same as in case of the plots for individual images, but the interesting phenomena are still observed here. It is very logical and obvious that SSIM is higher when we have good alignment, it is slightly declined when we induce 3-pixel misalignment, but interestingly, the misalignment extent does not matter that much, when we have 7- and more-pixel misalignment. In my opinion, the reason for this is the small size of the patches. Just to get back to the approach, we are processing 64-by-64 pixel patches. Therefore, misalignment of 7 or more pixels takes some key structures out of the corresponding patch and increases the dissimilarity dramatically regardless the fact,

how much was the misalignment further increased. The structures in two “corresponding” patches are already significantly different.

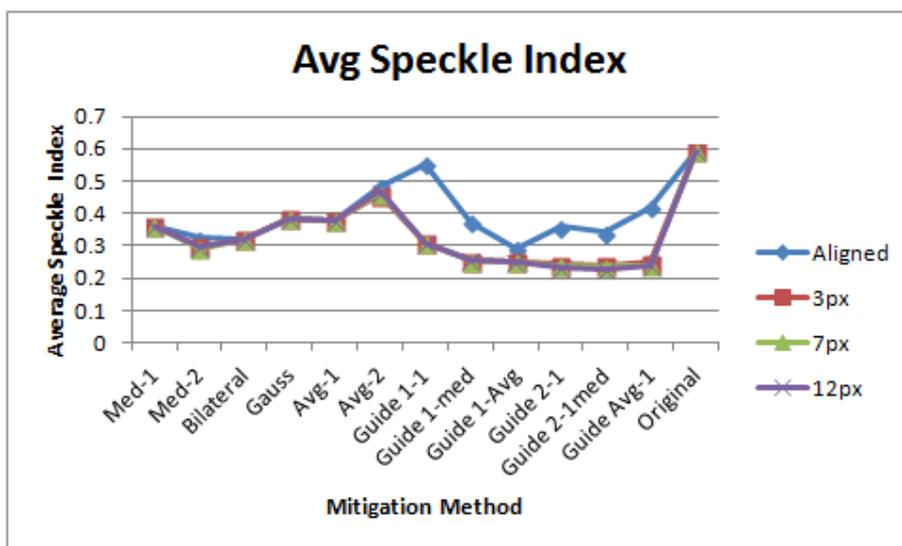


Figure 4.1.10.- speckle index – average of the 10 images

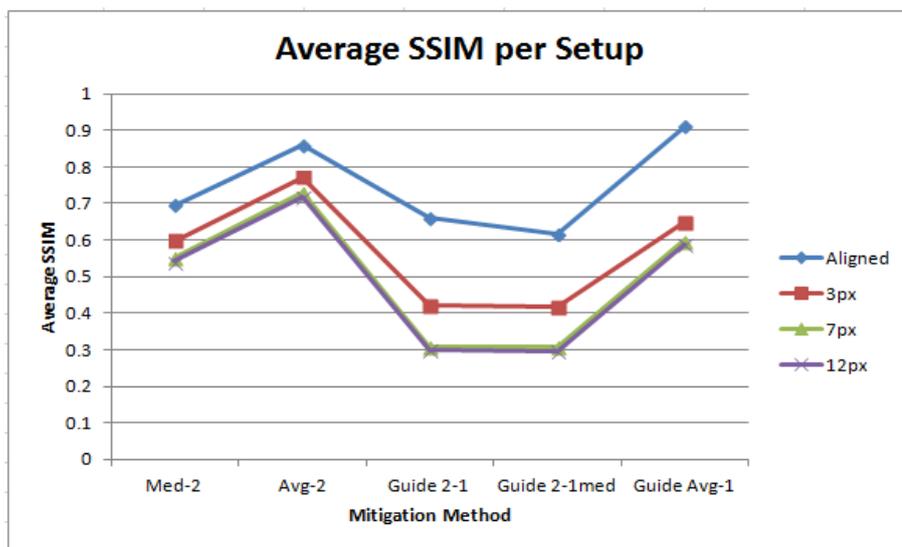


Figure 4.1.11.- structural similarity - average of the 10 images

The average values give us some interesting clues, but it is still interesting to look deeper into this data. The figures 4.1.12 and 4.1.13. illustrate not only average values, but maximal and minimal values among 10 images, as well as the standard deviation.

The majority of the approaches demonstrate quite consistent performance in terms of speckle index – while the mean of the original and the standard deviation are closer to the maximum meaning that we have more images with higher speckle index, this parameter is balanced in majority of the cases, where the mean is right in the halfway

between maximum and minimum – meaning proportionally more images with lower speckle index.

It is worth mentioning that the point-wise average and filtering the image guided by itself keep higher degree of speckle index. And wherever we have outliers for those approaches with minimum values, we can conclude that those are the images where the level of speckle was low even from the very beginning, in the original.

In terms of speckle index, filtering the image guided by the average of itself, as well as the simple 2D median filtering and bilateral filtering seem most promising. But for bilateral filter, we can observe slightly high standard deviation that means slightly higher instability.

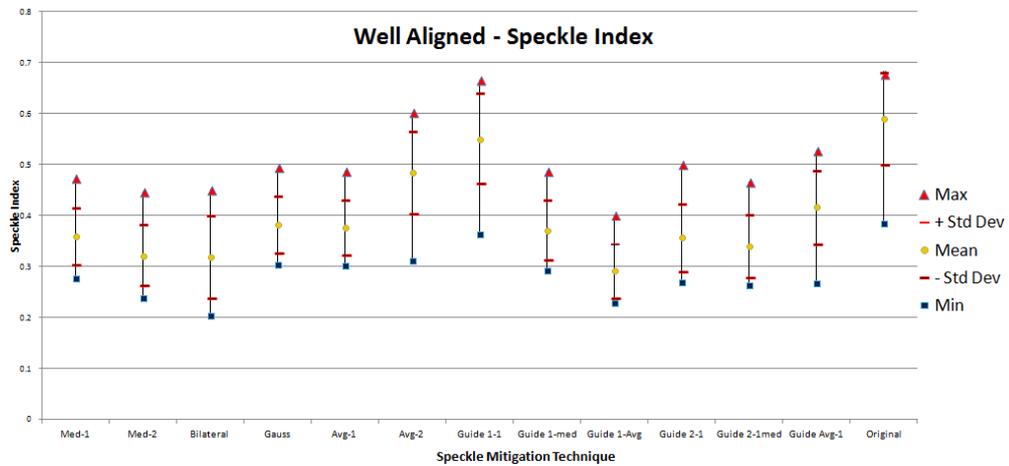


Figure 4.1.12 – Mean, maximum and minimum values of speckle index for each speckle mitigation, as well as the standard deviation, algorithm for well aligned images.

It is even more interesting to refer to the similar plot for SSIM.

The first fact that stands out of others is that while filtering the first image guided by itself, minimum and maximum values are almost equal, standard deviation is zero and all the values are close to 1. This means that filtering the image by itself leaves the structures almost intact, while still imposing some degree of blur. Filtering the point-wise average by the first frame also demonstrates very low standard deviation and high degree of keeping the structures.

Highest spread of the values and the largest standard deviation is observed in the cases where two different images are used as input and guidance in guided image filtering. This can be explained with the fact that the structures are prioritized sometimes from one, and sometimes from another image.

It is also worth mentioning that the performance of the bilateral filter is better than that of most of the approaches.

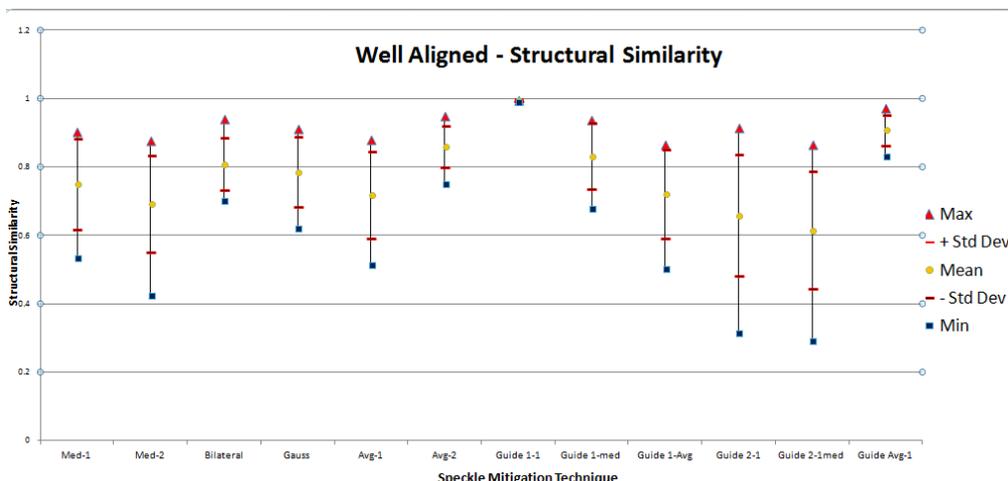


Figure 4.1.13 – Mean, maximum and minimum values of structural similarity, as well as the standard deviation, for each speckle mitigation algorithm for well aligned images.

Even though we were discussing performance in terms of SSIM and in terms of speckle reduction, the pivotal concept of the whole project was that the tradeoff between the two had to be found. Therefore, I think that the *Figure 4.1.14.* representing SSIM as a function of Speckle Index, is the most informative figure to evaluate the global picture.

The information is color coded on this plot.

- ◆ The red diamonds are for the approaches that use the information from two images.
- ◆ The sky blue diamonds are for the approaches that use the information from the single image only.

3D Average The red labels correspond to the pairs with satisfactory alignment.

3D Average The blue labels correspond to the pairs with 3-pixel misalignment.

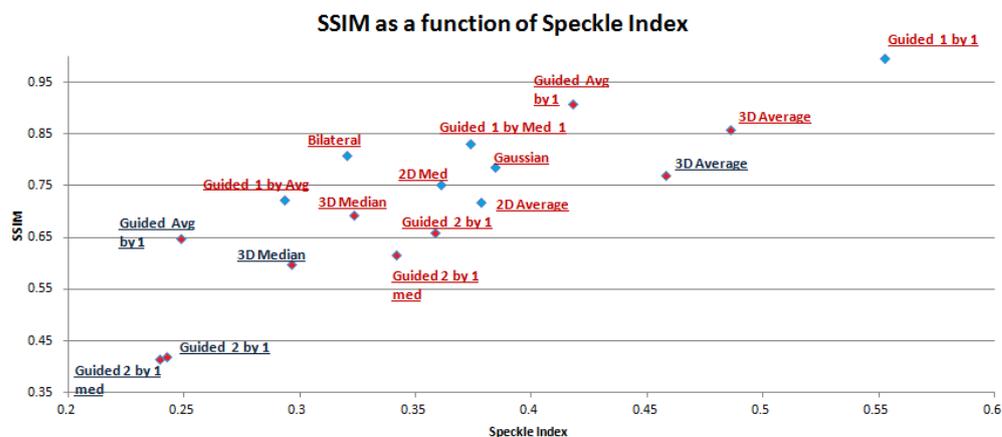


Figure 4.1.14. – SSIM as a function of Speckle Index

As mentioned earlier, low speckle index (horizontal axis) and high structural similarity (vertical axis) are optimal outcome and while improving one usually compromises the other - the best tradeoff had to be found.

This means that closer the upper left corner the approach's corresponding diamond is, better can it be considered, while on the other hand, the ones close to bottom right corner are the worst performing ones.

While filtering the image by itself provides almost intact structural similarity, it removes hardly any speckle noise being located on the upper right corner of the plot.

The "default" approach mentioned in the literature, the point-wise average, keeps the structures well enough but is not that good in speckle mitigation.

In the majority of the cases, the single image approaches perform better than double image approaches. Or from the other way around, introducing the second image in the process, is not always a better idea. Bilateral filter, simple 2D median filter and filtering image by the average of itself can be considered the best performing ones among single image techniques.

On the other hand, the point-wise average filtered under the guidance of the first frame can be considered best among the double image approaches. While reducing the speckle noise through blurring the low variance regions of the image and through applying the "default" and most commonly cited way of pixel-wise averaging, it manages to ignore the artifacts introduced by the pixel-wise averaging, prioritizes the structures from the original image and keeps structural similarity high enough.

3D median filtering can look also quite tempting due to its low speckle level, but its structural similarity is not really satisfactory.

The two blue labeled red diamonds in the very bottom left corner make us conclude that it is a bad idea to use different guidance and input images, when the alignment is not perfect.

Another interesting point can be observed on the plot: we have five techniques that use two images and five pairs of the red diamonds with blue and red labels, which signify well aligned and 3-pixel misaligned versions. It is quite apparent that in all five cases the shift from well-aligned to 3-pixel misaligned version is towards bottom left corner, meaning that misalignment degrades the structural similarity (that was quite logical), but at the same time slight misalignment decreases the speckle noise level. The explanation for this phenomenon was provided while discussing the *Figure 4.1.10*.

Even though low speckle noise is tempting, I think that we have to prioritize higher structural similarity of the well aligned images, in order to avoid the visual artifacts that degrade the image even more than slightly higher level of the speckle noise.

4.2. Full Image Approach

In some cases, it was possible to mosaic directly the large images without splitting into the patches. Therefore, manually imposed misalignment lost its importance and only automatically aligned images were considered.

10 images with varying dimensions were examined (starting from 256-by-256 pixels up-to 1000-by-1000 pixels). The images were selected from the same or maximally similar videos, as the ones for splitting them into 64-by-64 pixel sub-regions, in order to minimize the impact of the content difference, even though we demonstrated in the previous subchapter that the content does not have dramatic impact on the overall performance. But it was still important to keep the absolute values of speckle index and structural similarity close to each other.

The results are illustrated on the figure 4.2.1. and 4.2.2. They are similar to *Figure 4.1.12* and *Figure 4.1.13*.

The general trend is the same as in case of splitting the images into the patches. Guided image filtering for a single image filtered by itself is still performing worse in terms of speckle index and best in terms of structural similarity.

It is interesting that the speckle index values are relatively low for the originals, as well as for all techniques. This can be explained by the fact that larger homogenous black background areas are present on the larger images than it was in case of small patches that were selected to be covering the cellular structures. One of the illustrations of this kind of image can be seen on *Figure 4.2.3*. This can be considered a drawback of speckle index as a metric and the solutions can be explored in the future work.

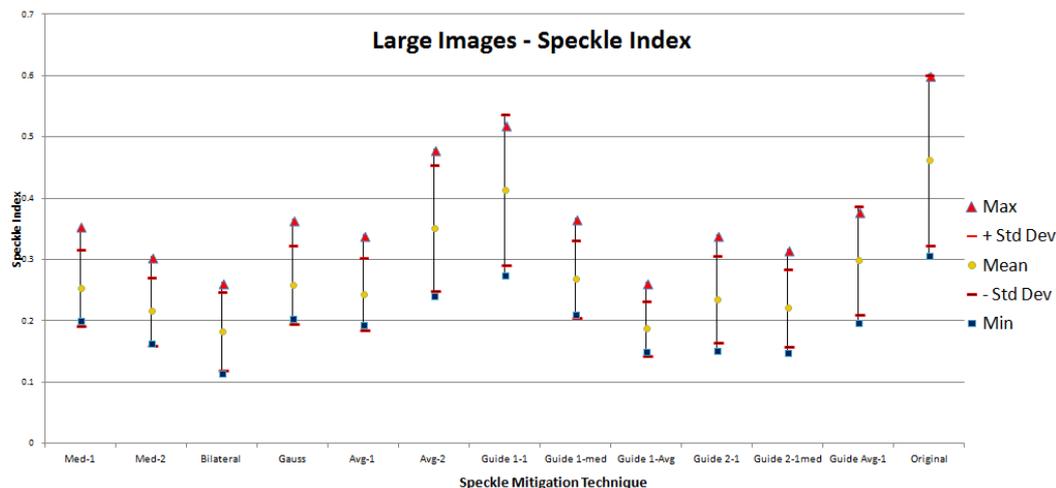


Figure 4.2.1.- maximum, mean, and minimum speckle indices, as well as the standard deviation for 10 large images.

Again the same trend is followed for the structural similarity as well, filtering the images with guidance of another image still leading to high standard deviations in structural similarity. The interesting point here is that the mean values for structural similarity are around 0.8, while minimal values are well above 0.6, except for the two just mentioned exceptions of the guided image filtering.

This structural similarity of the output images in some double-image based approaches, like point-wise average, that is one of the best performing in this case, means that the registration was in total satisfactory due to the challenges mentioned in subchapter 3.2.

This is the promising outcome that means that after further refinement of the automatic registration, the need of manual alignment will be minimized and reaching fully automatic robust model is a very realistic aim.

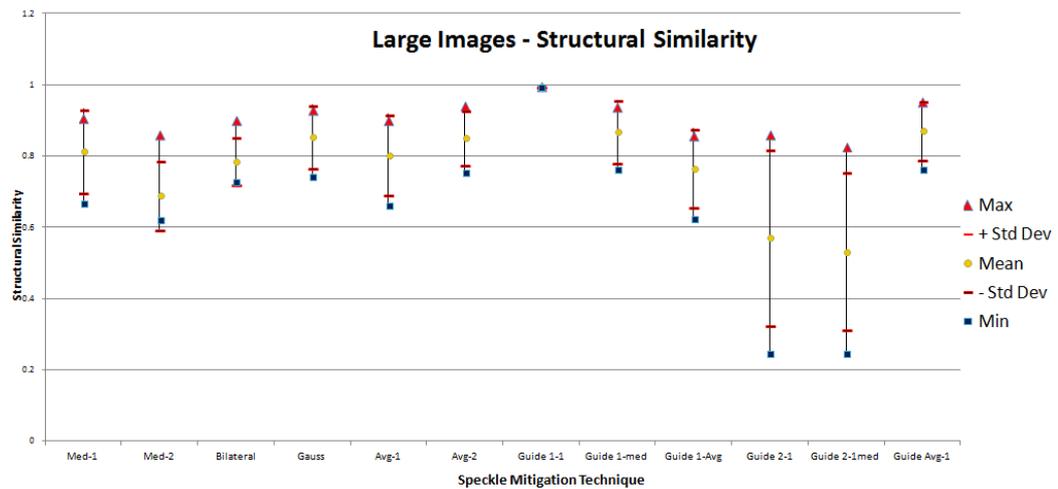


Figure 4.2.2.- maximum, mean, and minimum speckle indices, as well as the standard deviation for 10 large images.

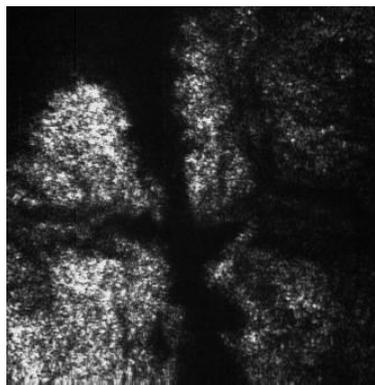


Figure 4.2.3. – Large homogenous black backgrounds lead to lower speckle index values.

4.3. Psychometric Scaling Experiments

The psychometric scaling experiments were held to compare best performing bilateral filtering and guided image filtering approaches, as well as conventional point-wise average and original image according to their appearance to the human observer. The details of the psychometric scaling experiments are explained in Chapter 2.

The results of the psychometric scaling experiments are summarized in the figures below. The vertical axis is the value of z-score, while the horizontal axis represents the Original and three different versions of it after speckle mitigation processing: Bilateral – bilateral filter applied on the first frame; Guided Filter – point-wise average of two consecutive frames filtered guided by the first frame; Pointwise Avg – pointwise average of the two consecutive frames.

For deeper insight into the data, each pair of the consecutive frames is represented separately. The legend below the plot lists 10 pairs processed as reference: Small_1 through Small_5 signify 64-by-64 pixel patches, while Large_1 through Large_5 signify larger, 512-by-512 pixel images. Each of them has its corresponding symbol (e.g. blue circle, blue diamond) signifying the mean z-score on the plot. The upper and lower limits of the error bar define the range of 95% confidence interval. If the name of the pair of the images is in bold on the legend, its z-scores are present on the plot, if not – then they are not.

It is very important to note that all pairs are assumed with best-possible alignment. Misalignment cases were not included in the psychometric experiments for avoiding extra-long experimental sessions. It can be considered and studied in future works.

Another important thing to note is that if the confidence intervals are overlapping, we cannot claim with 95% of the confidence that one approach or version is better than the other.

Let's refer to the Figures 4.3.1. through 4.3.5. below.

All but the third of the 64-by-64 patches demonstrate the similar trend: Bilateral filter has the least mean z-score, while the original has slightly higher mean z-score and guided filter and point-wise average have highest values almost equal to each other.

In most of the cases, there is no or minor overlap between the confidence intervals of the bilateral filter on the one hand and guided filter and point-wise average on the other hand. This allows us conclude that the latter are perceived more balanced in terms of speckle level and visualization of the key structures than the bilateral filter. The same can be said about bilateral filter and the unprocessed version. But the original, guided filtered version and point-wise average overlap with their confidence intervals a lot and it is difficult to conclude, which one better. While guided filtered version and point-wise average have generally higher mean z-score, in case of the fifth

pair the original is best performing with 95% of confidence. There is another interesting exception, pair 3, where bilateral filter is best performing one.

In order to explain those kinds of behavior, let's refer to Figure 4.3.6. which illustrates all the patches used in the experiment. The labels are located directly on the figure below the images.

The third image original is special in its nature. From the visual judgment, it has very high speckle noise level that is substantiated with high speckle index value as well. The structures of this patch, in contrast to other patches, is not contiguous and is segmented into several parts. Bilateral filter that demonstrated highest blurring capability, removes the extreme level of noise, while blurring small bright parts as well, leaving just larger segments of cellular structures - making the structures stand out of the background. That was not that apparent in other cases, because the structures were visible well enough from the very original.

It is even more challenging to explain the performance of the 5th pair, where original, thus, no processing at all, is considered the best. While the lack of speckle mitigation algorithm makes it the patch with highest speckle noise, it is a fact that the structures are what make it more preferable to the observers. The structures alongside the speckle are present on dark black background. While some of the structures are heavily segmented, speckle mitigation algorithms do not consider them as edges and blur them. This leads to the fact that the observer prefers the original, where the edges look apparently sharper, and speckle noise level is however high, but not disturbing to them, as the results show.

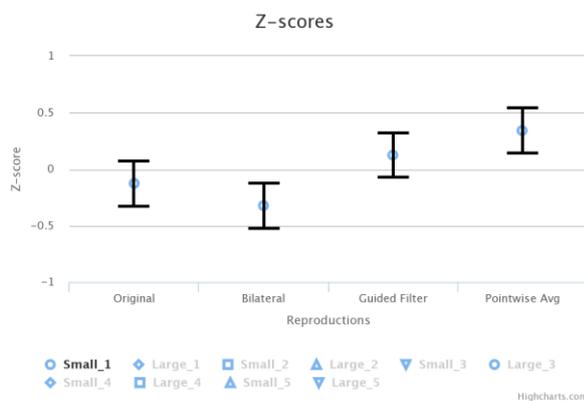


Figure 4.3.1. – z-scores for 1st pair of 64-by-64 pixel patches

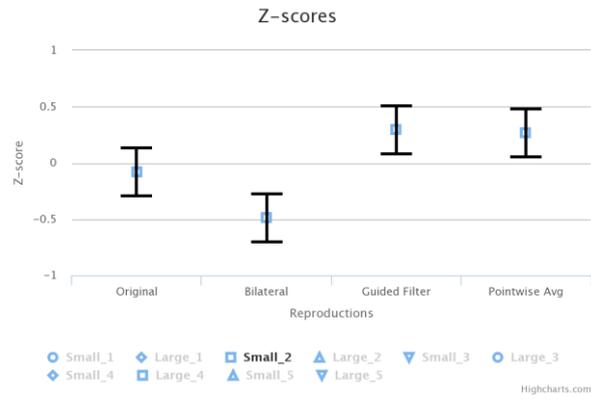


Figure 4.3.2. – z-scores for 2nd pair of 64-by-64 pixel patches

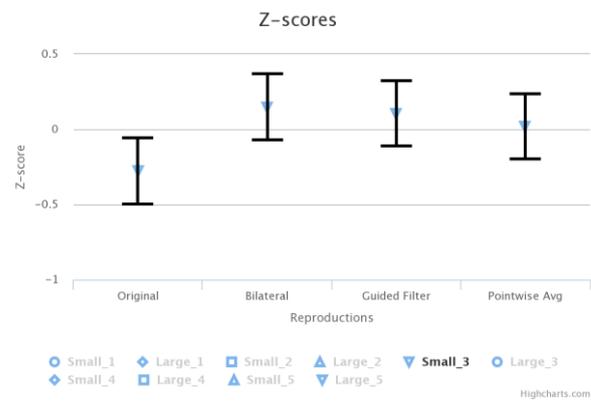


Figure 4.3.3. – z-scores for 3rd pair of 64-by-64 pixel patches

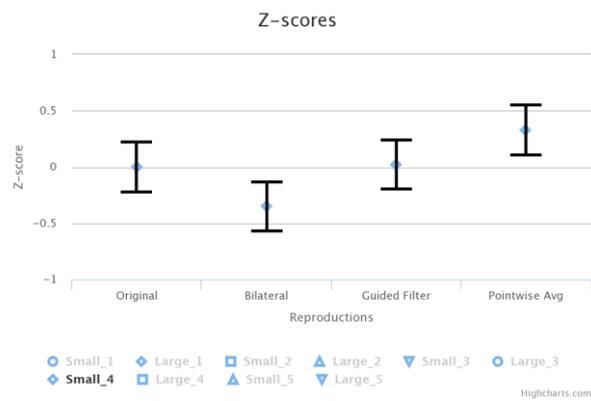


Figure 4.3.4. – z-scores for 4th pair of 64-by-64 pixel patches



Figure 4.3.5. – z-scores for 5th pair of 64-by-64 pixel patches

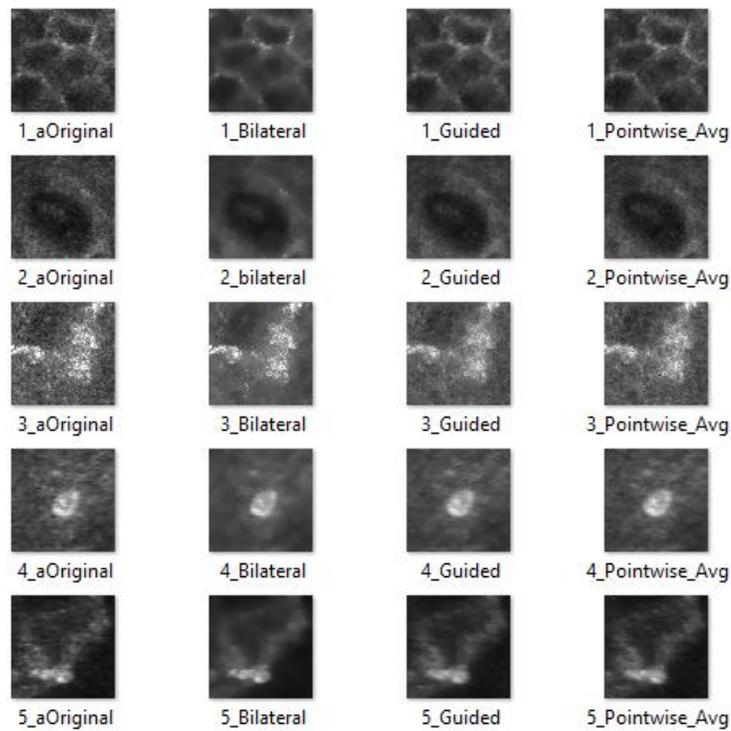


Figure 4.3.6. 64-by-64 patches used in the psychometric scaling experiments

The larger images processed at once were studied the same way. The original frames (1st one) are illustrated on *Figure 4.3.7.*, while the results are illustrated on the figures 4.3.8. through 4.3.12.

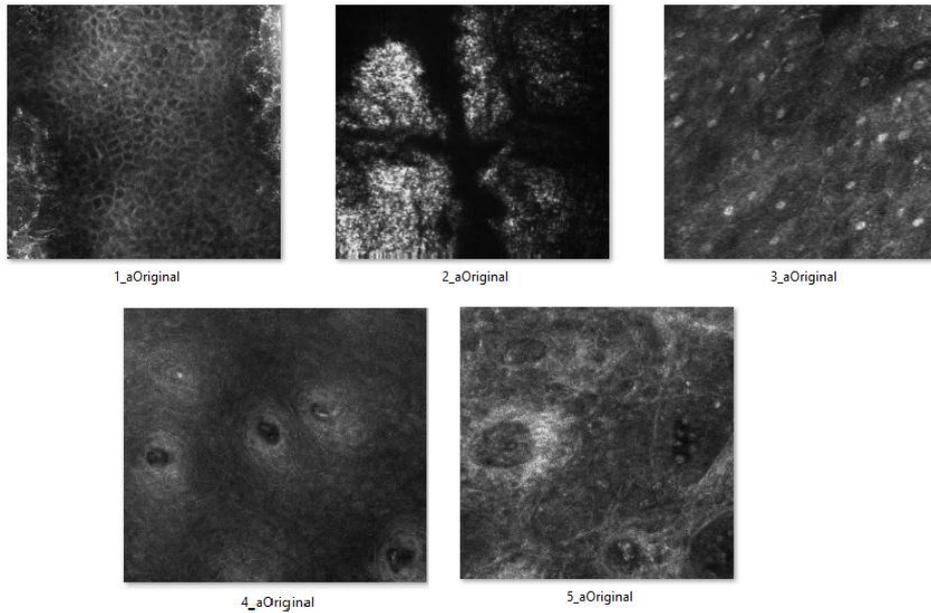


Figure 4.3.7. – the large frames used for psychometric scaling experiments

Those results are more consistent, bilateral filter being the worst performing one, without no, or insignificant overlaps with the confidence intervals of the rest of the approaches. On the one hand, it is a bit surprising, because bilateral filter uses single frame and it should be completely free from registration artifacts. On the other hand, its parameters discussed in Chapter 3, seemingly make it the most blurring approach among the three in question. At large scale, the observers feel the blur and miss the fine details.

It is difficult to conclude with confidence which one is considered better among the rest, due to the overlapping 95% confidence intervals. The only exception is Large_3 image on *Figure 4.3.10*.

I think the reason for this was the fact that registration of those pair of consecutive frames was a bit challenging and some minor artifacts present in point-wise average (*Figure 4.3.13*.) made the observers opt for the original. The similar could be the reason for smaller overlap for Large_4 image (*Figure 4.3.11*.) Its point-wise average is illustrated on *Figure 4.3.14*.

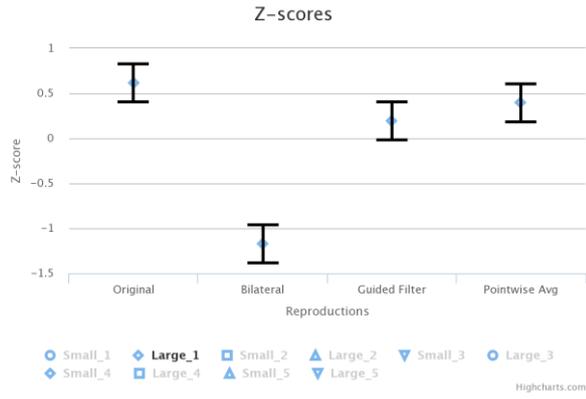


Figure 4.3.8. – z-scores for 1st pair of 512-by-512 images

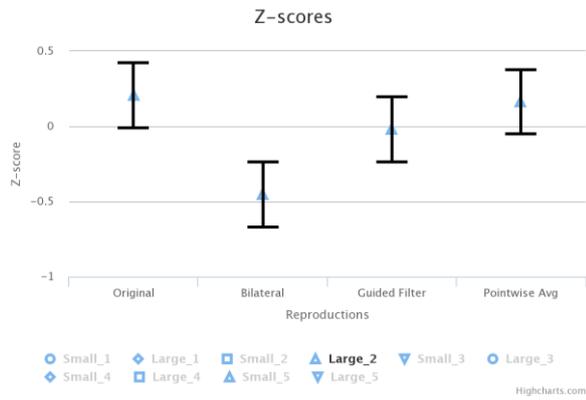


Figure 4.3.9. – z-scores for 2nd pair of 512-by-512 images

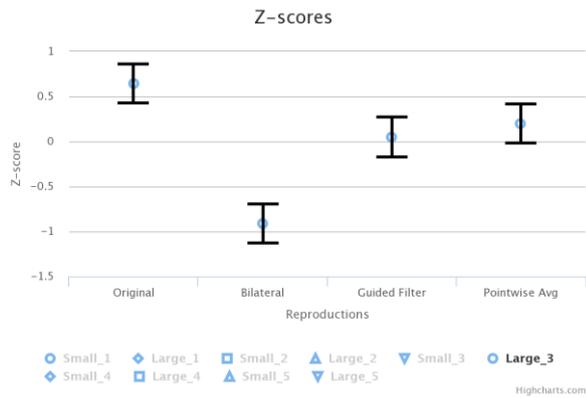


Figure 4.3.10. – z-scores for 3rd pair of 512-by-512 images

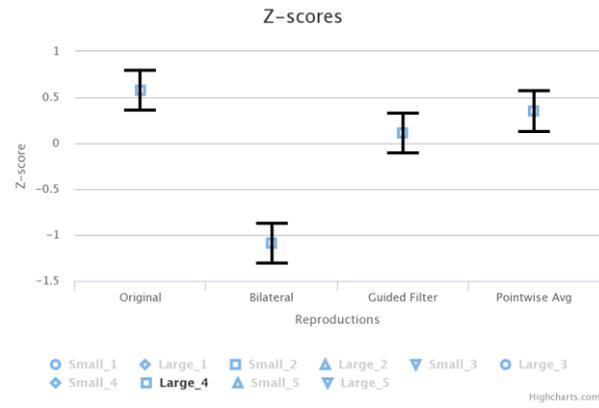


Figure 4.3.11. – z-scores for 4th pair of 512-by-512 images

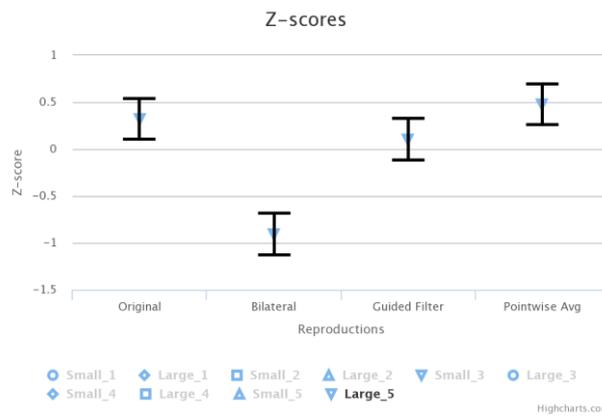


Figure 4.3.12. – z-scores for 5th pair of 512-by-512 images

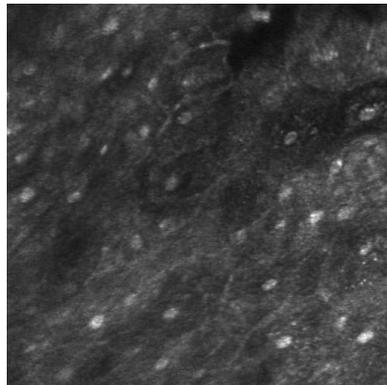


Figure 4.3.13. – minor artifacts present in point-wise average could be the explanation why the observers preferred the unprocessed version

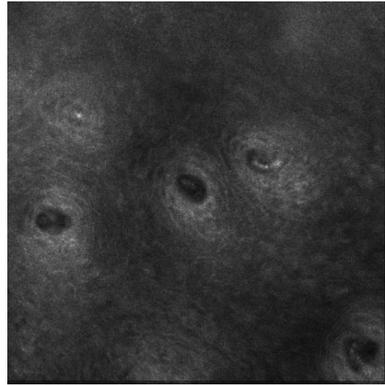


Figure 4.3.14. – minor artifacts present in point-wise average could be the explanation why the observers preferred the unprocessed version

Another interesting figure is Figure 4.3.15., illustrating the confidence intervals for all small patches except for the outlier Small_3. If this outlier is absent, we clearly see that the original, guided filter and point-wise average outperform the bilateral filter.

The same thing can be concluded without any doubt about large images, where the confidence intervals for all large images are present and none of the z-score 95% confidence intervals of bilateral filtered version overlap with that of any other “reproduction”. See the illustration of this fact on Figure 4.3.16.

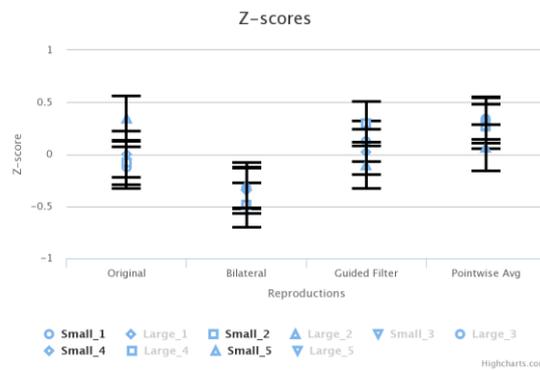


Figure 4.3.15. – all the 95% confidence intervals of bilateral filter are located below of the remaining three. The outlier image is not present on the plot.

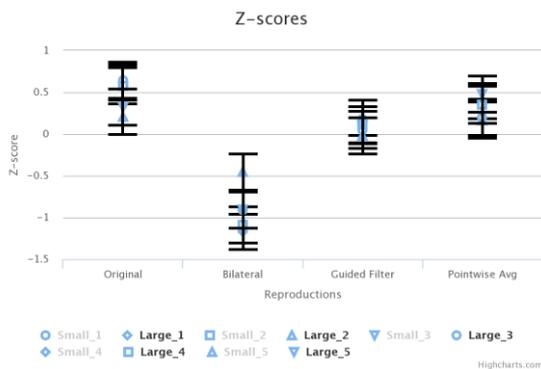


Figure 4.3.16. – none of the 95% confidence intervals of bilateral filter are overlapped with that of any other “reproduction”, making the bilateral filter worst performing one for large images.

If we put the confidence intervals of small and large images together (refer to Figure 4.3.17.), excluding the obvious outlier, we will still see the same trend that bilateral filter is worst performing one, while the rest are overlapping, making any claims and conclusions unreliable.

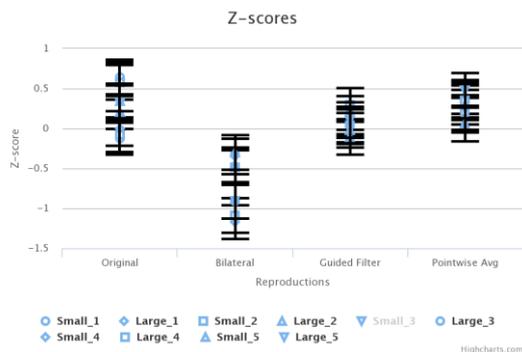


Figure 4.3.17. – 95% confidence intervals of small and large images put together

Chapter 5: Summary

To summarize, I have implemented and explored poorly studied way of speckle mitigation – usage of uncorrelated images, as I benefited from the overlapping regions and redundant data present in the videos acquired by the dual-axis confocal microscopes designed at the Molecular Biophotonics Laboratory of the University of Washington.

The key novelty of our approach was using videos, with automatic extraction of redundant data for this purpose and especially, techniques other than simple averaging after extraction of this redundant data.

Major problem, as already mentioned several times was the mosaicing of the images for extraction of the overlapping areas. The registration problem was quite demanding, as nearly perfect registration was needed for point-wise average. I tried to find the workaround for the lack of perfect alignment through applying several different techniques other than averaging, state-of-the-art guided image filtering being among the most interesting ones studied.

Limited dataset was also a big problem that makes generalization still a bit risky and leaves need for further credibility. In total 10 pairs of small patches and 10 pairs of large medical microscopy images were studied under twelve speckle mitigation techniques: seven of them were using only single image for speckle mitigation, while the remaining five used the data from two consecutive images. Commonly used application specific speckle index was used as an objective metric for quantifying the speckle noise, while image quality metric SSIM – Structural Similarity was used to measure how good the cellular structures were preserved. On the other hand, psychometric scaling experiments were held to further explore the performance of the most interesting (according to their objective metric performance) approaches.

The performance of the speckle mitigation techniques was quite consistent not only among different images, but between the two approaches – translation-assumed patches and automatically registered larger images as well.

While still larger datasets are needed for further generalization and more thorough case-by-case study is necessary to characterize the performance of a particular technique under give conditions, some conclusions still could be drawn:

The essential and most significant conclusion is the fact that the quality of registration is dramatic influence on the outcome for the techniques that use several consecutive frames. The recommendation would be the following: if no satisfactory alignment is possible, introducing the second image will just lead to further problems and therefore, it is better to use single-image-based approach.

Among single image based approaches, bilateral filtering and filtering the image under guidance of its own average performed the best. While conventional 2D median filtering can be also mentioned for its simplicity, flexibility and computational efficiency. On the other hand, filtering the image guided by itself should be avoided. Although it provides almost perfect preservation of the structures, its performance in terms of speckle noise mitigation is pretty poor.

When misalignment is observed, it is better to avoid using different images as input and guide in guided image filtering. Otherwise, you will end up with substantial artifacts, much blurriness and low structural similarity.

When the alignment is good enough, two- (or more in the future works) image approaches can be used. Pixel-wise average that is most commonly mentioned in the literature, can be used, but the registration should be really nearly-perfect, in order to avoid very apparent, image spoiling artifacts. Another option that works even in case of slight misalignment is filtering the pixel-wise average guided by the first frame of the sequence.

But the question, whether it is worth introducing the second channel even in case of satisfactory alignment remains open and needs to be applied to larger, more diverse databases to be answered decisively.

Even psychometric scaling experiments failed to answer this question decisively. Wherever minor artifacts due to misalignment were present, unprocessed original image was considered more preferable than double-image approaches by the observers.

According to psychometric scaling experiments, the bilateral filter was apparently least preferable among all small and large images, except for one outlier that had very high speckle noise level and specific content composition. We have seen other cases, where content seemingly played the role in determination of observer preferences. The psychometric scaling experiments increased interest in content dependence of the approach performances. Furthermore, if the selected parameters for bilateral filter were considered optimal for objective metrics, it turned out being too blurring for human observers. As it was the only single-image speckle mitigation technique studied within those experiments, future works should definitely expand to other same type of approaches. Considering the given setup, the recommendation would be to avoid the bilateral filter, unless the noise level is too high – masking the key structures. Another interesting finding was that if the misalignment is an issue, not only single-channel approaches outperform multi-channel ones, but in some cases, even no processing at all is considered preferable.

And last, but not least, it should be considered that those solutions could be intended for real-time applications, where except for structural similarity and noise, new constraint - computational efficiency is introduced. In such cases, slightly worse in

terms of speckle mitigation, but way faster approaches could be prioritized and simple and classical approach of 2D median filtering can be a nice example of this. It could be studied in psychometric scaling experiments as well.

Chapter 6: Further Work

Still large amount of work has to be done to reach the ultimate goal of creating fully automatized robust framework of speckle noise mitigation. Therefore, several follow-up research projects can be built upon this one.

First of all, the essential point of the approach and the most significant challenge was the automatic registration of the video sequence images. At the last stage of the project, the partners from the Stanford University provided the works by Loewke et al. [1][2] together with the software and its source code. It tries to solve in-vivo real-time image mosaicing problem for handheld dual-axes confocal microscopes and tries to tackle the problems that we faced during this project, e.g. tissue deformation. Unfortunately, no time was left to examine this tool into details within this project, but it could be a great starting point for the follow-up work. And generally, ways of automatic image mosaicing should be defined as one of the most significant challenges to be solved for successful implementation of this approach.

The second point of further work can be more refined ways of evaluation, as speckle index has demonstrated some shortcomings while applying for large microscopy images. Various ways of evaluation are described in the work of Suri et al. [3], classification tasks using kNN classifier being one of the most promising among them, if the dataset is built with symptomatic and asymptomatic images. This can be a good starting point in this direction.

And generally speaking, building and processing the larger datasets are essential to the success of the future work and its generalization possibilities, as limitedness in the proper datasets was one of the biggest challenges faced during this project.

Additional point worth mentioning about evaluation techniques can be applying machine learning approach to identify the optimal tradeoff between speckle noise removal and keeping the cellular structures, as it was challenging without medical expertise to identify the optimum.

Apart from this, it is important to note that manually defined parameters have enormous impact on the performance of some of the techniques. While I tried to find the optimal parameters empirically through my own subjective judgment, this question still can be considered open for further improvements and generalization.

Furthermore, I think that it is worth paying extra attention to the phenomenon, when misalignment decreases the speckle noise. This can lead to interesting findings in terms of speckle mitigation and in terms of finding the optimal degree of alignment, and generally, in terms of the nature of the algorithm applied.

Larger number of conventional approaches can be compared against ours, as well as larger number of techniques can be applied to extracted redundant data of the

overlapping regions, while commonly cited pixel-wise average needs perfect alignment, more sophisticated and refined techniques can be more robust and work in more limited conditions as well. Wavelet-based approach is among the interesting conventional techniques that our approach is worth comparing against.

Besides, there is one more interesting point worth mentioning for the follow-up: the whole project carried out within the framework of the master's thesis was focused on processing just two consecutive frames. While introduction of the second frame led me to lots of complications and challenges, the limitedness in time and resources did not allow me to go further within the scope of this project. But in fact, the same part of the tissue is present in more than two consecutive frames and even larger amount of the redundant data of the overlapping regions can be extracted. Therefore, one of the interesting directions for the future work can be usage of more than two consecutive frames, like three, four or more frames. This will indeed lead to further challenges, as the registration of 3- or more-layer data will be even more difficult, but pixelwise averaging of more uncorrelated data could be the way to further improvement.

Apart from this, psychometric scaling experiments demonstrated need for further directions. First of all, the content dependence of the approaches and the impact of the speckle amount in the original image need further study and understanding. Secondly, psychometric scaling experiments should be extended to misalignment cases and simple single-channel approaches, like 2D median filtering, to understand the need of introduction of multiple frames in speckle mitigation process. And most importantly, in order to gain extra credibility for psychometric experiments, especially, before actual implementation in real-life medical solutions, the experiments should be held on specific and calibrated medical devices under controlled conditions and the observers should be the people with the expertise in the medical field – they are the people final product is intended for and the average observers were used due to limitedness of our resources and opportunities, and just for general insight.

And finally, the collaboration should work both ways: while post-processing is one of the ways of solving the problems with speckle noise, the thorough study of issue can help us come up with some hardware improvement recommendations for the designers of the microscope, and on the other hand, the nature of the videos can be a good reason to create a guide of the best practices for the operator of the microscope, as high quality (e.g. smoothly moving) videos can simplify speckle mitigation task dramatically.

6.1. Chapter Bibliography

[1] Loewke, K., Camarillo, D., Piyawattanametha, W., Breeden, D. and Salisbury, K., 2008, February. Real-time image mosaicing with a hand-held dual-axes confocal microscope. In Proc. SPIE (Vol. 6851, pp. 68510F-68510F).

[2] Loewke, K.E., Camarillo, D.B., Piyawattanametha, W., Mandella, M.J., Contag, C.H., Thrun, S. and Salisbury, J.K., 2011. In vivo micro-image mosaicing. *IEEE Transactions on Biomedical Engineering*, 58(1), pp.159-171.

[3] Suri, JS, Chang, R, & Kathuria, C (eds) 2008, *Advances in Diagnostic and Therapeutic Ultrasound Imaging*, Artech House, Norwood. pp.37-63 Available from: ProQuest Ebook Central. [8 August 2017].

Appendix 1

Algorithm 1. Guided Filter.

Input: filtering input image p , guidance image I , radius r , regularization ϵ

Output: filtering output q .

```

1:  $\text{mean}_I = f_{\text{mean}}(I)$ 
    $\text{mean}_p = f_{\text{mean}}(p)$ 
    $\text{corr}_I = f_{\text{mean}}(I * I)$ 
    $\text{corr}_{Ip} = f_{\text{mean}}(I * p)$ 
2:  $\text{var}_I = \text{corr}_I - \text{mean}_I * \text{mean}_I$ 
    $\text{cov}_{Ip} = \text{corr}_{Ip} - \text{mean}_I * \text{mean}_p$ 
3:  $a = \text{cov}_{Ip} / (\text{var}_I + \epsilon)$ 
    $b = \text{mean}_p - a * \text{mean}_I$ 
4:  $\text{mean}_a = f_{\text{mean}}(a)$ 
    $\text{mean}_b = f_{\text{mean}}(b)$ 
5:  $q = \text{mean}_a * I + \text{mean}_b$ 
/*  $f_{\text{mean}}$  is a mean filter with a wide variety of  $O(N)$  time
methods. */

```

Algorithm taken from:

[1] He, K., Sun, J. and Tang, X., 2013. Guided image filtering. *IEEE transactions on pattern analysis and machine intelligence*, 35(6), pp.1397-1409.