Signal and Image Processing Techniques for Image-Based Photometry With Application to Diabetes Care

Vom Fachbereich 18
Elektrotechnik und Informationstechnik
der Technischen Universität Darmstadt
zur Erlangung der Würde eines
Doktor-Ingenieurs (Dr.-Ing.)
genehmigte Dissertation

von
M.Sc. Nevine Demitri
geboren am 6. August 1987 in Kairo, Ägypten

Referent:	Prof. Dr.-Ing. Abdelhak M. Zoubir
Korreferent:	Prof. Dr. Hamid Krim
Tag der Einreichung:	26.04.2016
Tag der mündlichen Prüfung:	14.06.2016

D 17
Darmstädter Dissertation 2016

'Woe the seekers of knowledge when they become satisfied with their achievements'
Taha Hussein

To Hany, Lilian, Youssef and Maggar

Acknowledgments

I would like to express sincere gratitude to my supervisor, Prof. Abdelhak M. Zoubir for giving me the opportunity to pursue my PhD at his institute. His excellent lectures initially sparked my interest for signal processing during my undergrad studies. Since then he has been my mentor, always offering his advice and support, both professionally and personally. Furthermore, I would sincerely like to thank my co-supervisor Prof. Hamid Krim for hosting me at the Vision, Information and Statistical Signal Theories and Applications Group in North Carolina State University (NCSU), Raleigh. I benefited greatly from my three-months stay at his group and gained new inspirations. Moreover, my gratitude goes to my examiners Prof. Klaus Hofmann, Prof. Heinz Köppl and Prof. Andy Schürr.

It has been a pleasure being part of the Signal Processing Group and I am thankful to all former and current members of the group for the great atmosphere and the countless feedback all through. Specifically, I would like to thank Stefan Leier for his sincere advice, Sara Al-Sayed for her friendship, Michael Muma for his feedback, as well as Jürgen Hahn and Di Jin for the long nights at the office and their moral support. I enjoyed very much sharing conferences with Michael Fauss, Gökhan Gül, Christian Weiss, Ivana Perna and Wassim Suleiman. Furthermore, it was a pleasure working with Mark Balthasar, Sahar AlKhawatmi, Lala Khadidja Hamaidi, Michael Leigsnering, Raquel Fandos, Ann-Kathrin Seifert, Feng Yin, Philipp Heidenreich, Freweyni Teklehaymanot, Ahmed Moustafa, Fiky Suratman, Patricia Binder, Simon Rosenkranz, Zhihua Lu, Gebremichael Teame. Among all students I supervised, I specially enjoyed working with Karim Wahby. I am sincerely grateful to Renate Koschella for her constant support and friendship, as well as Hauke Fath for his technical support. Furthermore, I extend my thanks to my colleagues at Roche Diagnostics GmbH, Mannheim for the valuable practical insight, specially Bernd Limburg and Klaus Wettengel. Thank you to my colleagues at the equal opportunities office.

Great thanks goes to my family in Egypt for encouraging me, each in their own way. My gratitude also goes to my friends in Darmstadt, who have been more like family. Janine for being a sister from day one; Monica for the inspiration to do things differently; Moura, for the sincere confidence and for all the laughs; Sara Hallouda, for the positivity and optimism; Amr Rizk, for the sincere advice. I would also like to express my thanks to my friends in Egypt for their support from the distance.

My deepest gratitude goes to my family. Hany and Lilian, you have provided

me with support and guidance in every possible way, as well as an invaluable friendship. Most of all you have motivated me to pursue my dreams, even if it meant often sacrificing your own comfort. I do not believe there is a greater form of love ! Thank you Youssef for constantly inspiring me, as well as for the telepathy and all the fun. Miggi, thank you for the happiness and love. And for bringing back my old self.

Darmstadt, 30.06.2016

Nevine Demitri

Kurzfassung

Diese Dissertation befasst sich mit dem Problem der bildbasierten, photometrischen Blutzuckermessung in tragbaren Geräten, die eine Blutmenge im Nanoliterbereich erfordern, d.h. eine um mehrere Größenordnungen kleinere Menge als in herkömmlichen Blutzuckermessgeräten. Die chemische Reaktion zwischen der Blutprobe und dem Reagens wird in dem betrachteten Aufbau mithilfe einer Kamera beobachtet. Allerdings kann das präsentierte Konzept für den Einsatz in beliebigen bildbasierten, photometrischen Messkonzepten basierend auf moderner Bildsensorik verallgemeinert werden.

In dieser Arbeit wird ein Konzept zur Ermittlung der Glukosekonzentration anhand von Rohbildern der Kamera entwickelt. Zunächst wird ein Vorverarbeitungsschema präsentiert, das die Rohbilder für die weitere Verarbeitung vorbereitet. Außerdem wird ein Detektionsalgorithmus zum Erkennen des Reaktionsbeginn entwickelt, der unnötige Berechnungen während der konstanten Phase der chemischen Reaktion vermeidet. Des Weiteren werden Verfahren zur Texturanalyse vorgestellt zur Erfassung fehlerhafter Glukosemessungen.

Bildsegmentierung ist ein wichtiger Beitrag dieser Arbeit. Die Segmentierung von Glukosebildern wird als Clustering-Problem interpretiert und mit zwei Ansätzen gelöst. Im ersten Ansatz werden die Mean-Shift und die Medoid-Shift-Algorithmen auf das Glukosesegmentierungsproblem angepasst. Darüber hinaus werden zwei neue Varianten entwickelt, die auf Robustheit und Sparsity beruhen. Die erste Variante ist in der Lage mit heavy-tailed-verteilten Bildern umzugehen. Die zweite Variante reduziert den Rechenaufwand des Mean-Shift Algorithmus ohne die Genauigkeit der Schätzwerte zu beeinträchtigen, indem nur eine Teilmenge der Datenvektoren verarbeitet wird. Darüber hinaus werden robuste und sparse Variationen des Medoid-Shift Algorithmus hergeleitet und die Konvergenzbeweise für alle entwickelten Algorithmen präsentiert. Der zweite Ansatz zur Bildsegmentierung basiert auf den Model der Vereinigung von Teilmengen (Union of Subspaces). Er passt die Methoden Sparse Subspace Clustering und Niedrigrangdarstellung an das Problem der Segmentierung der Glukosebilder an. Zwei neue Formulierungen von Sparse Subspace Clustering werden hergeleitet, die genauere Segmentierungsergebnisse als den Stand der Technik liefern. Zudem wird eine sequentielle Version von Subspace Clustering hergleitet, die sowohl genauere als auch schnellere Segmentierungsergebnisse bereitstellt.

Die Konvergenzdetektion der chemischen Reaktion stellt einen weiteren wichti-

gen Beitrag des präsentierten Konzepts dar. Hierfür wird ein nichtlineares Modell der chemischen Kinetik aufgestellt, das die lineare Beziehung zwischen der Reaktionsgeschwindigkeit und dem Konvergenzwert der chemischen Reaktion ausnutzt, um die Anwendung von Zustandsschätzung und Prediktionsmethoden zu erleichtern. Diese Methoden sind in der Lage mittels eines Extended-Kalman-Filter zum einen und eines regionsbasierten Partikelfilters zum anderen, die Genauigkeit der Schätzwerte zu erhöhen, sowie die Messzeit drastisch zu kürzen.

Eine umfangreiche Messreihe von Echtdaten wird verwendet, um die Gültigkeit der entwickelten Ansätze in Bezug auf die Segmentierungsergebnisse, die klinische Genauigkeit der geschätzten Glukosewerte und die benötigte Messzeit aufzuzeigen. Ferner wird eine Untersuchung zur Bestimmung einer unteren Grenze für das Blutprobenvolumen mittels Echtdaten durchgeführt.

Abstract

This PhD thesis addresses the problem of measuring blood glucose from a photometric measurement setup that requires blood samples in the nano litre-range, which is several orders of magnitude less than the state of the art. The chemical reaction between the blood sample and the reagent in this setup is observed by a camera over time. Notably, the presented framework can be generalised to any image-based photometric measurement scheme in the context of modern biosensors.

In this thesis a framework is developed to measure the glucose concentration from the raw images obtained by the camera. Initially, a pre-processing scheme is presented to enhance the raw images. Moreover, a reaction onset detection algorithm is developed. This eliminates unnecessary computation during the constant phase of the chemical reaction. To detect faulty glucose measurements, methods of texture analysis are identified and employed in a detection scheme.

Image segmentation constitutes an important contribution of this thesis. In this work, the segmentation of glucose images is interpreted as a clustering problem and is tackled using two approaches. In the first approach the mean-shift and the medoid-shift algorithms are adapted to the glucose segmentation problem. Moreover, two novel variations are derived: the robust mean-shift and the sparse mean-shift. The former is able to deal with heavy-tailed noise in the images. The latter reduces computation of the mean-shift algorithm, while not compromising in accuracy, by only processing a subset of the data vectors. The robust and sparse variations are applied to the medoid-shift. Convergence proofs for all developed algorithms are provided.
The second approach to tackle the image segmentation problem is based on the union of subspaces model. To this end, sparse subspace clustering and low-rank representation methods are adapted to the glucose segmentation problem. Two novel formulations of sparse subspace clustering are derived that outperform the state of the art in producing more accurate segmentations of the video data. Additionally, a sequential version of subspace clustering is derived that achieves more accurate and earlier segmentations results.

Convergence detection of the chemical reaction constitutes a further major contribution of the presented framework. For this purpose, a non-linear model of the chemical kinetics is established. Thereby, a linear relation between the reaction rate and the convergence value is exploited to facilitate the application of state estimation and prediction techniques. These succeed at increasing the accuracy of the converged

remission value and decreasing measurement time drastically. To this end, the first technique is based on the extended Kalman filter and the second technique is a region-based particle filter.

An extensive set of real glucose measurements is used to prove the validity of the developed approaches w.r.t segmentation results, clinical accuracy and measurement time. Furthermore, a study is performed using real data, determining a lower limit for the blood sample volume.

Contents

Chapter 1

Introduction

1.1 Motivation

Terms such as *Patient Empowerment* [AF05, AF10] or *The Healthcare Internet of Things* have been trending in recent years. These terms indicate an apparent paradigm shift in healthcare, such that the diagnosis and treatment is not only confined to the physicians clinic, but also constitutes an ongoing exchange of information between the patient and the physician. With the development of easy-to-use and accurate self-measurement devices, the patients are gaining an integral role in the success of their treatment, particularly in the case of chronic diseases. Smartphones are being employed as health sensing devices that can measure different analytes contained in fluids, e.g., blood or saliva, using internal and external sensors [FWO+11, TVS+15]. Additionally, new companies have been emerging that promise early cancer detection from a single blood drop [Tew15]. Others promise to use pinprick blood samples in hand-held devices to streamline laboratory blood tests [HRHW15]. Thereby, patients are obtaining higher responsibility in their treatment process.

The importance of regular self-monitoring has long been identified as a necessary component of successful treatment for diabetes patients. The first photometric hand-held blood glucose measurement device was developed around 40 years ago. It constitutes one of the oldest technologies in point-of-care testing and patient self-management [CF12]. At this time, the device required a blood drop of 20 µl to 30 µl (micro litre) to be extracted from the finger and placed on a chemical reagent that changes its colour according to the amount of glucose in the blood sample. The patient had to compare the change of colour on the test strip to a provided colour scale. Since that time, much research has been done to improve the usability of these devices. The development of small and accurate optical sensors has allowed automatic reading of the colour change by the device. This has led to an increase in measurement accuracy of the glucose concentration.

Today, devices on the market are either based on the described photometric approach, or an electrochemical approach. The required blood sample volume has decreased to a range of 0.3 µl to 3 µl. Although these devices have come a long way since their emergence, they leave much room for improvement in terms of usability

and accuracy. Therefore, research in blood glucose self-measurement devices remains an important topic of interest [JAT02].

1.2 Aims and Research Objectives

Recently, a photometric blood glucose measurement approach has been developed [AHKP11, PW13] that drastically reduces the blood sugar sample volume to a range of 10 nl to 100 nl (nano litre). It is based on capturing the colour change on the chemical test strip over time by a camera, obtaining a video of the chemical reaction. The decreased blood sample volume in this setup leads to a reduction of the pain sensation for the patient, while the use of the camera enhances accuracy compared to state-of-the-art devices that use a photodiode to measure the colour change. It is important to note that most methods developed for glucose measurement devices are proprietary. To the best of the author's knowledge very little published works exists in terms of signal and image processing techniques for image-based photometric measurement systems in the medical context.

This doctoral thesis aims at developing methods and techniques to enhance the usability of photometric blood sugar measurement devices that require blood samples in the nl-range. This embodies achieving high accuracy of measurements, in spite of the very small blood sample, as well as reducing measurement time. For this purpose three objectives are sought in this thesis:

1. The development and presentation of a full signal and image processing framework: The developed framework takes as an input the raw glucose images obtained from the described novel measurement setup and outputs an accurate estimate of the glucose concentration. Hereby, the needed measurement time is minimised. It is important to note, that the presented framework is not intended to be confined, solely, to the measurement of glucose in blood, but can be used to measure other analytes in a fluid using a photometric measurement approach.

2. The accurate segmentation of the glucose images: This can be a challenging problem due to the small volume of the blood drop, which leads to a disconnected region of interest. Particularly in low glucose scenarios the colour change is weak and barely identifiable, further complicating the segmentation. Additionally, the reaction area is often occluded by artefacts that may disrupt the reaction region.

3. The reduction of the measurement time: Typically, glucose measurements using the described setup may take up to 20 s to saturate. Reducing the measurement time is an important step towards increasing usability and motivating regular self-monitoring.

1.3 Summary of the Contributions

The original contributions of this doctoral thesis are:

The development and presentation of a full framework as described by the first objective in Sec. 1.2. This framework measures the glucose concentration from raw glucose images obtained by the photometric measurement setup, requiring blood samples in the nl-range. The original contributions in the context of the framework are constituted by:

- the development of a pre-processing scheme to enhance the glucose images and facilitate further processing [DZ13].
- the development of a reaction onset detection algorithm based on the Neyman-Pearson hypothesis test. This approach allows saving unnecessary computation during the constant phase of the chemical reaction [DZ16].
- the identification of faulty glucose measurements using a texture analysis approach that is based on grey-level co-occurrence matrices [DZ14a].

The contributions pertaining to the second objective, namely accurate image segmentation constitute a major part of this thesis. The developed segmentation techniques are part of the framework described previously. They are treated separately due to their importance. The original contributions made for segmentation of the glucose images are given by:

- the application of the mean-shift algorithm and the medoid-shift algorithm for segmentation of the glucose images and subsequent estimation of the colour change representing the glucose concentration [DZ13]. To this end, the kernel and bandwidth selection problems are discussed in detail and adapted to the glucose estimation problem.

- the development of a novel robust variation of the mean-shift algorithm and the medoid-shift algorithm that are able to deal with contaminated data scenarios and increase the estimation accuracy [DZ14b]. The derivation of the robust variations is based on an alternative interpretation of the mean-shift algorithm in terms of the reproducing kernel Hilbert space (RKHS). Additionally, the convergence proofs of these algorithms are derived [DZ16].

- the development of a novel sparse variation of the mean-shift algorithm and the medoid-shift algorithm that achieves comparable accuracy using only a fraction of the data [DZ15]. As for the robust variation of the mean-shift, the sparse variation is based on an interpretation in terms of the RKHS. The convergence proofs for these algorithms are derived [DZ16].

- the application of subspace clustering methods that take into account not only the colour change of each frame but also the temporal evolution of the colour change across the frames to identify the reaction region [DKZ16].

- the derivation of two novel formulations of sparse subspace clustering that are tailored to the segmentation of video data by enforcing the similarity of neighbouring pixels and favouring the choice of only few cluster centres to represent the remaining data vectors [DKZ16]. Two different noise models are analysed in this context.

- the derivation of a novel sequential scheme for subspace clustering that estimates the error contained in previous frames and uses it to enhance segmentation of consequent frames [WDZ16].

Finally, the schemes developed to detect the convergence of the chemical reaction at an early stage and minimise measurement time are presented in the sequel. They represent the last block in the previously described framework. The contributions relating to convergence detection are given by:

- the extension of a non-linear model of the chemical kinetics by establishing a linear correlation between the reaction rate and the saturation value [DZ16]. This facilitates further use in a state estimation and prediction framework.

- the application of an extended Kalman filter approach to predict the saturation value of the chemical reaction ahead of time, using the derived non-linear model of the chemical kinetics [DZ16].

- the derivation of a region-based particle filter approach that tracks the image statistics, hereby approximating the saturation value of the chemical reaction and reducing measurement time [SDZ16].

1.4 Publications

The period of doctoral candidacy has culminated in following publications:

Internationally Refereed Journal Articles

- Demitri, N. ; Krim, H. ; Zoubir, A. M. An Ordered Subspace Clustering Approach for Temporal Segmentation of Blood Glucose Measurements. *IEEE Transactions on Biomedical Engineering,* submitted.

- Demitri, N. ; Zoubir, A. M. Measuring Blood Glucose Concentrations in Photometric Glucometers Requiring Very Small Sample Volumes. *IEEE Transactions on Biomedical Engineering*, March 2016, early access.

Internationally Refereed Conference Papers

- Wahby, K. ; Demitri, N. ; Zoubir, A. M. Sequential Low Rank Representation For Blood Glucose Measurements.*Proceedings 24th European Signal Processing Conference(EUSIPCO).* Budapest, Hungary. August 2016, submitted.

- Seifert, A.-K. ; Demitri, N. ; Zoubir, A. M. Decreasing the Measurement Time of Blood Sugar Tests using Particle Filtering.*IEEE 41st International Conference on Acoustics, Speech, and Signal Processing (ICASSP).* Shanghai, China, March 2016.

- Demitri, N. ; Zoubir, A. M. Estimating Glucose Concentration Using A Sparse Scalable Mean-Shift Algorithm. *2015 IEEE 11th International Symposium on Biomedical Imaging (ISBI).* New York City, USA, April 2015.

- Demitri, N. ; Zoubir, A. M. Detection of Faulty Glucose Measurements Using Texture Analysis.*Proceedings 22nd European Signal Processing Conference(EUSIPCO).* Lisbon, Portugal, September 2014

- Demitri, N. ; Zoubir, A. M. A Robust Kernel Density Estimator Based Mean-Shift Algorithm. *IEEE 39th International Conference on Acoustics, Speech, and Signal Processing (ICASSP).* Florence, Italy, May 2014.

- Demitri, N. ; Zoubir, A. M. Mean-shift Based Algorithm for the Measurement of Blood Glucose in Hand-Held Device.*Proceedings 21st European Signal Processing Conference(EUSIPCO).* Marrakech, Morocco, September 2013

Workshop Contributions

- Demitri, N. ; Zoubir, A. M. Bildbasierte photometrische Blutzuckermessung in tragbaren Geräten. *Entwicklerforum Medizinelektronik: "Telemedizin, Smart Health und Ambient Medicine".* Munich, Germany, October 2015.

1.5 Thesis overview

Following this introduction, the dissertation is organised as follows:

In **Chapter 2** an overview is given on diabetes care and the importance of blood sugar self-monitoring for diabetes patients. The state of the art in blood glucose meters is reviewed and the novel photometric setup, which constitutes the basis of the work in this thesis is detailed. The real glucose images that are used throughout this thesis for validation are presented and glucose-specific validation methods are discussed. The real data sets were provided, kindly, by Roche Diagnostics GmbH, Mannheim.

In **Chapter 3** the developed framework for the estimation of the blood glucose concentration from the glucose images is introduced. This constitutes the development of a pre-processing scheme to simplify further processing. Additionally, a reaction onset detection method is developed to save unnecessary computation prior to the onset of the reaction. Furthermore, a faulty measurement detection algorithm is presented that is based on texture analysis. Finally, a model of the chemical kinetics is derived. Using this model, a prediction and state estimation-based convergence detection scheme is developed to increase the accuracy of the estimates and reduce measurement time. Extensive experiments using both simulated and real data sets validate the developed framework. Furthermore, a study is performed using real data sets, establishing a lower limit for the permissible blood volume of the blood drop

w.r.t accuracy.

In **Chapter 4** the application of the mean-shift and medoid-shift algorithms for segmentation of the glucose images is discussed. To this end, the problem of kernel and bandwidth selection is highlighted. Furthermore, robust and sparse mean-shift and medoid-shift variations are derived. They are based on expressing the kernel density estimator in terms of the reproducing kernel Hilbert space. Convergence proofs for all novel algorithms are derived. Simulation and experiment results support the validity of the presented approaches.

In **Chapter 5** subspace clustering methods are applied to the glucose video segmentation problem. To enhance the performance of these methods for video segmentation, two novel formulations of sparse subspace clustering are derived. To this end, two different noise models are compared. Additionally, a sequential approach for subspace clustering methods is developed. Real data experiments validate the performance of the developed approaches.

Finally, in **Chapter 6** conclusions are drawn and recommendations for future work are presented.

Chapter 2

Diabetes Mellitus and Photometric Glucose Measurement

This chapter provides an overview of diabetes care, specifically of hand-held glucose measurement devices and related concepts that affect the work presented in this thesis.

First, a brief overview of Diabetes Mellitus is given in Sec. 2.1. Then the attention is turned to the discussion of hand-held blood glucose measurement devices. For this purpose, a review of the state-of-the-art methods in blood glucose measurement is given in Sec. 2.2. The photometric measurement principle, in particular, is discussed in more detail in Sec. 2.3. It constitutes the basis of the framework proposed in this thesis. Finally, the novel glucometer, from which the real data is attained, is discussed in Sec. 2.4. Hereby, special attention is given to the kinetic behaviour of the underlying chemical reaction. Moreover, the real glucose data sets as well as the glucose-specific validation methods are presented.

2.1 Diabetes Mellitus

Diabetes Mellitus describes a group of metabolic diseases affecting 422 million people worldwide (data from 2016) [P$^+$10, WHO16]. It occurs when the pancreas cannot produce enough insulin or when the body cannot use the insulin it produces [WHO]. Insulin is a hormone released in the body that helps regulate the blood sugar levels, also termed blood glucose levels, acquired by food intake [JPD07, SS00]. The digestion of food produces so-called monosaccharides, such as glucose that enter the bloodstream. These are used in the cells to provide energy and growth. This process is enabled by the hormone insulin, and thus, when insulin is not produced or recognised by the cells, the blood contains a high concentration of unused glucose. Diabetes is, thereby, characterised by chronically high blood glucose levels.

Generally, diabetes is classified into three categories [Ame10]:

1. *Type 1 Diabetes*: accounts for only $5\% - 10\%$ of all diabetes cases and is often diagnosed in children or young adults. It occurs when the pancreas is unable

to produce insulin. Patients with this kind of diabetes require regular insulin intake using an insulin pump. To this end, the patients need to self-monitor their glucose levels to be able to inject themselves with an appropriate amount of insulin.

2. *Type 2 Diabetes*: accounts for $90\% - 95\%$ of all diabetes cases. While in Type 2 diabetes the pancreas produces insulin, the cells are inefficient at absorbing this insulin. This condition is termed *insulin resistance*. Leading a healthy lifestyle, as well as the intake of medication, can strongly slow down the progression of Type 2 diabetes. Patients diagnosed with this category of diabetes rely strongly on self-management and regular monitoring of their blood glucose levels.

3. *Gestational diabetes mellitus (GDM)*: is defined as any glucose intolerance that has its onset or is recognised during pregnancy. It occurs when the placenta supporting the baby produces hormones that block the action of the mother's insulin in her body. Almost 15% of pregnant women are affected by GDM [GDM]. The treatment of GDM relies strongly on the self-control of glucose levels and careful adjustment of lifestyle as well as, in some cases, the intake of insulin.

The healthy glucose range when fasting lies between 70 mg/dl to 130 mg/dl (milligram/decilitre) [P$^+$10]. Fasting *hyperglycaemia* describes a condition where the blood sugar levels are higher than 130 mg/dl. This condition can have long-term effects on the body, such as the failure of various organs e.g., the kidneys, the eyes, the heart and the blood vessels. *Hypoglycaemia*, in contrast, presents itself for fasting blood sugar levels under 70 mg/dl. It can occur when the patient injects too much insulin into the body, misses a meal, or exercises too much, leading to a decreased level of glucose in the blood. The risks of this condition are short-term and can lead to losing consciousness, coma or seizures in severe cases [JAT02]. According to the International Diabetes Foundation (IDF), diabetes is the eighth leading cause of death worldwide, accounting for 4.9 million deaths annually [IDF14]. It mainly prevails in low- and middle-income countries. Projections for 2035 indicate that the number of people diagnosed with diabetes will rise to 592 million people worldwide.

One indispensable component of diabetes management is self-management and regular self-control. Several studies have been able to statistically associate a significant delay of the onset or slowing down of the progression of complications through intensive treatment guided by frequent blood glucose self-monitoring [Dia93, GDG$^+$03]. The injection of insulin or the adjustment of food intake and exercise strongly rely on the results of self-measurement. For this purpose, invasive hand-held devices containing glucose biosensors are used by patients to regularly and reliably self-monitor their

glucose levels. The patients typically perform up to five measurements daily to keep track of their blood sugar levels.

Another very critical application of glucose meters is point-of-care (POC) testing [RRF12]. This denotes medical diagnostic testing that is performed at the patient's bedside and stands in contrast to medical laboratory testing. Many studies have shown that glucose POC testing significantly improves morbidity and mortality in critically ill patients by regularly monitoring the blood glucose levels of the patient. This is typically performed using glucometers intended for self-monitoring.

Current hand-held devices on the market are mainly invasive, as will be presented in the sequel. The measurement process requires the patient to extract a blood drop from the body, typically from the finger, the earlobe, or the forearm, using a lancet. This can be prohibitive and often leads to decreased self-control by the patients. It is of great importance to increase the usability of these devices, while maintaining high accuracy, as they are essential to successful diabetes treatment. High usability features include ensuring reduced pain when extracting the blood drop, as well as providing devices that perform fast measurements. Furthermore, it is required that the devices be small and easy to carry as well as exhibit low power consumption characteristics.

2.2 State of the Art in Glucometers

Traditionally, glucose self-monitoring devices are based on an invasive procedure that uses a photometric or an electrochemical approach to infer the glucose concentration from a blood sample [Lau15, CF12, JAT02, JYS13]. The photometric approach relies on a chemically induced colour change of the reaction area and is presented in more depth in Sec. 2.3. In the electrochemical approach the chemical reaction results in an electric current that can be measured and related to the underlying blood glucose concentration [TN09]. In this thesis, only the photometric approach will be considered as it is the basis of the proposed framework. Both electrochemical and photometric approaches are invasive, which implies the extraction of a blood drop. The devices on the market today require the volume of the blood drop to be between 0.3 µl to 3 µl [CF12].

The last years have seen a lot of research in the area of non-invasive blood glucose monitors to reduce the pain caused by invasive methods and motivate more

regular testing. These fall mainly in three categories: continuous glucose monitoring using implanted sensors, transdermal sensors that measure through the skin, and spectroscopic meters [JAT02]. The problem with implanted sensors is that up until today the sensors' performance decreases rapidly with increased exposure to the internal environment of the body. Other challenges using this technique exist include the corrosion of sensors inside the body, or the reaction with other substances in the body leading to spurious signals. Furthermore, the drawback of implanting a sensor in the body remains an obstacle for many patients [JAT02]. Transdermal techniques present a more convenient alternative as they only require sticking a patch on the skin. However, they collect fluid samples that are much smaller in magnitude than the volume of blood drops collected in invasive devices and contain strong noise. The success of these approaches strongly relies on the development of innovative sensors that enhance the permeation of glucose through the skin [JAT02]. In addition, many of these devices require calibration using traditional invasive devices. Finally, the amount of glucose contained in sweat or tears is not only much smaller than in blood, but also delayed. Patients who need to monitor their glucose concentration several times a day, as well as POC patients, need to rely on methods that measure the blood glucose concentration. The third category constituted by near-infrared techniques, which marked the birth of non-invasive blood sugar measurement research, remain too inaccurate [JAT02].

While much research is being performed to replace the current devices by non-invasive alternatives [KAK+91, YMK98, CC97], it seems that the traditional technologies will continue to maintain their position [JAT02], at least in the near future. Reducing the blood volume needed, and hereby the induced pain, remains a necessary and important field of research [JAT02].

There has been some research to reduce the volume of the blood drop in order to minimise the patient's pain sensation. The challenge herein is to maintain the required accuracy despite the smaller blood sample at hand. Some companies have managed to develop electrochemical measurement devices that require a blood drop volume of up to 0.3 µl [CF12, Alv08, JAT02]. In [AHKP11, PW13] a photometric measurement device is presented that requires blood sample volumes in the range 0.1 nl to 500 nl. The framework presented in this thesis is based on this measurement approach. It will be presented in detail in Sec. 2.4

2.3 Photometric Measurement of Glucose in Blood

Photometry is a very common method to measure chemical compounds in a fluid. The current consensus on the definition of photometry is the following: [BDE+09]

Photometry is the measurement of the ability of electromagnetic radiation to produce a visual sensation in a physically realisable manner, via a defined stimulation of human vision.

This means that photometry is a way to measure light as it is received by the human eye. For that, only wavelengths of light between 390 nm to 750 nm perceived by the human eye are considered in photometry.

In photometric glucose measurement, reflective photometry is used. It relies on chemical reagents that can react with specific ions, causing the reagents to change colour. The greater the concentration of the compound, the stronger the colour change. To detect the colour change and be able to numerically quantify it, the chemical reagent is illuminated by a light source. According to the Beer-Lambert law [Bee52], there is a relation between the absorbance behaviour of a solution and the concentration of light-absorbing compounds in that solution. This allows for quantitative photometric measurements that can be used to estimate the concentration of a compound on the basis of the measured light intensity or colour. A light source illuminates the chemical reagent from one side and a photodetector is placed on the other side to collect the reflected light. The latter is measured in so-called relative remission, an indicator of the amount of absorbance of the chemical substance. The glucose concentration, often given in mg/L (milligrams/litre), can then be calculated from the measured relative remission values.

The chemical reaction, on which photometric glucose measurement, is based, typically relies on glucose oxidation [TN09, GM13]. When the blood sample gets into contact with the chemical reagent on the test strip, it rehydrates the enzymes on the test strip and glucose reacts with the catalysing enzyme, e.g., glucose oxidase. This enzyme oxides the glucose to an ion that can be detected according to the Beer-Lambert law.

Glucose meters usually contain three components, as illustrated in Fig. 2.1: 1) the main body, which contains the light source and the photodetector. Computation is performed in the main body and it typically has a screen to show the result to user;

Figure 2.1: An example of an invasive, hand-held glucose meter containing 1) the main body, 2) the lancet, and 3) the test strip. Photo credit: ©iStock.com/ratmaner.

2) a lancet for extracting the blood drop; 3) the chemical test strip on which the blood drop is placed.

Figure 2.2 presents a model of a multilayer chemical test strip typically used in this context [BM03]. The upper layer contains an aperture to let the light from the light source in. The lower support layer at the bottom of the test strip contains a viewing point for the photodetector to capture the reflections. The layers in the middle constitute a semi-porous material that separates the red blood cells from the solution, in order not to interfere with the photometric measurement. Furthermore, a spreading layer ensures a homogeneous wetting of the whole test strip. The chemical reagent is applied to the mid and lower layers.

2.4 Novel Glucometer

As discussed in Sec. 2.2, reducing the blood sample volume needed for blood sugar measurement devices remains an important field of research, as it allows the reduction of the measurement inhibition and pain sensation for the patient. In [AHKP11, PW13] a photometric measurement approach is presented that requires a blood sample volume in the nano litre-range (nl-range), which is up to a magnitude of 100 smaller than is common in current state-of-the-art devices [AHKP11, JAT02]. Traditional photometric

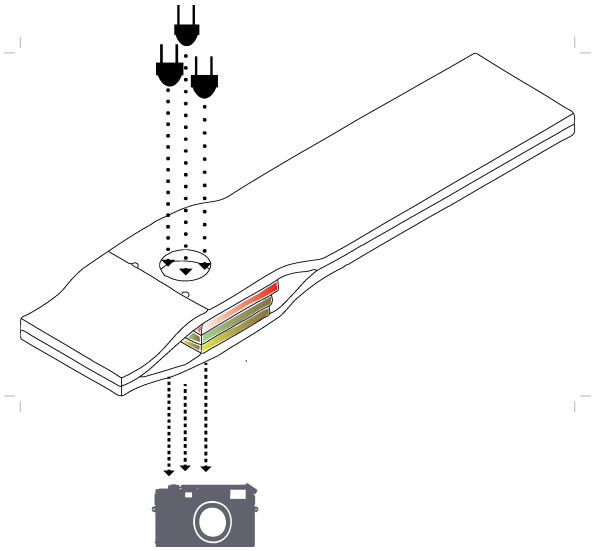


Figure 2.2: An example of a glucose test strip, illuminated by LEDs from the top layer, and containing a viewing point at the bottom layer to capture the reflections by a photosensor.

measurement principles use a blood sample in the µl-range that completely covers up the test strip. In such a setup a photodiode is used to capture the resulting reflections [BOW95, Lei09]. The use of much smaller blood samples renders this approach inaccurate, as the reaction region becomes much smaller than the observation area of the photodiode and the signal-to-noise-ratio rises. This motivates using a camera to observe the chemical reaction [PW13]. The measurement procedure illustrated in Fig. 2.3 constitutes the basis of the work performed in this thesis. The underlying photometric principle requires placing the extracted blood drop on the chemical test strip, so that the chemical reaction can be initiated. The resulting colour change is captured by illuminating the test area with a monochromatic light source and observing the reflections with a camera over time, e.g., a charge-coupled device (CCD) array. To counter the common problem of ambient light noise [Lei09] in photometry, the measurement area is placed completely inside the device and, thus, protected from ambient light. The camera observing the test strip results in frames at discrete time instants $n = n_0, n_1, \ldots$ where $n = tf_s$, $t \in \mathbb{R}$ is the continuous time and f_s is the sampling frequency. The frames show both the region where the reaction takes place as well as surrounding areas. The former represents the region of interest (ROI) and has to be delineated. The underlying intensity value representing the colour change, also termed relative remission $r \in \mathbb{R}$, is mapped to the underlying glucose concentration $g \in \mathbb{R}$. The estimated glucose concentration can, finally, be delivered to the user on a screen.

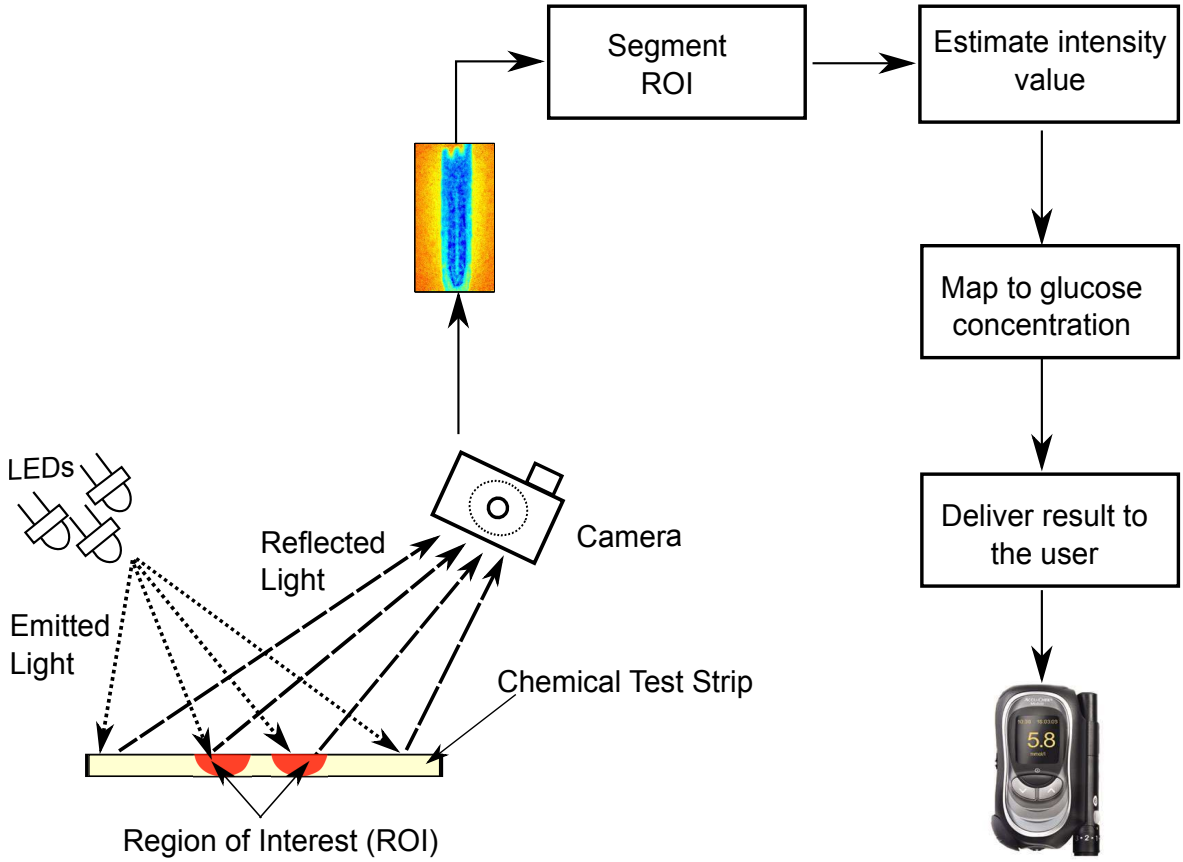


Figure 2.3: The photometric measurement principle used in this thesis to measure the glucose concentration in a blood sample.

2.4.1 Kinetic Behaviour in Glucometry

The term *enzyme kinetics* generally describes chemical reactions caused by the catalysis of enzymes. The main two parameters describing the kinetics behaviour are the saturation value r_C and the reaction rate τ.

The chemical reaction taking place on the test strip is an enzymatic reaction and can be modelled by the Michaelis-Menten model of enzyme catalysis [EK88, Bis08]. Typically, the chemical kinetic behaviour exhibits three different stages:

1. *Constant relative remission* where the reaction between the glucose and the chemical agent has not yet started.

2. *The moistening period* starts at $n = n_D$ and is characterised by a rapid drop of the relative remission value r_D followed by a slow decrease. It can be modelled by an exponential decay.

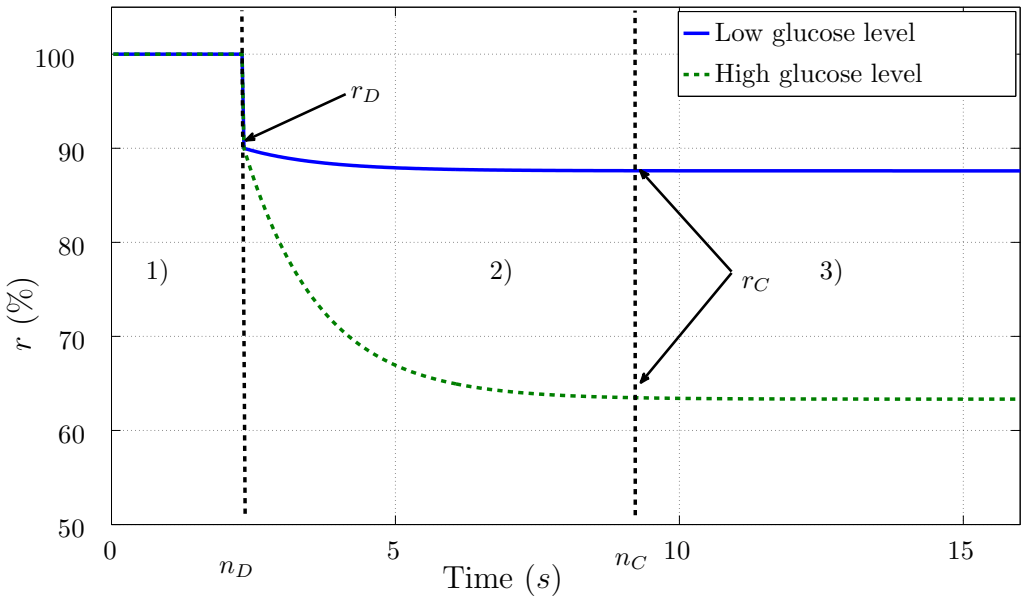


Figure 2.4: The typical course of the relative remission value r over time, termed kinetic curve, for a low (blue solid) and a high (green dashed) glucose level. The three distinct stages of the chemical reaction are illustrated.

3. *Saturation* of the chemical reaction is reached at $n = n_C$ at a saturated remission value r_C. The saturation time n_C is a characteristic of the chemical reaction and depends on the underlying glucose concentration, as well as other factors such as humidity, temperature, and haematocrit level. Typically, it lies between 10 s to 15 s.

Figure 2.4 depicts an idealised course of a typical chemical reaction for a low and a high glucose case. It is obtained by measuring the relative remission value for the whole duration of the chemical reaction. It is important to note that the remission value in this curve is normalised to the value 100% and does not represent an absolute remission value. The normalisation is performed with respect to the initial remission value of the starting frame and will be explained in detail in Sec. 3.2.1. For the high glucose range, the drop in relative remission is much larger than for the low glucose range. This can be traced back to the Beer-Lambert law [Bee52], which states that the degree of colour change is proportional to the concentration of the compound being tested. Furthermore, the reaction rate, i.e., the decay of the kinetic curve, is steeper for high glucose ranges than for lower ones. This is well-founded in the Michaelis-Menten model [EK88, Bis08].

2.4.2 Real Glucose Data Sets

Table 2.1: The real data sets used in the experiments.

Set	N_M	N_g	Υ	Sample Volume
A	48	5	6.45 µm /Pixel	10 nl to 100 nl
B	78	16	6.45 µm /Pixel	10 nl to 100 nl
C	78	16	6.45 µm /Pixel	10 nl to 100 nl
D	48	4	30 µm /Pixel	10 nl to 100 nl
E	200	10	30 µm /Pixel	around 1 nl
F	200	10	30 µm /Pixel	Standard

A major contribution of this thesis is the development of image processing techniques (Chapters 4 and 5) to segment the glucose images and estimate the underlying relative remission value. Accordingly, the image characteristics have to be well understood. In the following, we present the available real glucose data sets provided by Roche Diagnostics GmbH, Mannheim. These data sets will be used throughout the thesis to evaluate the proposed framework and the developed methods. They are obtained from a setup as in Fig. 2.3 using whole blood samples from blood donations injected with a glucose solution corresponding to the amount of glucose needed. Altogether, six different data sets are used, comprising 452 different measurements, which corresponds to a total of $263,064$ processed images. Information on the different data sets is given in Table 2.1. The number of measurements in the data sets is denoted by N_M, the number of different glucose concentrations tested by N_g, and the resolution of the images by Υ. The volume of the blood drop "Standard" indicates state-of-the-art ranges around 1 µl to 25 µl [JAT02]. Each measurement contains $N_f = 580$ frames obtained at a frame rate of $f_s = 30\,\text{fps}$.

The obtained frames can differ strongly depending on the stage of the chemical reaction, the underlying glucose concentration, and the measurement setup. Figure 2.5 shows some examples of the measurement frames at different time instants of the chemical reaction. In Fig. 2.5 (a) the reaction has not started yet and the observed frame is very homogenous. As soon as the chemical is hydrated by the blood drop, as seen in Fig. 2.5 (b), the chemical reaction is initiated and a slight colour change is already visible. With time, the reaction becomes more prominent and the reaction region can clearly be seen in Fig. 2.5 (c) as the area inside the dotted lines. In Fig. 2.5 (d) the reaction has saturated and the final colour change, indicating the underlying glucose concentration, has been reached. In the last two images the dry background area has also experienced a colour change. However, it is much weaker

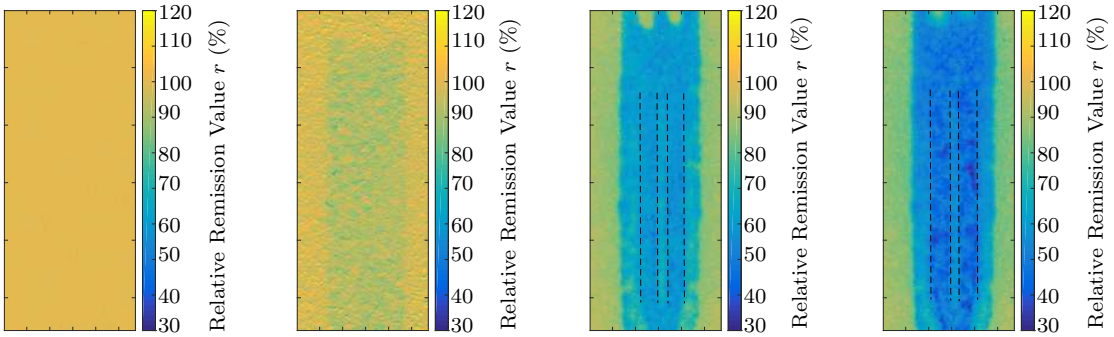


Figure 2.5: Examples of observed images at different time instants: (a) $t = 0\,\mathrm{s}$, (b) $t = 0.3\,\mathrm{s}$, (c) $t = 5\,\mathrm{s}$, (d) $t = 18\,\mathrm{s}$.

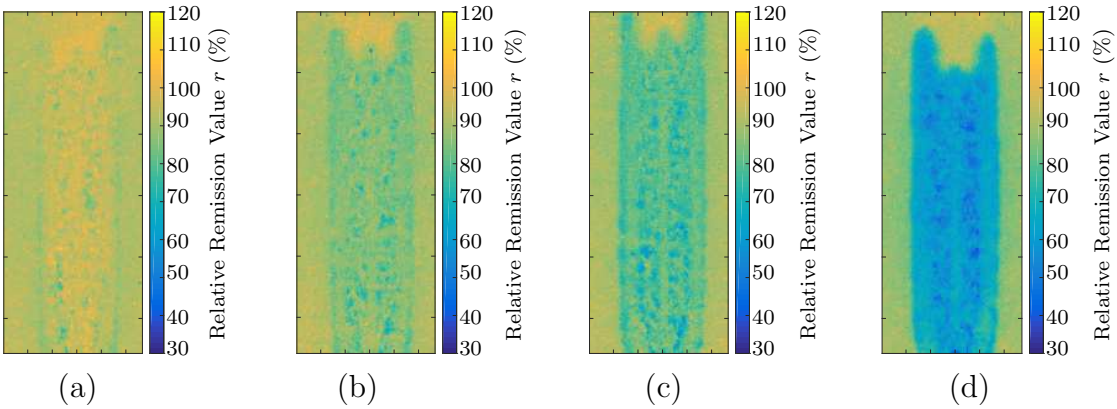


Figure 2.6: Examples of saturated images for different underlying glucose concentrations: (a) $g = 30\,\mathrm{mg/dl}$, (b) $g = 75\,\mathrm{mg/dl}$, (c) $g = 105\,\mathrm{mg/dl}$, (d) $g = 300\,\mathrm{mg/dl}$.

than the one exhibited by the reaction region. This is due to humidity effects in the device that lead to a hydration of the dry area and a slight reaction colour change. A third area is also clearly visible, namely, the region between the background and the reaction region, as well as the bridge between the two parts of the reaction region. The bridge shows intensities that lie in between both regions and is physically characterised by an inhomogeneous blood flow from one side of the reaction area to the other side. The border area between the reaction region and the background shows the same behaviour. Therefore, it induces a much weaker colour change than the reaction region. Figure 2.6 shows saturated frame examples from different underlying glucose concentrations. Very low glucose concentrations, as in Fig. 2.6 (a), are homogenous and comparable to the behaviour of dry frames as in Fig. 2.5 (a). Higher glucose concentrations show a stronger colour change and more structure in the images.

2.4.3 Validation Methods

Throughout the thesis, evaluation methods that are specific to glucose monitoring will be used to evaluate the proposed approaches [JD00, DFFC12]. They are presented in the sequel.

Coefficient of Variation (CV)

The coefficient of variation (CV) is a very popular measure to assess the precision of pharmacokinetic measurements [JD00]. The remission coefficient of variation $\mathrm{CV}_{\hat{r}}$ measures the variance of the relative remission value for measurements of equal underlying glucose concentrations, i.e., it is a measure of how precise the relative remission estimates are. Ideally, the $\mathrm{CV}_{\hat{r}}$ should go to zero, indicating that all relative remission estimates $\hat{r}$ of one underlying glucose concentration are equal. It is given for several measurements of a specific glucose concentration $g(\gamma)$ by

$$\mathrm{CV}_{\hat{r}_{g(\gamma)}} = \begin{cases} \hat{\sigma}_{\hat{\mathcal{R}}_{g(\gamma)}} & \text{if } g(\gamma) \leq 100\,\mathrm{mg/dl} \\ 100 \cdot \frac{\hat{\sigma}_{\hat{\mathcal{R}}_{g(\gamma)}}}{\hat{\mu}_{\hat{\mathcal{R}}_{g(\gamma)}}} & \text{if } g(\gamma) > 100\,\mathrm{mg/dl} \end{cases} \tag{2.1}$$

where $\gamma = 1, \ldots, N_g$, and $\hat{\sigma}_{\hat{\mathcal{R}}_{g(\gamma)}}$ and $\hat{\mu}_{\mathcal{R}_{g(\gamma)}}$ are the respective sample standard deviation and sample mean of the elements of the set $\hat{\mathcal{R}}_{g(\gamma)}$. $\hat{\mathcal{R}}_{g(\gamma)}$ denotes the test set of relative remission estimates obtained from several measurements of the same underlying glucose concentration $g(\gamma)$. To evaluate the precision of several measurements from several different glucose concentration values, a mean $\mathrm{CV}_{\hat{r}}$ is calculated as the sample mean of Eq. (2.1) over all underlying glucose concentrations

$$\mathrm{CV}_{\hat{r}} = \frac{1}{N_g} \sum_{\gamma=1}^{N_g} \mathrm{CV}_{\hat{r}_{g(\gamma)}}. \tag{2.2}$$

Clarke's Error Grid (CEG)

Clarke's Error Grid (CEG) [DFFC12] is a standard method for evaluating glucose measurement performance. It plots the actual glucose concentrations g against the estimated glucose concentrations $\hat{g}$ and classifies the error according to its medical severity. To this end, different regions are defined in the CEG as depicted in Fig. 2.7. Points lying within regions A and B are considered clinically acceptable. Points lying in the other regions represent errors that could lead to false medication and treatment. A common specification is to have at least 95% of all points in region A, maximally 5% of all points in region B [KC10] and no points in the other zones.

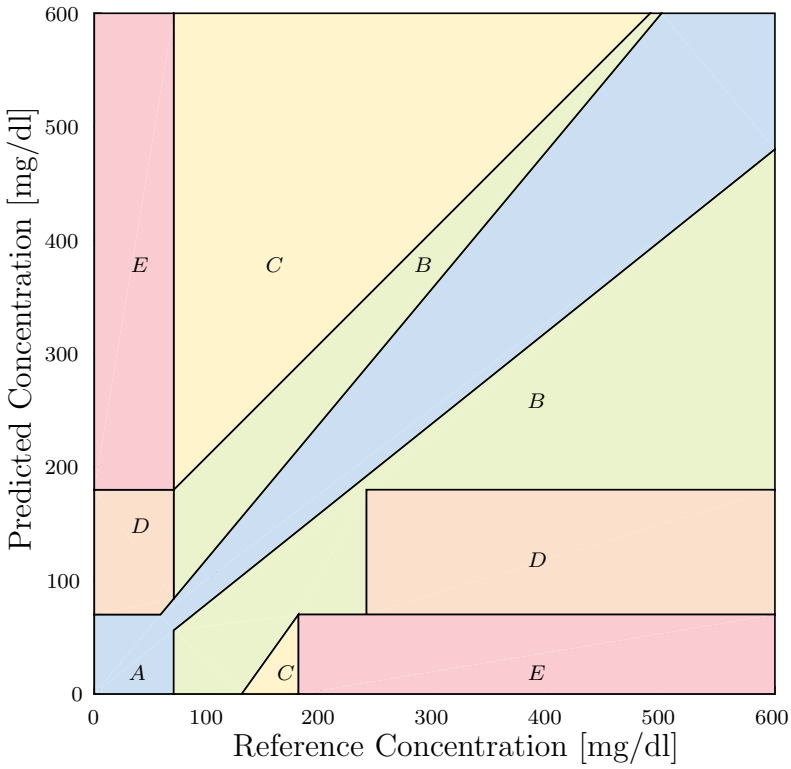


Figure 2.7: An Example of Clarke's Error Grid Analysis.

Glucose-Specific Mean Absolute Deviation (gMAD)

To quantify the accuracy of glucose concentration estimates a glucose-specific mean absolute deviation (MAD) inspired by [DFFC12] is adopted. Instead of using the standard MAD between the true and the estimated glucose level, a glucose-specific MAD (gMAD) is defined. The gMAD weights the errors according to their medical severity using a penalty function on the basis of the CEG. Hereby, errors made for hypoglycaemic cases below 70 mg/dl are weighted with the highest factor w_{gMAD} as they present critical short-term risks; hyperglycaemic cases presenting longer-term risks are weighted with a slightly smaller factor w_{gMAD}, and normal cases are given the factor $w_{\mathrm{gMAD}} = 1$. The gMAD reads

$$\mathrm{gMAD} = \frac{1}{N_M} \sum_{\gamma=1}^{N_M} |g(\gamma) - \hat{g}(\gamma)| \cdot w_{\mathrm{gMAD}}(g(\gamma) - \hat{g}(\gamma)), \tag{2.3}$$

where the penalty function is given by

$$\begin{aligned} w_{\mathrm{gMAD}}(g,\hat{g}) =& 1 + \alpha_L \overline{\sigma}_{g\leq T_L,\beta_L}(g)\sigma_{\hat{g}\geq g,\gamma_L}(\hat{g},g) \\ &+ \alpha_H \sigma_{g\geq T_H,\beta_H}(g)\sigma_{\hat{g}\leq g,\gamma_H}(\hat{g},g). \end{aligned} \tag{2.4}$$

The sigmoidal functions $\overline{\sigma}(\cdot)$ and $\sigma(\cdot)$ ensure a smooth transition and are given by

$$\sigma_{g\geq a,\epsilon}(g) = \begin{cases} 0 & \text{if } g \leq a \\ -\frac{1}{2}\xi^4 - \xi^3 + \xi + \frac{1}{2} & \text{if } a \leq g \leq a + \frac{\epsilon}{2} \\ \frac{1}{2}\xi^4 - \xi^3 + \xi + \frac{1}{2} & \text{if } a + \epsilon \leq g \\ 1 & \text{if } a - \epsilon \leq g \end{cases}, \tag{2.5}$$

and

$$\overline{\sigma}_{g\leq a,\epsilon}(g) = \begin{cases} 1 & \text{if } g \leq a - \epsilon \\ \frac{1}{2}\overline{\xi}^4 - \overline{\xi}^3 + \overline{\xi} + \frac{1}{2} & \text{if } a - \epsilon \leq g \leq a - \frac{\epsilon}{2} \\ -\frac{1}{2}\overline{\xi}^4 - \overline{\xi}^3 + \overline{\xi} + \frac{1}{2} & \text{if } a - \frac{\epsilon}{2} \leq g \leq a \\ 0 & \text{if } a \leq g, \end{cases}, \tag{2.6}$$

where

$$\xi := \frac{2}{\epsilon}\left(g - a - \frac{2}{\epsilon}\right); \quad \overline{\xi} := \frac{2}{\epsilon}\left(g - a + \frac{2}{\epsilon}\right). \tag{2.7}$$

The parameters contained in Eq. (2.4) are presented in Table 2.2, as applied in this thesis.

Table 2.2: Parameter values for the gMAD calculation.

α_L	α_H	β_L	β_H	γ_L	γ_H	T_L	T_H
1.5	1	30	100	10	20	85	155

According to the most recent ISO standards [Int13], the maximal permissible error is $\pm 15\,\mathrm{mg/dl}$ within a reference range of $0\,\mathrm{mg/dl}$ to $75\,\mathrm{mg/dl}$ and 20% for a reference range above $75\,\mathrm{mg/dl}$.

2.5 Summary

This chapter has introduced *Diabetes Mellitus* and has highlighted the importance of regular self-monitoring for effective diabetes treatment. The state of the art in hand-held glucose monitoring devices has been reviewed, as well as recent trends in reducing the blood sample volume needed for measurement.

Subsequently, the photometric blood glucose measurement principle used in this thesis has been presented. In this context, attention has been given to the temporal progression of the chemical reaction underlying the glucose measurement. Furthermore, the resulting glucose images have been presented and described, as well as the data sets that will be used for validation throughout the thesis. Finally, glucose-specific validation criteria have been laid out that will be used to assess the performance of the developed methods.

Chapter 3

A Framework for Image-Based Photometry in Glucose Measurements

This chapter introduces the framework developed in this thesis for measuring the blood glucose concentration from a photometric hand-held device, that requires very small blood samples, as presented in Chap. 2.

The main contributions in this chapter are constituted by: (i) the development of a pre-processing scheme for the glucose frames which is presented in Sec. 3.2.1; (ii) the detection of the onset of the chemical reaction, using a Neyman-Pearson hypothesis test as shown in Sec. 3.2.2; (iii) the application of a grey-level co-occurrence matrix-based texture approach to identify faulty glucose measurements and the selection of suitable texture features for this purpose as depicted in Sec. 3.2.3; (iv) the derivation of a model for the chemical kinetics that is incorporated in a temporal tracking and prediction setup to enhance accuracy and drastically reduce measurement time as explained in Sec. 3.2.5; (v) and finally the validation of the proposed framework by means of experiments conducted using real data sets as provided in Sec. 3.3. Finally, a conclusion is drawn in Sec. 3.4.

The material presented in this chapter is partly taken from [DZ13,DZ14a,SDZ16,DZ16].

3.1 Introduction and Motivation

The main contribution of this chapter is the development of a framework to measure the concentration of an analyte in a fluid from an image-based measurement procedure. Specifically, this thesis applies the developed methods to measure the glucose concentration in blood. It is noteworthy, that this framework is not confined to this sole application but is suitable for other medical applications, e.g. that rely on photometry.

Invasive hand-held glucometers typically require the patient to extract a blood drop from their finger and place it on a chemical test strip. This causes hydration of the enzyme on the test strip, and a chemical reaction between the blood glucose and the catalyzing enzyme is induced. As described in Sec. 2.3, the result of the

reaction is a colour change of the test strip, that is proportional to the amount of glucose contained in the blood drop. To measure the degree of the colour change, the reaction area is illuminated by a light source and the reflections are captured by photosensors. The underlying measurement setup used in this thesis reduces the blood sample volume from the micro litre-range (µl-range) to the nano litre-range (nl-range). The chemical test strip, as well as the chemical reaction taking place on it, are observed by a camera producing a set of frames that describe the measurement. Due to the very small size of the blood drop, the observed frames do not only show the reaction region but several other areas, as illustrated in Sec. 2.4.2. Furthermore, they may contain artefacts or anomalies that occlude the reaction region. The final colour, after saturation of the chemical reaction, is an indicator of the underlying glucose concentration level.

The framework takes as an input the raw glucose frames obtained by the camera during the measurement and outputs the estimated glucose concentration of the examined blood drop. The raw frames are typically deteriorated by Gaussian noise and contain strong illumination artefacts. Therefore, a pre-processing scheme, as developed in Sec. 3.2.1 is essential to facilitate further processing and image understanding.

The hydration of the chemical test strip does not happen instantly, and as illustrated in Fig. 2.4 a drop time n_D is exhibited before the chemical reaction actually begins. This typically amounts to around 1.3 s to 1.7 s. To avoid unnecessary processing of frames captured when the chemical reaction has not yet begun, a drop time detection method is proposed, that is based on a Neyman-Pearson hypothesis test.

In some cases, the blood drop can be insufficient or badly placed on the test strip such that it becomes much more difficult to observe the chemical reaction. Also, artefacts such as large dust particles or air bubbles lying on the surface can occlude the reaction region leading to an insufficient size of the region of interest. Such cases require the immediate termination of the measurement process and repetition of the measurement by the user, so as to not deliver a false measurement value to the patient. Moreover, early detection of these measurements leads to reducing unnecessary energy consumption caused by processing faulty frames. A texture analysis approach is proposed to detect such cases, which is based on Haralick's grey-level co-occurrence matrices (GLCM) [HSD73, Har79].

Once a measurement has been detected as correct, the processing continues with the segmentation of the images to identify the reaction region, termed in the

rest of this thesis as the region of interest (ROI). The approaches developed for segmenting the glucose images constitute a major contribution of this thesis and will be presented in detail in Chap. 4 and 5. The final colour change exhibited in the ROI gives inference on the underlying concentration of glucose in the blood sample. Therefore, the relative remission of the ROI after saturation has to be estimated.

The early detection of the saturation time of the chemical reaction is a challenging issue. Ideally, the chemical kinetic curve evolves as in Fig. 2.4. Final saturation is typically reached after 10 s to 15 s. On the one hand, waiting for the saturation of the chemical reaction implies long measurement times for the patient and, hereby, reduced usability of the glucometer, which can be a cause of irregular self-control. On the other hand, estimating the glucose value prematurely can result in erroneous measurements of blood glucose concentration. So far, the convergence of kinetic curves in hand-held devices has been defined by the slope of the kinetic curve over a predefined time period [AHKP11]. This chapter develops a model of the chemical kinetics, and applies it in a state estimation and prediction framework, to detect convergence earlier and enhance estimation accuracy. Once the converged relative remission value has been estimated, it is used to estimate the underlying glucose concentration that is presented to the user.

Section 3.3 elucidates results of the proposed framework, validating that the framework exhibits high accuracy and reduces the measurement time, while using drastically smaller blood sample volumes than state-of-the-art methods.

3.2 Proposed Framework

The proposed framework is summarised in Fig. 3.1. The input is given by the raw frames $\mathbf{I}_{\text{raw}}^{(n)}$ of size $\tilde{M}_x \times \tilde{M}_y$, obtained by the camera at discrete time instances $n = 1, 2, \ldots$. These represent the reflectivity behaviour of the observed area, as described in Sec. 2.3.

First, the frames are pre-processed allowing for a better segmentation and estimation in later stages. A hypothesis test is performed on the pre-processed, grey-scale reflectance image $\mathbf{I}^{(n)}$ of size $M_x \times M_y$, obtained at frame n. The purpose of this step is to detect whether the chemical reaction has started, i.e., if $n \geq n_D$. If the onset of the chemical reaction has been detected, an artefact detection test is performed, that is based on the calculation of grey-level cooccurrence-matrix (GLCM) features. If a

measurement is detected as faulty, the measurement procedure is broken off and the user gets a message to re-perform the measurement. If, however, the measurement is detected as correct, the framework proceeds with the next steps.

The segmentation of the glucose images results in the identification of the region of interest (ROI), i.e. the region where the blood drop is placed and where the reaction has taken place. The ROI contains the reflectivity information of interest and allows inference on the underlying glucose concentration.

After segmenting the image, the relative remission values of the different regions are estimated and the relative remission value corresponding to the reaction region is identified. The estimated relative remission values $\hat{r}(n)$ are tracked over time and used in a state estimation and prediction setup to estimate the convergence value. The result of this stage is the converged relative remission value $\hat{r}_C$. Once this value is found, it is mapped to its corresponding glucose level $\hat{g}$. If the reaction has not converged yet, the next frame is processed in the same manner. The different blocks as depicted in Fig. 3.1 will be described in further detail in the sequel.

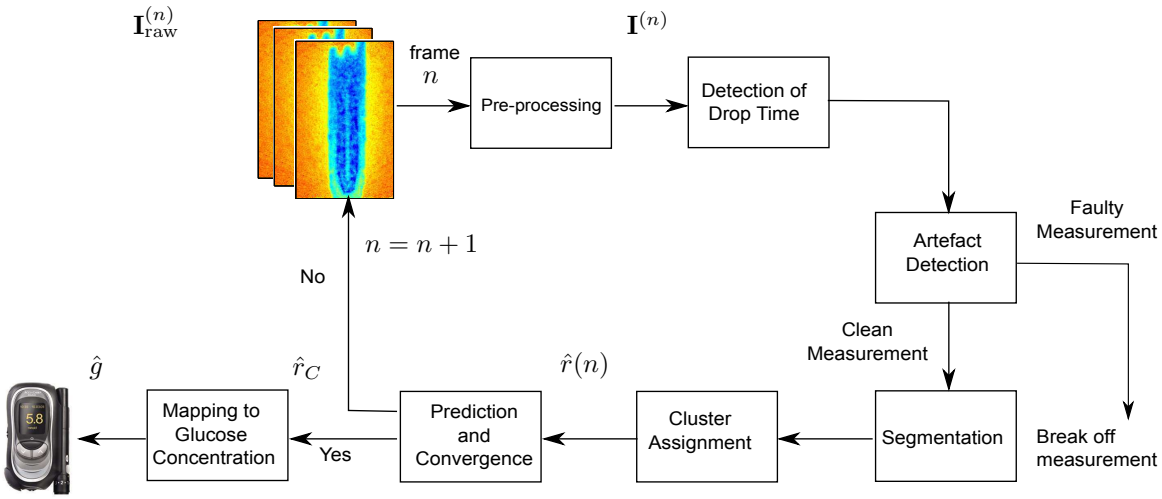


Figure 3.1: The proposed framework for estimation of glucose concentrations from photometric glucose images.

3.2.1 Pre-processing of Glucose Images

Figure 3.2 (a) depicts a typical raw frame $\mathbf{I}_{\text{raw}}^{(n)}$ obtained by the camera. Not only does the frame show an area larger than the test strip, but it is also evident that extracting information from the test strip area, depicted in turquoise, in this state could prove to

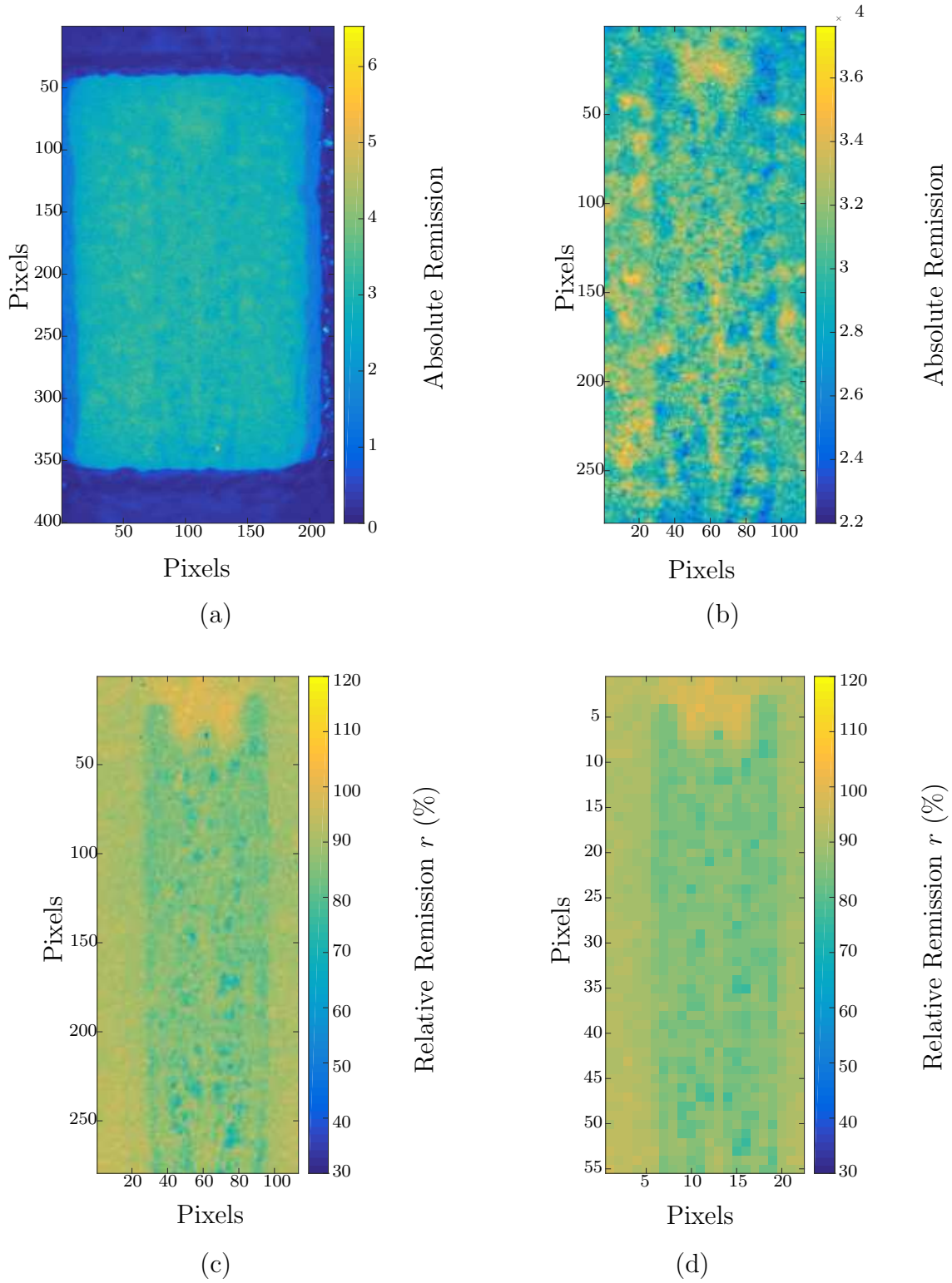


Figure 3.2: The pre-processing stages of glucose frames: (a) The raw frame $\mathbf{I}^{(n)}_{\text{raw}}$, (b) the cropped frame $\mathbf{I}^{(n)}_{\text{crop}}$, (c) the normalised frame $\mathbf{I}^{(n)}_{\text{norm}}$, and (d) the binned frame $\mathbf{I}^{(n)}$.

be a difficult task. For that a raw frame needs to be pre-processed to simplify further processing. The pre-processing chain suggested in this thesis constitutes three main phases [DZ13]. They will be discussed in the sequel.

The first step is to extract the test strip from the frame, i.e. to crop the turquoise foreground from the dark blue background in Fig. 3.2 (a). This is done using Otsu thresholding [Ots75]. Otsu's method is a very intuitive and effective grey-level thresholding technique, which aims at separating pixels of an image into two classes by maximizing the separability of the resulting classes. To this end, only the zeroth and the first order moments of the histogram of the image under test are needed. Otsu assumes that the grey-levels $\gamma = 1, 2, \ldots, N_\gamma$ in an image can be classified into two classes. He relies on the following discriminant criterion to separate the classes: $\eta_{\text{Otsu}} = \hat{\sigma}_B^2/\hat{\sigma}_T^2$, where $\hat{\sigma}_B^2$ is the between-class variance and $\hat{\sigma}_T^2$ is the total variance. By maximizing $\hat{\sigma}_B^2(\delta_{\text{Otsu}})$ with respect to a certain threshold level δ_{Otsu} the optimal grey-level threshold can be found

$$\hat{\sigma}_B^2(\delta_{\text{Otsu}}^*) = \max_{1 \leq \delta_{\text{Otsu}} \leq N_\gamma} \hat{\sigma}_B^2(\delta_{\text{Otsu}}) \ , \tag{3.1}$$

where $\hat{\sigma}_B^2(\delta_{\text{Otsu}})$ is calculated as

$$\hat{\sigma}_B^2(\delta_{\text{Otsu}}) = \frac{\Big(\hat{\mu}_T \cdot \hat{w}(\delta_{\text{Otsu}}) - \hat{\mu}(\delta_{\text{Otsu}})\Big)^2}{\hat{w}(\delta_{\text{Otsu}}) \cdot \Big(1 - \hat{w}(\delta_{\text{Otsu}})\Big)}, \tag{3.2}$$

where

$$\begin{aligned} \hat{\mu}_T &= \sum_{\gamma=1}^{N_\gamma} \gamma \hat{p}_\gamma, \quad \hat{\mu}(\delta_{\text{Otsu}}) = \sum_{\gamma=1}^{\delta_{\text{Otsu}}} \gamma \hat{p}_\gamma, \\ \hat{w}(\delta_{\text{Otsu}}) &= \sum_{\gamma=1}^{\delta_{\text{Otsu}}} \hat{p}_\gamma. \end{aligned} \tag{3.3}$$

$\hat{\mu}_T$ is the total sample mean of grey-levels in the image and $\hat{p}_\gamma$ is the probability of occurrence of a certain grey-level γ. $\hat{\mu}$ denotes the sample mean of the first class and $\hat{w}$ is the probability of occurrence of said class. The grey-level probability of occurrence $\hat{p}_\gamma$ is calculated as the ratio n_γ/N_T , n_γ being the number of pixels with this grey-level and N_T the total number of pixels.

The binary image obtained after thresholding is further processed by using the morphological operation of opening. By using a disk-shaped structuring element [GW02] small remaining structures are removed from the foreground. Opening describes the dilation of the erosion of an image $\mathbf{I}_{\text{raw}}^{(n)}(\tilde{m}_x, \tilde{m}_y)$ by a structuring element

$\mathbf{A}_{\mathrm{SE}}(\tilde{m}_x, \tilde{m}_y)$, where $\tilde{m}_x = 1, \ldots, \tilde{M}_X$ and $\tilde{m}_y = 1, \ldots, \tilde{M}_y$ are the row and column indices of the image $\mathbf{I}^{(n)}_{\mathrm{raw}}$. The structuring element used here is disk-shaped with a radius of $R_{\mathrm{SE}} = 5$. Finally, the cropping indices are found as the edges of the detected foreground object in the first frame. These cropping indices are then used to crop all consecutive frames. An example of a cropped frame is given in Fig. 3.2 (b). Note that the frame size has changed compared to Fig. 3.2 (a).

The cropped frames $\mathbf{I}^{(n)}_{\mathrm{crop}}$ obtained by the camera are normalised w.r.t. the initial reflectivity prior to the start of the chemical reaction. To this end, a set of calibration frames are recorded before the chemical reaction starts. The so-called dark frames $\mathbf{I}^{(n)}_{\mathrm{dark}}$ are obtained prior to the insertion of the test strip into the device without illuminating the LEDs. The so-called light frames $\mathbf{I}^{(n)}_{\mathrm{light}}$, are obtained before the test strip enters the device with illuminated LEDs to account for illumination effects. All in all, $N_{\mathrm{F,dark}} = 4$ dark frames and $N_{\mathrm{F,light}} = 25$ light frames are recorded for calibration. The normalisation is performed using the calibration frames and results in a grey-level image $\mathbf{I}^{(n)}_{\mathrm{norm}}$ with pixels in the range $[0 - 100]$. The frames are normalised as follows

$$\mathbf{I}^{(n)}_{\mathrm{norm}}(m_x, m_y) = \frac{\mathbf{I}^{(n)}_{\mathrm{crop}}(m_x, m_y) - \overline{\mathbf{I}}^{(n)}_{\mathrm{dark}}(m_x, m_y)}{\overline{\mathbf{I}}^{(n)}_{\mathrm{light}}(m_x, m_y) - \overline{\mathbf{I}}^{(n)}_{\mathrm{dark}}(m_x, m_y)}, \tag{3.4}$$

where $m_x = 1, \ldots, M_x$ and $m_y = 1, \ldots, M_y$ are the row and respectively, column sizes after cropping. $\overline{\mathbf{I}}^{(n)}_{\mathrm{dark}}$ is the frame calculated as the average over all dark frames, while $\overline{\mathbf{I}}^{(n)}_{\mathrm{light}}$ is the average frame of all light frames. Figure 3.2 (c) shows that the normalisation enhances the ROI, simplifying further processing.

The last pre-processing step constitutes binning the frames, i.e. averaging over windows of size $B \times B$ [DZ13]. The benefit of this step is twofold: 1) it serves as a noise reduction step, and contributes to an improved image segmentation and relative remission estimation as will be shown in Sec. 3.3; 2) the image size is reduced leading to more efficient computation. Finally, the pre-processed image is denoted by $\mathbf{I}^{(n)}$. It is evident when comparing Fig. 3.2 (a) to Fig. 3.2 (d) that the pre-processing significantly enhances the frame such that the reaction area is better recognisable. For the rest of this thesis, the x- and y-axis of the glucose images will be omitted to enhance readibility. The image sizes will always correspond to the image size of Fig. 3.2 (d).

3.2.2 Detection of Drop Time

As illustrated in Fig. 2.4, the chemical reaction exhibits a constant stage for $n < n_D$, as the glucose in the blood sample has not yet been detected. The time n_D when the drop occurs, i.e., when the chemical reaction starts, depends on on the particular chemical reagent, the blood sample at hand, and the underlying temperature and humidity conditions in the device. This period does not contribute to the estimation of the underlying relative remission value, and therefore does not need to be processed. By detecting the drop time n_D, unnecessary computation is saved.

To this end, an approach based on the Neyman-Pearson hypothesis test of variance is proposed [DZ16]. The advantage of using the Neyman-Pearson test is, that it provides a meaningful way to set a threshold by incorporating the image statistics and controlling the false-alarm rate.

Figure 3.3 illustrates the probability density functions $f_{\tilde{\mathbf{x}}^{(n)}}$ of the intensities of two exemplary images taken at time $n_1 < n_D$ and $n_C > n_2 > n_D$, i.e. directly after the occurrence of the drop. Extensive analysis of real data showed that it is reasonable to assume that the mean-adjusted frames for $n < n_D$ follow a zero-mean normal distribution with variance σ_1^2, whereas after the drop the mean-adjusted frames can be modelled by a zero-mean normal distribution with variance $\sigma_2^2 > \sigma_1^2$. This increase in variance can be explained by the onset of the chemical reaction and thereby the onset of the colour change on the test strip. Hereby, σ_1^2 can be estimated as the sample variance from available data sets. However, σ_2^2 is unknown.

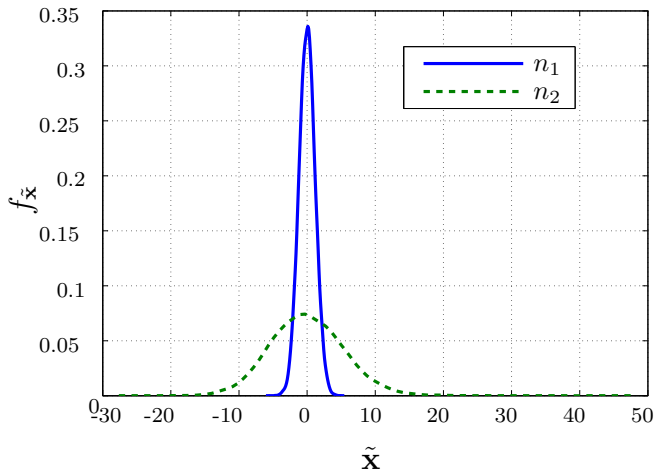


Figure 3.3: Pdf estimate of mean-adjusted intensities of a frame at $n_1 < n_D$ (blue solid) and $n_2 \geq n_D$ (green dashed).

Here, the frames are dealt with in vectorised form and will be denoted by

$$\mathbf{x}^{(n)} = \mathrm{vec}\left(\mathbf{I}^{(n)}\right), \tag{3.5}$$

where $\mathbf{x}^{(n)}$ is of size $1 \times L, L = M_x \cdot M_y$. Now, the hypothesis test can be formulated as

$$\begin{aligned} \mathcal{H}_0 &: f_{\tilde{\mathbf{x}}^{(n)}} \sim \mathcal{N}(0, \sigma_1^2), \quad n < n_D \\ \mathcal{H}_1 &: f_{\tilde{\mathbf{x}}^{(n)}} \sim \mathcal{N}(0, \sigma_2^2), \quad n \geq n_D \end{aligned} \tag{3.6}$$

where $\tilde{\mathbf{x}}^{(n)}$ is the mean-adjusted version of $\mathbf{x}^{(n)}$ and $f_{\tilde{\mathbf{x}}^{(n)}}$ denotes the pdf of $\tilde{\mathbf{x}}^{(n)}$. Calculating the likelihood-ratio leads to the following test statistic

$$T\big(\tilde{\mathbf{x}}^{(n)}\big) = \sum_{l=1}^{L} \left(\tilde{x}_l^{(n)}\right)^2 \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \underbrace{L \frac{\frac{2}{L}\ln(\delta_D) + \ln(\frac{\sigma_2^2}{\sigma_1^2})}{\frac{1}{\sigma_1^2} - \frac{1}{\sigma_2^2}}}_{\delta_D'}, \tag{3.7}$$

where the threshold δ_D' is set to ensure a nominal false alarm rate. Assuming $\tilde{\mathbf{x}}^{(n)}$ to be spatially i.i.d., the distribution of $\frac{T(\tilde{\mathbf{x}}^{(n)})}{\sigma_1^2}$ under $\mathcal{H}_0$ and $\frac{T(\tilde{\mathbf{x}}^{(n)})}{\sigma_2^2}$ under $\mathcal{H}_1$ can be shown to follow a χ^2-distribution with L degrees of freedom [Kay93]. The spatial i.i.d. assumption is justifiable in this case as frames prior to the onset of the reaction are observed, where no structure is present in the image. L represents the number of pixels in an image and is, therefore, quite large. Thus, we can approximate the χ^2-distribution by a Gaussian distribution [Kay93]. The probability of false alarm P_{FA} then becomes

$$P_{\mathrm{FA}} = Q\left(\frac{\delta_D'}{\sigma_1^2}\right), \tag{3.8}$$

where $Q(\cdot)$ is the complementary cumulative distribution function of the standard Gaussian distribution given as [Kay93]

$$\begin{aligned} Q(x) &= \frac{1}{2}\mathrm{erfc}\left(\frac{x}{\sqrt{2}}\right) \\ &= \frac{1}{\sqrt{2\pi}} \int_x^\infty \exp\Big(-\frac{u^2}{2}\Big) du, \end{aligned} \tag{3.9}$$

$\mathrm{erfc}(\cdot)$ being the error function. Given a fixed probability of false alarm P_{FA}, the threshold can be calculated as

$$\delta_D' = Q^{-1}(P_{\mathrm{FA}}) \cdot \sigma_1^2. \tag{3.10}$$

3.2.3 Faulty Measurement Detection Using Texture Analysis

Figures 2.5 and. 2.6 depict examples of the observations, obtained by the camera, at the different time stages of the chemical reaction and for different underlying glucose concentrations. Fig. 2.5 (a) shows an observation at time $n < n_D$, where the reaction has not yet taken place. Figures 2.6 show observations after convergence ($n > n_C$) for low and high glucose values, respectively. Errors may occur in the measurement process that do not lead to these typical observations, hereby leading to false estimates of the underlying glucose values. Figure 3.4 exemplifies such cases obtained for $n > n_C$, where in Fig. 3.4 (a) and 3.4 (c) the blood drop volume is either insufficient or incorrectly placed on the test strip. This can happen when the patient draws too little blood from the finger. In Fig. 3.4 (b) the test strip is not placed correctly in the observation area of the camera, leading to an occlusion of the reaction region.

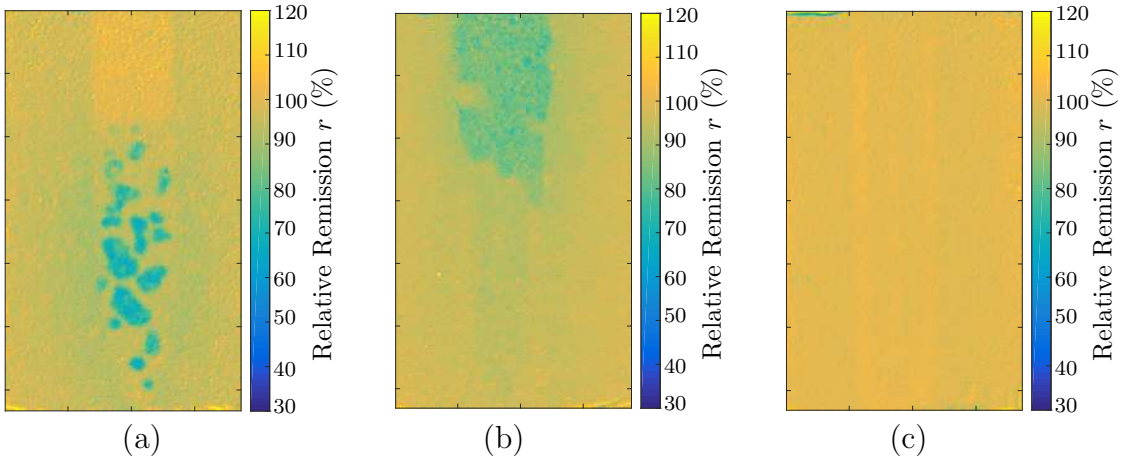


Figure 3.4: Examples of faulty images. (a) and (c) are examples of an insufficient blood drop, in (b) the test strip is not placed correctly in the observation area.

The analysis of the given frames results in the observation that the different cases are not only uniquely marked by their intensity behaviours, but also by their texture. The advantage of texture is that it shows unique characteristics for faulty measurements irrespective of the underlying glucose concentration. This is opposed to intensity that is closely related to the underlying glucose concentration. Therefore, an approach is developed that uses the texture information in the frames to identify erroneous measurements [DZ14a]. The aim is to detect faulty measurements at the beginning of the moistening period at $n = n_D$, instead of having to wait for the final

convergence value at $n = n_C$.

In general, texture describes the spatial and statistical distribution of grey-levels in a frame. Many methods exist to describe texture; they can be divided into four categories according to [TJ93]: statistical, geometrical, model-based and signal processing. The most common method was proposed by Haralick in [HSD73, Har79]. It assumes that the texture information in an image is contained in the average spatial relationship that different grey-levels have to one another. The so-called grey-level co-occurrence matrix (GLCM) is defined to describe this and statistical features are derived from it. Haralick's texture analysis method has been used for a wide range of applications such as biomedical and aerial image segmentation as well as quality control, to describe different regions of an image or detect anomalies [HSD73,Har79,CCAG09,ST99].

In the photometric application it is intuitive to analyse the images w.r.t. texture as different underlying physical textures exist in the frames: (i) the dry chemical test strip, (ii) the wet test strip when the blood sample is applied, (iii) the presence of artefacts or air bubbles, (iv) as well as different textures for different stages of the chemical reaction and different underlying glucose concentrations.

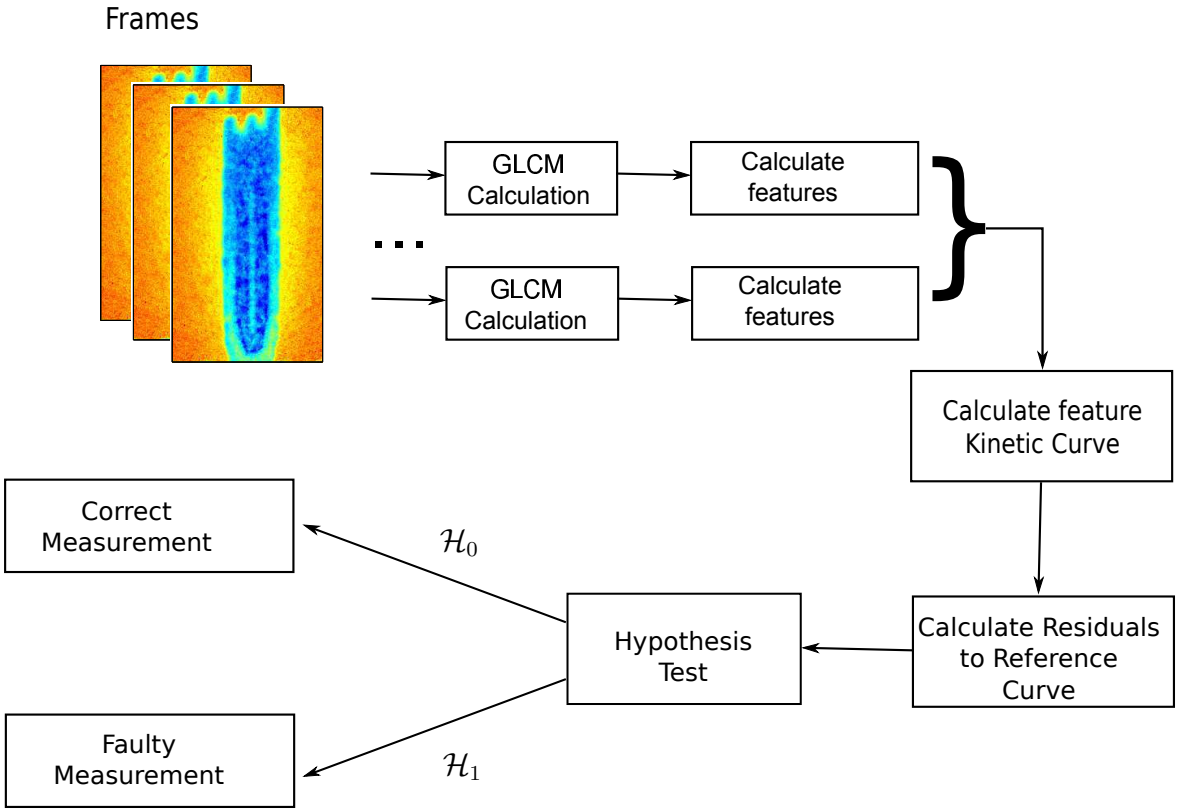


Figure 3.5: Flowgraph of the proposed GLCM-based algorithm for detection of faulty glucose measurements.

In [HSD73], Haralick proposes a grey-level co-occurrence texture method, that has become very popular for texture segmentation applications. Given a frame $\mathbf{I}^{(n)}$ of size $M_x \times M_y$ with the grey-levels $\gamma = \{1, 2, ..., N_\gamma\}$, the method analyses the spatial relationships of the occurring grey-levels. The underlying assumption is that texture information can be specified using a square matrix of relative frequencies with which two neighbouring pixels with defined grey-level values (γ_i, γ_j) occur at a certain distance d_γ and angle α_γ. This is described using the so-called grey-level co-occurrence matrix (GLCM) $\varrho(\gamma_i, \gamma_j; d_\gamma, \boldsymbol{\alpha}_\gamma, L_\gamma)$. 8-pixel neighbourhoods are considered such that the angles used are $\boldsymbol{\alpha}_\gamma = \{0°, 45°, 90°, 135°\}$ as well as their symmetric counterparts. The size of the matrix L_γ is typically set to the number of occurring grey-levels N_γ.

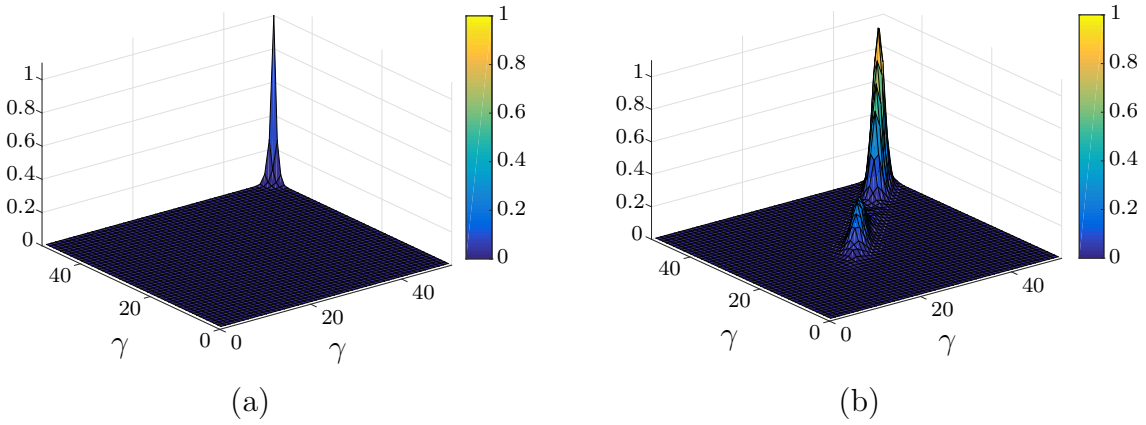


Figure 3.6: Examples GLCMs for (a) a faulty and (b) a correct measurement.

However, coarser quantization can be used to accelerate computation and reduce noise [CD05]. After calculating the GLCM, statistical features can be calculated using the co-occurring relative frequencies, i.e. the entries of the GLCM. The features can be defined to be directional calculating the GLCM for one angle α or rotationally invariant by calculating the GLCMs of all angles and averaging over the resulting features [HSD73].

For the detection of faulty glucose measurements, features are needed that can distinguish between a correct measurement and a false measurement, regardless of the underlying glucose value. All GLCM-based features proposed in [HSD73] and [ST99] are analysed, and three features are identified that proved to be sensitive to faulty measurements: the energy (E), the maximum probability (M) and the correlation (C).

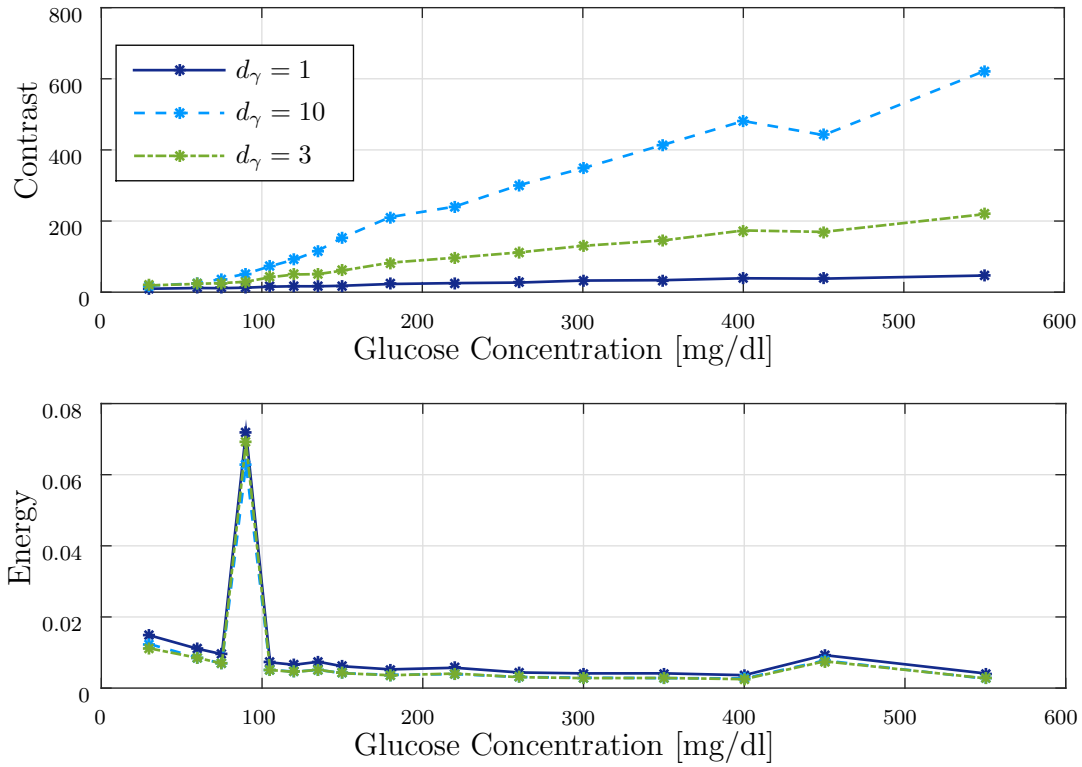


Figure 3.7: Comparison of the features contrast and energy for different values of d_γ applied to glucose images with different underlying glucose concentrations.

They are given as follows [HSD73]:

$$s^E = \sum_{\gamma_i=1}^{N_G} \sum_{\gamma_j=1}^{N_G} \varrho(\gamma_i, \gamma_j; d_\gamma, \boldsymbol{\alpha}_\gamma, L) \tag{3.11}$$

$$s^M = \max_{\gamma_i \gamma_j} \varrho(\gamma_i, \gamma_j; d_\gamma, \boldsymbol{\alpha}_\gamma, L) \tag{3.12}$$

$$s^C = \frac{\sum_{\gamma_i=1}^{N_G} \sum_{\gamma_j=1}^{N_G} \gamma_i \gamma_j (\varrho(\gamma_i, \gamma_j; d_\gamma, \boldsymbol{\alpha}_\gamma, L) - \hat{\mu}_x \hat{\mu}_y)}{\hat{\sigma}_x \hat{\sigma}_y}, \tag{3.13}$$

where $\hat{\mu}_x, \hat{\mu}_y, \hat{\sigma}_x, \hat{\sigma}_y$ are the sample means and standard deviations of the marginal distributions of $\varrho(\gamma_i, \gamma_j; d_\gamma, \boldsymbol{\alpha}_\gamma, L)$.

For frames such as the ones depicted in Fig. 3.4 the GLCM is characterized by a very high peak at the first couple of entries of the diagonal as can be seen in Fig. 3.6 (a). Contrarily, for cases such as Fig. 2.5 and Fig. 2.6 the GLCM shows values spread over the diagonal as can be seen in Fig. 3.6 (b). This behaviour is observed regardless of the underlying glucose value. The features s^E, s^M, s^C are well suited to detect this behaviour and hence to identify incorrect measurements.

Figure 3.7 compares the feature values of contrast and energy using real glucose images of different underlying glucose concentrations. The different curves correspond to varying the parameter d_γ used in the calculation of the GLCMs. Two faulty measurements exist in these sets, one for 90 mg/dl which is depicted in Fig. 3.6 (c) and another for 450 mg/dl depicted in Fig. 3.6 (b). The feature contrast is clearly correlated with the underlying glucose concentration and cannot reliably be used to detect faulty measurements. The feature energy, however, shows a nearly constant behaviour over correct measurements and is able to identify faulty ones, by detecting major deviations from this behaviour.

The proposed algorithm for the detection of faulty glucose measurements is summarised in Fig. 3.5. It is presented in detail in the sequel.

Step 1.: The first step is the calculation of the GLCMs $\varrho(\gamma_i, \gamma_j; d_\gamma, \boldsymbol{\alpha}_\gamma, L)$ for each incoming frame. An extensive analysis has shown that using $L = N_\gamma$ produces the same accuracy of results as $L = N_\gamma/2$. To accelerate computation $L = N_\gamma/2$ is used. The GLCMs are calculated using a distance of $d_\gamma = 1$ as the glucose images are characterized by a micro-structured texture. A total of $N_{\boldsymbol{\alpha}_\gamma} = 4$ GLCMs are calculated for each frame, corresponding to the $N_{\boldsymbol{\alpha}_\gamma} = 4$ angles $\boldsymbol{\alpha}_\gamma = [0°, 45°, 90°, 135°]$.
Step 2.: The features from Eq. (3.11) - (3.13) are calculated for each GLCM, resulting in a feature vector

$$\mathbf{s}^i = [s^i_{\boldsymbol{\alpha}_\gamma(1)}, s^i_{\boldsymbol{\alpha}_\gamma(2)}, s^i_{\boldsymbol{\alpha}_\gamma(3)}, s^i_{\boldsymbol{\alpha}_\gamma(4)}], \quad i \in \{E, M, C\}$$

for each frame. To make the features rotationally invariant the mean of the features w.r.t the angles is calculated, hereby, building the following feature vector for each frame

$$\mathbf{s} = \begin{bmatrix} \frac{1}{N_{\boldsymbol{\alpha}_\gamma}} \sum_{k=1}^{N_{\boldsymbol{\alpha}_\gamma}} s^E_{\boldsymbol{\alpha}_\gamma(k)} \\ \frac{1}{N_{\boldsymbol{\alpha}_\gamma}} \sum_{k=1}^{N_{\boldsymbol{\alpha}_\gamma}} s^M_{\boldsymbol{\alpha}_\gamma(k)} \\ \frac{1}{N_{\boldsymbol{\alpha}_\gamma}} \sum_{k=1}^{N_{\boldsymbol{\alpha}_\gamma}} s^C_{\boldsymbol{\alpha}_\gamma(k)} \end{bmatrix}. \tag{3.14}$$

The feature vector $\mathbf{s}$ is normalised using linear scaling to unit range as described in [AH01]

$$\tilde{\mathbf{s}} = (\mathbf{s} - \mathbf{l}) \oslash (\mathbf{u} - \mathbf{l}),$$

where $\mathbf{l} = [l^E, l^M, l^C]^T$ and $\mathbf{u} = [u^E, u^M, u^C]^T$ contain the lower and upper values of the features (E), (M) and (C) and $\oslash$ denotes element-wise division of the vectors.
Step 3.: Steps 1-2 are performed for every frame in a measurement, producing a feature kinetic curve $k^i(n)$ with $i \in \{E, M, C\}$ and $n = 1, ..., N$, where N is the total number of frames. Figure 3.8 shows examples of feature kinetic curves for the feature maximum probability, comparing a correct and a faulty measurement.

Step 4.: The residuals of $k^i(n)$ and a reference kinetic curve for the features, $k^i_{\text{ref}}(n)$ are calculated

$$\rho^i(n) = k^i(n) - k^i_{\text{ref}}(n). \tag{3.15}$$

The reference kinetic curve is obtained using the real data. This is done by calculating

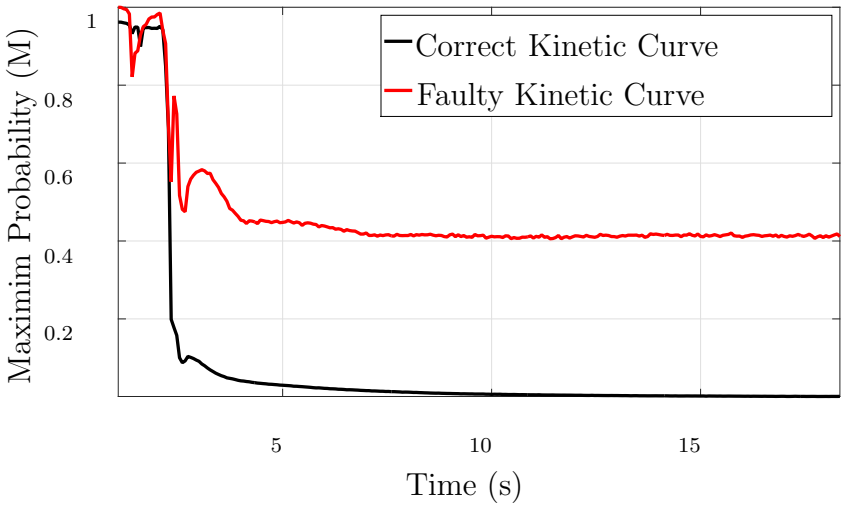


Figure 3.8: Comparison of the feature kinetik curves $k^M(n)$ for a correct and a faulty measurement w.r.t. the feature maximum probability.

the feature kinetic curves of all the correct measurements and taking their median to be the reference curve

$$k^i_{\text{ref}}(n) = \underset{m \in \mathcal{M}^i_C}{\text{median}}\{k^i_m(n)\},$$

where $\mathcal{M}^i_C$ is a set containing all feature kinetic curves of correct measurements of the corresponding features. As the real data set contains measurements of the lowest and highest physically possible glucose concentrations as well as many in between, the reference curve is representative.

The residuals $\rho^i(n)$ are assumed to be i.i.d.. They can be approximated by a random process $\boldsymbol{\rho}^i$ drawn from a Gaussian distribution according to $f_{\boldsymbol{\rho}^i} \sim \mathcal{N}(0, (\sigma^i_{\text{C}})^2)\ \forall\ n \in \mathbb{R}$, where σ^i_{C} is unknown and has to be estimated using the given data. For the case of faulty measurements the residuals follow a Gaussian distribution according to $f_{\boldsymbol{\rho}^i} \sim \mathcal{N}(\mu^i_{\text{F}}, (\sigma^i_{\text{F}})^2)\ \forall\ n \in \mathbb{R}$, where $\mu^i_{\text{F}} > 0$ and σ^i_{F} is unknown. Fig 3.9 shows an example of the distribution of residuals for correct measurements and faulty measurements for the feature (M).

Step 5.: A hypothesis test is performed, where

$$\begin{aligned} \mathcal{H}_0 &: f_{\boldsymbol{\rho}^i} \sim \mathcal{N}(0, (\sigma^i_{\text{C}})^2), \quad \text{correct measurement} \\ \mathcal{H}_1 &: f_{\boldsymbol{\rho}^i} \sim \mathcal{N}(\mu^i_{\text{F}}, (\sigma^i_{\text{F}})^2), \quad \text{faulty measurement.} \end{aligned} \tag{3.16}$$

Here, $\mu_{\mathrm{F}}^i > 0$, and σ_{C}^i and σ_{F}^i are both unknown, but assumed to be equal $\hat{\sigma}^i = \sigma_{\mathrm{C}}^i = \sigma_{\mathrm{F}}^i$

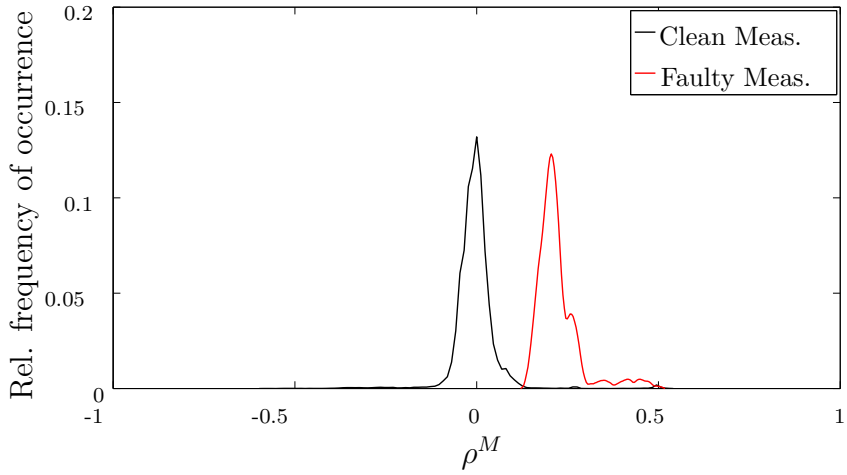


Figure 3.9: Relative frequency of occurrence of the residuals of clean and faulty measurements.

and can be estimated from the data at hand. Given the i.i.d. assumption, a Neyman-Pearson approach is used for the hypothesis testing that results in the following test statistic

$$T(\rho^i(n)) = \frac{1}{N}\sum_{n=1}^{N}\rho^i(n) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \underbrace{\frac{\ln \delta_{\mathrm{GLCM}}^i}{\mu_{\mathrm{F}}^i N}(\hat{\sigma}^i)^2 + \frac{\mu_{\mathrm{F}}^i}{2}}_{\tilde{\delta}_{\mathrm{GLCM}}^i}, \tag{3.17}$$

where δ_{GLCM}^i is determined using a nominal value for the probability of false alarm

$$\begin{aligned} P_{\mathrm{FA}} &= \int_{\tilde{\delta}_{\mathrm{GLCM}}^i}^{\infty} P(\rho^i(n)|\mathcal{H}_0) \\ &= Q\left(\frac{\tilde{\delta}_{\mathrm{GLCM}}^i}{\sqrt{(\hat{\sigma}^i)^2/N}}\right). \end{aligned} \tag{3.18}$$

Here, $Q(\cdot)$ is given by Eq. (3.9). Finally, the threshold $\tilde{\delta}_{\mathrm{GLCM}}^i$ is calculated as

$$\tilde{\delta}_{\mathrm{GLCM}}^i = Q^{-1}(P_{\mathrm{FA}}) \cdot \sqrt{(\hat{\sigma}^i)^2/N}. \tag{3.19}$$

If the null hypothesis is rejected, the measurement is assumed to be faulty. The measurement procedure is broken off and the user is given an error message and is asked to repeat the measurement. As long as the null hypothesis is accepted, the test is performed for each new incoming sample, to be able to cover the cases where the fault happens during the measurement. This can occur for example, when due to a strong movement of the device, the chemical test strip shifts and is not under the observation area of the camera.

3.2.4 Image Segmentation and Relative Remission Estimation

The underlying assumption on the glucose images is that they contain a ROI where the blood sample is distributed. The colour change of the ROI represents the reflectivity caused by the underlying glucose concentration. This area is not necessarily completely spatially connected as it may contain artefacts where the reaction has not taken place, or granularities in the chemical that do not contribute to a proper reaction. Other regions in the image do not correspond to the ROI, either, because the blood sample is unevenly distributed in these areas, or because the areas are dry, i.e., not covered with blood. The regions in the image are not characterised by clear edges, as is usual in typical image segmentation problems. The different regions, in this problem, are interpreted as clusters with specific cluster centres overlaid by noise. This way a pixel-wise segmentation can be performed. The data points converging to a certain cluster center are assumed to belong to the same region. The noise contaminating the images is assumed to be Gaussian. This assumption is based on the calibration frames where no reaction has started. Two different clustering approaches are proposed in this thesis and will be presented in Chap. 4 and 5. They result in an estimate of the underlying number of clusters and a set of labels defining the cluster membership of the pixels.

The segmentation results in a finite number of cluster centres corresponding to the different image regions. The cluster center corresponding to the ROI is identified and its relative remission value is tracked until saturation of the chemical reaction. This will be explained in more detail in subsequent chapters. For the time being, the assumption is made that a segmentation has taken place and the region of interest has been identified, providing knowledge of the relative remission estimate over time $\hat{r}(n)$.

3.2.5 Kinetic Filtering and Convergence Detection Using State Estimation Techniques

As portrayed in Sec 3.1, the standard procedure for saturation detection is detecting the point at which the slope of the kinetic curve maintains a certain threshold δ_{slope} over a pre-defined time interval [AHKP11]. Formally, this criterion is described as

$$\begin{aligned} r'(n) &\triangleq \frac{dr(n)}{dn} = \frac{r(n+1) - r(n)}{\Delta n} > \delta_{\text{slope}} \\ r''(n) &\triangleq \frac{d^2r(n)}{dn^2} = \frac{r'(n+1) - r'(n)}{\Delta n} \approx 0, \end{aligned} \tag{3.20}$$

where $r(n)$ describes the relative remission value at frame n and $\Delta n = 1$. The threshold is typically set to $\delta_{\text{slope}} = -0.015$ for the setup used in this thesis. This point is reached after around 3 s to 15 s, depending on the underlying measurement. The assumption made, hereby, is that in the ideal case, measurements of equal glucose concentration will follow a similar progression at all times. Therefore, the relative remission estimate at the predefined slope threshold will be related to the underlying glucose concentration. The mapping function between intensity and glucose can be adjusted to account for the inaccuracy. While this can lead to satisfactory results when no information is given on the kinetic curve, it is shown in this thesis that this criterion can also result in erroneous, premature estimates of the converged relative remission $\hat{r}_C$.

To improve the performance of this approach drastically, the incorporation of a model of the kinetic behaviour of the chemical reaction [Lei09] is proposed. This model is used in a state estimation framework to predict the actual convergence value r_C ahead of time, ensuring a faster and more accurate measurement. Hereby, stages 2) and 3) of the kinetic curve, as depicted in Fig. 2.4, are modelled. Stage 2) is the moistening period where the chemical reagent is hydrated and the reaction begins. Stage 3) constitutes the saturation phase where convergence is reached.

The glucose oxidase taking place on the chemical test strip can be modelled by the differential equation [Lei09]

$$r(n) = (r_D - r_C) \cdot e^{-n\tau} + r_C + v(n), \quad n > n_D, \tag{3.21}$$

where $r(n)$ is the remission measurement value at time n, r_D is the initial remission value after the drop, and r_C is the convergence value of the relative remission after the chemical reaction has saturated. The saturated remission value r_C is defined as the state, that is to be predicted. Furthermore, τ is the reaction rate, and $v(n)$ is a zero-mean, white, Gaussian noise process describing the measurement noise. It is distributed according to $v(n) \sim \mathcal{N}(0, \sigma_v^2)\ \forall\ n \in \mathbb{R}$.

Due to the nature of the chemical reaction and based on the Michaelis-Menten model [EK88, Bis08] explained in Sec. 2.4.1, it is realistic to assume that τ and r_C are correlated. Building up on [Lei09], a regression analysis using a real data set of estimated convergence values $\hat{\mathbf{r}}_C = [\hat{r}_{C,1}, ..., \hat{r}_{C,N_M}]$ and corresponding rates $\hat{\boldsymbol{\tau}} = [\hat{\tau}_1, ..., \hat{\tau}_{N_M}]$ is performed. Here, N_M is the number of available measurements in the data set. Hereby, a linear relation between τ and r_C is established to be the most suitable least-squares (LS) fit [DZ16], as can be seen in Fig. 3.10

$$\hat{\boldsymbol{\tau}} = \Delta\tau \cdot \hat{\mathbf{r}}_C + \tau_0, \tag{3.22}$$

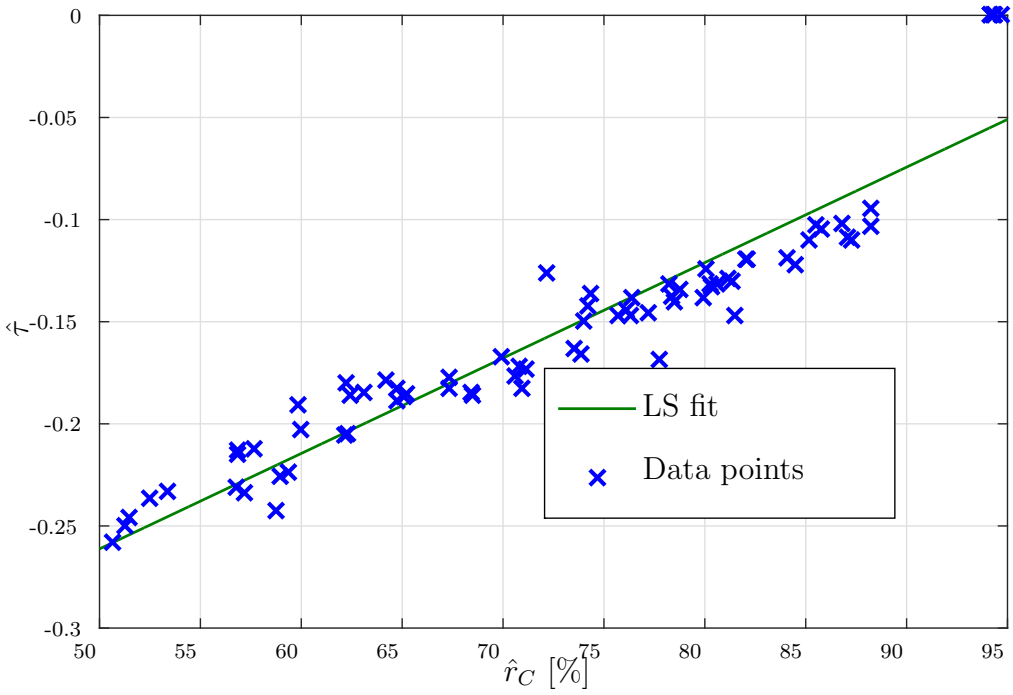


Figure 3.10: Regression analysis for the relation between the estimated reaction rate $\hat{\tau}$ and the convergence value of the relative remission $\hat{r}_C$.

where $\Delta\tau > 0$ and $\tau_0 < 0$. Using a LS fit, the regression parameters $\Delta\tau$ and τ_0 are estimated, obtaining from Eq. (3.21) the following nonlinear relation for the kinetic curve

$$\begin{aligned} r(n) &= (r_D - r_C) \cdot e^{-n\cdot(\Delta\tau \cdot r_C + \tau_0)} + r_C + v(n) \\ &= f_{\text{Kin}}(r_C, v(n)) \ , \quad n > n_D. \end{aligned} \tag{3.23}$$

The Extended Kalman Filter Approach

An Extended Kalman Filter (EKF) [Sim02, WB95] is proposed to process the measured kinetic curve [DZ16] and perform an online prediction of the remission convergence value r_C. The advantage of using the EKF, is that it is able to deal with the non-linear nature of the observation model given in Eq. (3.23). Hereby, the saturation relative remission value r_C is assumed to be the underlying static state of the system. The observations are given by the incoming relative remission estimates $\hat{r}(n)$. The relative remission estimate $\hat{r}(n)$ is given as the cluster center of the ROI segmented from the image $\mathbf{I}^{(n)}$ at frame n. To this end, the prediction and measurement equations are

formulated as follows

$$\hat{r}_C(n) = a(n)\hat{r}_C(n-1) + w(n), \tag{3.24}$$

$$\hat{r}(n) = h(\hat{r}_C(n), v(n)) \ , \tag{3.25}$$

where $a(n)$ is the transition matrix describing the transition of the state estimate $\hat{r}_C$ from time $n-1$ to n. In this case, $a(n) = 1$, as r_C is a static state of the chemical reaction. The process noise is modelled by the random variable $w(n)$ and can be used to account for uncertainty in the model. It is assumed that $w(n)$ is an i.i.d. additive zero-mean Gaussian noise component distributed according to $w(n) \sim \mathcal{N}(0, \sigma_w^2)\ \forall\ n \in \mathbb{R}$. $h(\hat{r}_C(n), v(n))$ describes the observation model, relating the estimated state $\hat{r}_C(n)$ to the incoming measurement $\hat{r}(n)$. It is given by the partial derivative of the process model w.r.t. the state $\hat{r}_C$

$$\begin{aligned} h(\hat{r}_C(n), v(n)) \ &= \frac{\partial f(\hat{r}_C(n), V(n))}{\partial \hat{r}_C(n)} \\ &= 1 - e^{-n\cdot(\Delta\tau\cdot\hat{r}_C(n)+\tau_0)} \\ &\quad \times (1 + r_D \cdot \Delta\tau \cdot n - \hat{r}_C(n) \cdot \Delta\tau \cdot n)\,. \end{aligned} \tag{3.26}$$

An alternation between the prediction and the correction step of the EKF, including a new measurement $\hat{r}(n)$ in each iteration is carried out. The algorithm for the prediction of $\hat{r}_C$ is presented in the sequel, taking as an input the relative remission estimates $\hat{r}(n), n = 1, \ldots, N$.

- **Step 1.** Initialise the algorithm: $\hat{r}_C(0) = \hat{r}(n_D), P(0), \sigma_v^2, \sigma_w^2$
- **Step 2.** While $\hat{r}_C(n) - \hat{r}_C(n-1) \leq \epsilon$
  - **Prediction Step**
    - a. $\hat{r}_C^-(n) = \hat{r}_C(n-1)$
    - b. $P^-(n) = P(n-1) + \sigma_w^2$
  - **Correction Step**
    - c. $\hat{r}_C(n) = \hat{r}_C^-(n) + K \cdot (\hat{r}(n) - f_{\text{Kin}}(\hat{r}_C(n)))$
    - d. $P(n) = (1 - K \cdot h(\hat{r}_C(n))) \cdot P^-(n)$
- **Step 3.** Stop when the algorithm converges, i.e. $\hat{r}_C(n) - \hat{r}_C(n-1) \leq \epsilon$, obtaining the estimated convergence value $\hat{r}_C$ and the time of convergence n_c.

The Particle Filter Approach

An alternative to the EKF is solving the prediction problem using a particle filter approach [SDZ16, Sei15], where the image statistics are tracked over time. The

advantage of particle filters is that they are able to deal with the multi-modal image statistics. The approach used in this work, is termed region-based particle filter (RBPF), as it assumes two image regions, one corresponding to the background and one to the foreground. Two parallel, identical particle filters that operate on the two assumed image regions are designed to solve the state estimation problem. The image regions are chosen such that one region is most likely to include the ROI, and the other region most probably does not contain ROI pixels. Overlapping regions are chosen to account for uncertainty about the exact location of the ROI in the captured images.

As in the EKF case, the static state of the system is assumed to be r_C. A dynamic is artificially introduced to this static state, making it follow a first order Markov process, where the current state only depends on the previous state given a random walk dynamic with small variance. This application-driven solution, formally known as artificial evolution of parameters [LW01, GSS93], eases the need of perfect initialisation of the particle filter and decreases the problem of sample degeneracy as potentially more particles obtain a significant weight in the update stage. The system model is, thus, given by

$$\hat{r}_C(n) = \hat{r}_C(n-1) + w(n), \tag{3.27}$$

where $w(n)$ is an i.i.d. additive zero-mean Gaussian noise component distributed according to $w(n) \sim \mathcal{N}(0, \sigma_w^2)\ \forall\ n \in \mathbb{R}$. Unlike in the EKF formulation, the observations in this case are given by the pixel intensities of the image $\mathbf{I}^{(n)}$. The observation model relates the observations $\mathbf{I}^{(n)}(m_x, m_y)$ at frame n to the state $\hat{r}_C(n)$ by

$$\mathbf{I}^{(n)}(m_x, m_y) = (r_D - \hat{r}_C(n)) \cdot \mathrm{e}^{-n\cdot(\Delta\tau\cdot\hat{r}_C(n)+\tau_0)} + \hat{r}_C(n) \ \ \forall\ (m_x, m_y) \in M_x \times M_y. \tag{3.28}$$

The filtering distribution is approximated by

$$p(\hat{r}_C^i(n)|\mathbf{I}^{(n)}) \approx \sum_{i=1}^{N_p} w^i(n)\delta\big(\hat{r}_C(n) - \hat{r}_C^i(n)\big)\ , \tag{3.29}$$

where N_p is the number of particles, $\delta(\cdot)$ denotes the Dirac delta function and the weights are normalised such that $\sum_{i=1}^{N_p} w^i(n) = 1$. The system model in Eq. (3.27) is used to formulate the prior as

$$p(\hat{r}_C^i(n)|\hat{r}_C^i(n-1)) = \mathcal{N}(\hat{r}_C^i(n); \hat{r}_C^i(n-1), \sigma_w^2), \tag{3.30}$$

which is used as the proposal distribution such that the weights of the particle filters can be calculated as

$$w^i(n) \propto w^i(n-1)\ p(\mathbf{I}^{(n)}|\hat{r}_C^i(n)). \tag{3.31}$$

The likelihood $p(\mathbf{I}^{(n)}|\hat{r}_C^i(n))$ is found by evaluating a kernel density estimate $\hat{f}$ of the pdf of observed pixel intensities contained in $\mathbf{I}^{(n)}$, which is obtained using Eq. (3.28)such that

$$p(\mathbf{I}^{(n)}|\hat{r}_C^i(n)) = \hat{f}_h(\mathbf{x}^{(n)}), \tag{3.32}$$

where $\mathbf{x}^{(n)}$ is the vectorised image at frame n and a kernel density estimator $\hat{f}_h(\cdot)$ based on a Gaussian kernel function with bandwidth parameter h is used. Here, h is set according to Silverman's rule of thumb [Sil86]. Chap. 4 will discuss the effect of h in more detail.

The minimum mean squared error (MMSE) state estimate is given by

$$\hat{r}_C(n) = \sum_{i=1}^{N_p} w^i(n)\hat{r}_C^i(n). \tag{3.33}$$

In order to mitigate sample degeneracy, i.e. all but a few samples will have negligible weight after a few iterations, the particles are resampled at each time step [AMGC02].

The region-based particle filter is summarised in the sequel, taking as an input the relative remission estimates $\hat{r}(n), n = 1, \ldots, N$.

- **Step 1.** Initialise the algorithm: $\hat{r}_C(0) = \hat{r}(n_D), N_p, p(\hat{r}_C^{(j)}(0)|\hat{r}_C^{(j)}(-1)), \sigma_w^2$.
- **Step 2.** While $\hat{r}_C(n) - \hat{r}_C(n-1) \leq \epsilon$
  - **Prediction Step**
    - a. Propagate weighted particle set $\{(\hat{r}_C^i(n-1), w^i(n-1)\}_{i=1}^{N_p}$ to approximate filtering distribution as in Eq. (3.29).
    - b. Draw N_p new particles from the proposal distribution given by Eq. (3.30).
  - **Update Step**
    - c. Receive new measurement $\mathbf{I}^{(n)}$.
    - d. Calculate importance weights as in Eq. (3.31) using the likelihood in Eq. (3.32).
    - e. Calculate the state estimate as in Eq. (3.33).
    - f. Perform resampling with replacement with probability $w^i(n)$.
- **Step 3.** Stop when the algorithm converges, i.e. $\hat{r}_C(n) - \hat{r}_C(n-1) \leq \epsilon$, obtaining the estimated convergence value $\hat{r}_C$ and the time of convergence n_c.

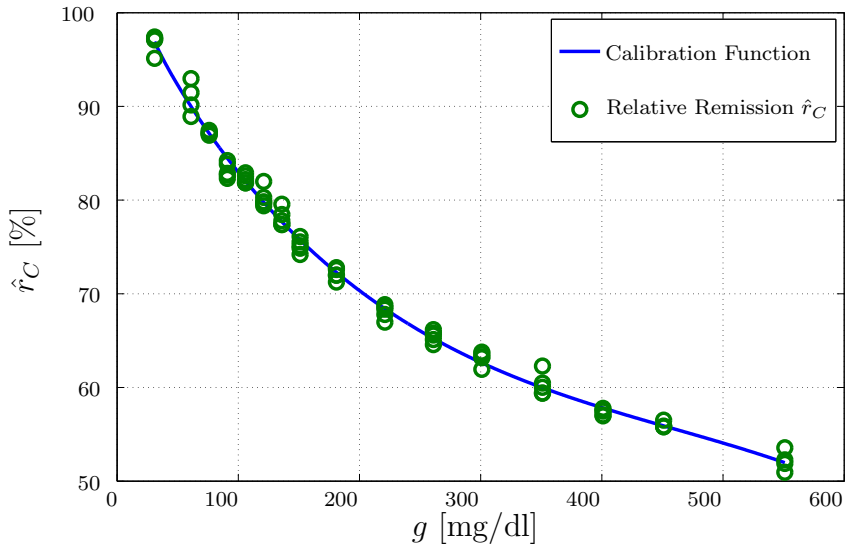


Figure 3.11: Example of a calibration curve generated using real glucose measurements. The relative remission estimates $\hat{r}_C$ are plotted against the true glucose concentration values g.

Mapping to Glucose Concentration

Finally, the estimated convergence relative remission value $\hat{r}_C$ is mapped to the underlying glucose concentration, which is delivered to the user. To this end, a calibration function $f_{\text{Calib}} : \hat{r}_C \rightarrow \hat{g}$ is coded into the glucometer to perform the mapping during measurement. The inverse function of the calibration function is generated *a priori* by means of photometric measurements performed under lab conditions using blood samples with known underlying glucose concentrations. To ensure both accuracy and precision, different glucose concentrations covering the possible range of blood glucose levels are needed, as well as several samples of the same glucose concentration. The inverse calibration function is given by

$$\hat{r}_C(g) = a_1 + a_2 \cdot g + a_3 \cdot g^{1.5} + a_4 \cdot g^{-5} + a_5 \cdot e^{a_6 \cdot g}, \tag{3.34}$$

where $a_i, i = 1, \ldots, 6$ are the parameters of the model. They are determined for each specific batch of chemical test strips produced using a robust non-linear least squares fit. An example of such a mapping function is illustrated in Fig. 3.11. It is noteworthy, that the mapping function typically becomes flatter for high glucose concentrations, hereby allowing a better distinction of nearby glucose concentrations in the lower region.

3.3 Experimental Results and Discussion

In this section the different blocks of the framework are validated using the real data set from Table 2.1. First, the effect of binning is evaluated and an optimal binning value is found for the provided real data. Next the detection of faulty glucose measurements using the GLCM-based texture analysis approach is assessed. An algorithm for sequential detection of the faulty frames is presented to speed up the detection and avoid unnecessary computation. Finally, the state estimation approaches, EKF and RBPF, are evaluated both w.r.t. to measurement time and quality of estimates.

3.3.1 Binning Effect in Pre-processing

For testing the effect of binning, and choosing a suitable binning value B, a synthetic and a real data set are used. The aim of using a synthetic data set, is to validate the performance with an available ground truth. It is developed by creating a mask from the real data set using a frame at $n = n_D$. The area of the test strip is modulated with the model of the chemical reaction given in Eq. (3.23), using $N_g = 20$ different glucose concentrations, resulting in synthetic measurement videos of the same size as the real data videos. Additive zero-mean Gaussian noise with a variance σ^2_{syn} is estimated from the real data samples.

To validate data binning on real data, Set A from Table 2.1 is used. It contains $N_M = 48$ measurements of $N_g = 5$ different glucose concentrations. The resolution of the frames used in this Set is $\Upsilon = 6.45\,\mu\text{m/Pixel}$ and the blood samples have a volume of 10 nl to 100 nl.

Image segmentation is performed using the mean-shift algorithm presented in Chap. 4. This approach identifies the ROI and simultaneously estimates its relative remission estimate $\hat{r}(n)$ over time. The convergence of the chemical kinetic is performed using the standard convergence criterion given in Eq. (3.20). The converged relative remission value is $\hat{r}_C = \hat{r}(n_C)$, i.e. the relative remission estimate at the detected convergence time instant.

The results of the synthetic data set are evaluated in terms of the averaged mean-squared error (MSE) of the relative remission estimates over $N_{\text{run}} = 100$ runs per glucose value. To assess the binning effect on the real data, the coefficient of variation $\text{CV}_{\hat{r}}$, given in Eq. (2.2), is computed. Table 3.1 presents the results of

binning using binning values between $B = 1$ to 6. Clearly, the estimation performance improves with increasing B. The reason for this, is that binning acts as smoothing filter that reduces noise in the frames. Furthermore, binning reduces the computation time due to the reduced frame size. We identify the best binning value to be $B = 5$. For $B > 5$, the image structure is lost, which leads to a degraded performance.

Table 3.1: Evaluation of different binning values B for both the synthetic and real data.

B	1	2	3	4	5	6
$\mathrm{MSE}_{\mathrm{syn}}$	3.12	0.51	0.04	0.08	0.02	0.05
$\mathrm{CV}_{\hat{r}}$	16.2	3.73	2.03	1.62	0.96	1.19

These results motivated the use of a camera with a lower resolution in the framework presented in Fig. 2.3. The real data sets D-F given in Table 2.1 are obtained using the low resolution setup. For the rest of this work, measurements with a resolution of $\Upsilon = 6.45\,\mu\mathrm{m/Pixel}$ are always be binned with a binning factor of $B = 5$ in the pre-processing step, while measurements with a resolution of $\Upsilon = 30\,\mu\mathrm{m/Pixel}$ are binned with a binning factor of $B = 1$, i.e. will not be binned.

3.3.2 Faulty Measurement Detection

Section 3.2.3 introduces an approach to detect faulty measurements using GLCM-based texture analysis. To validate the performance of this proposed approach, Sets B and C from Table 2.1 are deployed. They contain $N_T = 162$ measurements of $N_\gamma = 16$ different underlying glucose values. All in all, three of these measurements are faulty.

The approach is validated, first, using the batch processing version of the algorithm described in Sec. 3.2.3. Two different nominal probabilities of false alarm P_{FA} are compared w.r.t the actually obtained false alarm and correct detection rates. The results are presented in Table 3.2. Observe that faulty frames are always correctly detected. For a nominal $P_{\mathrm{FA}} = 10^{-3}$ false alarms occur for very low glucose value measurements. They are confused with faulty measurements as the texture is quite similar. The feature (M) shows the best performance w.r.t to achieving the nominal P_{FA}. This is due to the fact that under $\mathcal{H}_0$ the approximation of the residuals $\rho^i(n)$ by $f_{\boldsymbol{\rho}^i} \sim \mathcal{N}(0, (\sigma_1^i)^2)$ is most accurate.

Table 3.2: Percentage of false alarms and correct detections for the features (E), (M) and (C) for the nominal false alarm rates $P_{\mathrm{FA}} = 10^{-3}$ and $P_{\mathrm{FA}} = 10^{-4}$.

P_{FA}	Feature	False alarms rate	Correct detection rate
10^{-3}	E	1.8%	100%
	M	0.63%	100%
	C	2.5%	100%
10^{-4}	E	0%	100%
	M	0%	100%
	C	0%	100%

As long as the reaction has not yet taken place ($n < n_D$), the features behave similarly for all measurements irrespective of them being faulty or not. Once the moistening period starts, the feature values quickly reach their final convergence value. In this experiment the test statistic is computed sequentially for each new incoming frame $\mathbf{I}^{(n)}, n = 1, 2, \ldots$ as follows

- **Step 1.** Calculate GLCMs $\varrho(\gamma_i, \gamma_j; d_\gamma, \boldsymbol{\alpha}_\gamma, L)$ for each incoming frame.
- **Step 2.** Build feature vector $\mathbf{s}$ as in Eq. (3.14).
- **Step 3.** Calculate the feature kinetic curves $k^i(n)$.
- **Step 4.** Calculate the residual kinetic $\rho^i(n)$ curve as in Eq. (3.15).
- **Step 5.** Perform the hypothesis test as in Eq. (3.35).

The output of this approach is given by the decision if the measurement is faulty or not and the time index n_F, when it has been detected as faulty. Following test statistic is used for the sequential test

$$T_j(r^i(n)) = \frac{1}{n_{\mathrm{F}}} \sum_{n=1}^{n_{\mathrm{F}}} \rho^i(n) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \delta^i_{\mathrm{GLCM}}, \tag{3.35}$$

where $n_{\mathrm{F}} < N$ is the frame at which the measurement is detected to be faulty. δ^i_{GLCM} is calculated using a nominal $P_{\mathrm{FA}} = 10^{-4}$. The frame is detected at which the measurement is detected as faulty. This simulation is performed for three faulty measurements contained in the data set, which are denoted by M_1, M_2, M_3. The correct measurements in the data set are all detected as being correct, as shown in the previous section. The results are listed in Table 3.3, as well as the estimated convergence times $\hat{n}_C$ using the convergence criterion in Eq. (3.20).

Table 3.3: Detection times for the three faulty measurements in the real data set.

	n_{F}^{E}	n_{F}^{M}	n_{F}^{C}	$\hat{n}_C$
M_1	1.067 s	1.067 s	0.76 s	3.33 s
M_2	0.86 s	0.967 s	0.76 s	3.0 s
M_3	1.4 s	1.3 s	0.70 s	5.33 s

Clearly, using the proposed method faulty frames are detected long before the convergence times $\hat{n}_C$. This means that the user does not have to wait until the device has calculated the estimate of the underlying glucose value to be told that the measurement is false, but gets an error message after around 1 s and is asked to repeat the measurement. It is also noteworthy that feature (C) converges the fastest among the examined features, with a convergence time between 0.7 s to 0.76 s for each of the tested samples.

The key point in this experiment is the convergence time of the features. For the faulty measurements, it has been proven that they converge after the values given in Table 3.3. To circumvent the small sample size of faulty measurements and prove that this result holds for larger sample sizes, the convergence times of the features for correct measurements are detected. The results of this study are feature convergence times between 0.8 s to 1.5 s. Hence, for any kind of measurement a steady feature value will be attained closely after n_D and the measurement can be classified as faulty or not.

3.3.3 Convergence Detection using the EKF approach

The convergence of the chemical reaction is typically detected using the criterion in Eq. (3.20). In this thesis, state estimation approaches are proposed to this end leading to faster and more accurate glucose concentration estimates. This experiment evaluates the EKF-based approach proposed in Sec. 3.2.5. The kinetic curves are calculated according to the proposed framework. The segmentation algorithm used is the mean-shift approach which will be explained in further detail in Chap. 4. The estimated kinetic curves $\hat{r}(n), n = 1, \ldots, N$ are fed to the EKF, and the output is given by the estimated convergence time $\hat{n}_C$ and the estimated converged relative remission value $\hat{r}_C$, as illustrated in Sec. 3.2.5. Table 3.4 lists the parameter values used for this experiment and Table 3.5 presents the calculated regression parameters from Eq. (3.22) for the different data sets.

Table 3.4: Simulation parameters of the EKF approach.

$\hat{r}_C(0)$	$P(0)$	σ_v^2	σ_w^2	ϵ
$\hat{r}(n_D)$	1	10	9.6	10^{-3}

Table 3.5: Regression parameters $\Delta\tau, \tau_0$ for the different real data sets.

Set	A	B	C	D	E	F
$\Delta\tau$	0.0031	0.0031	0.0033	0.0029	0.003	0.003
τ_0	0.5	0.48	0.46	0.5	0.38	0.43

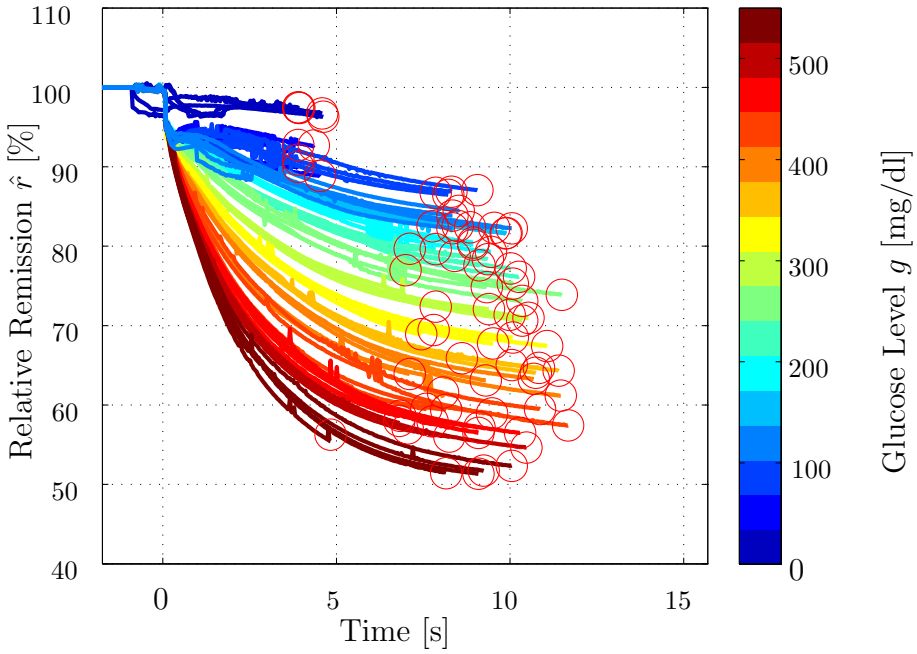


Figure 3.12: Kinetic curves for measurements of Set C using the mean-shift approach discussed in Chap. 4. The red circles signify convergence points $\hat{r}_C^{(\text{Stand. Conv})}$.

Figure 3.12 depicts estimated kinetic curves over time of the measurement Set C from Table 2.1. Evidently, convergence of same-glucose-level measurements is not always reached at the same time which results in strongly varying remission estimates, although the course of the kinetic curve is very similar.

Taking a closer look at the 550 mg/dl measurements in said figure, this behaviour is very pronounced. Convergence is found at times between $\hat{t}_C \approx 4\,\text{s}$ and $\hat{t}_C \approx 9\,\text{s}$, leading

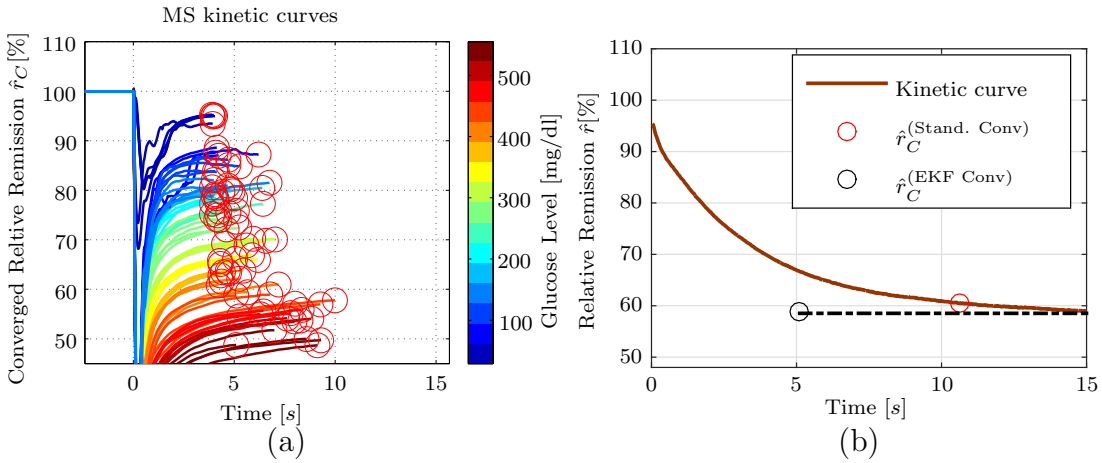


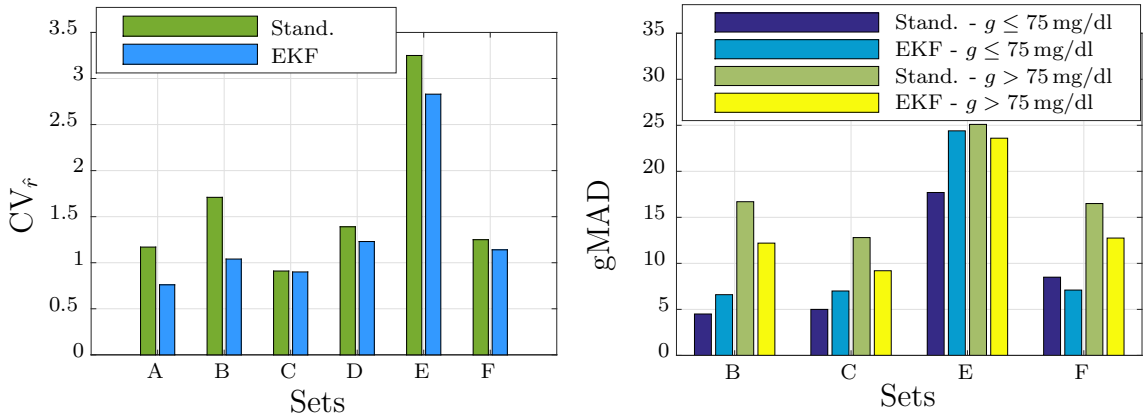
Figure 3.13: (a) State estimation of r_C using the EKF. The red circles indicate the estimated convergence values. (b) A comparison of the convergence estimates using the standard convergence criterion and the EKF.

to remission estimates from $\hat{r}_C \approx 50\%$ to $\hat{r}_C \approx 59\%$. Figure 3.13 (a) shows the EKF state estimates $\hat{r}_C^{(\text{EKF Conv})}$ for this example. Evidently, the EKF converges to reliable state estimates quickly and therefore, the resulting $\hat{r}_C$ estimates not only lie in the same range, but match the final convergence value of the kinetic behaviour more accurately. This can be seen clearly in Fig. 3.13 (b). The convergence value reached by the standard method $\hat{r}_C^{(\text{Stand. Conv})}$ occurs after around 11 s and does not match the actual saturation value. The convergence value reached by EKF $\hat{r}_C^{(\text{EKF Conv})}$, however, occurs after 5 s and is much closer to the actual saturation value. Hence, it can be asserted, that the EKF-based convergence detection leads to both a gain in accuracy and time.

Table 3.6: Gain in time and reduced error obtained using EKF.

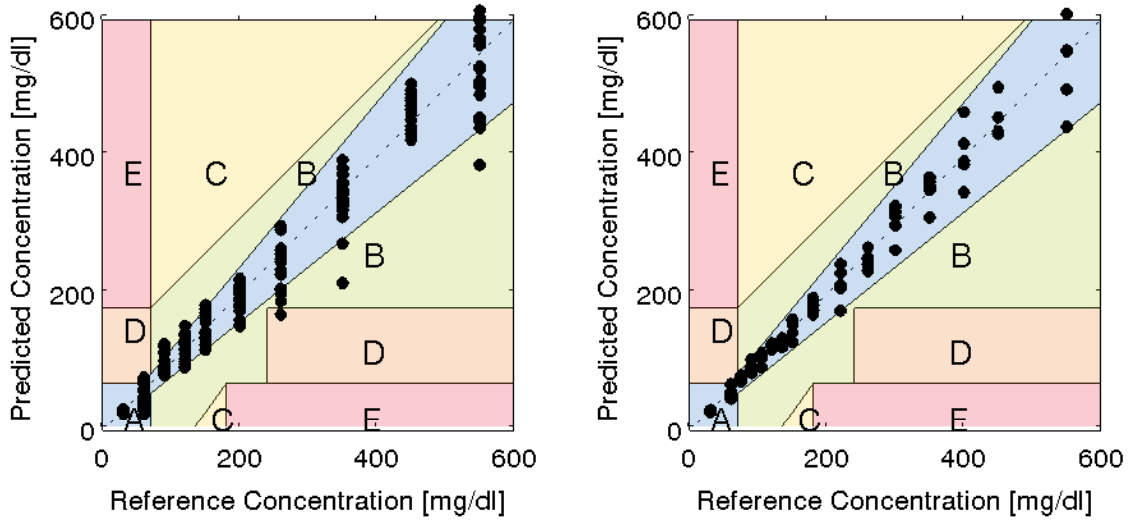
	Gain in time	Error Benefit
Low Glucose Range ($g \leq 75$ mg/dl)	1.41 s	2.23 %
High Glucose Range ($g > 75$ mg/dl)	0.97 s	2.1 %

Table 3.6 shows the gain in time obtained using the EKF. These results are averaged over all measuremeent sets in the low and high glucose range, respectively. The EKF enables a gain in time of about 1 s to 1.5 s, which corresponds to saving computation of 30 frames to 45 frames. Comparing the remission estimates $\hat{r}_C^{(\text{Stand. Conv})}$ and $\hat{r}_C^{(\text{EKF Conv})}$ to the saturation values of the chemical reaction $r_C(N)$, it can be asserted that the EKF produces more accurate results, with a gain of more than 2% in remission.



(a) (b)

Figure 3.14: (a) Comparison of the $CV_{\hat{r}}$ values for the standard convergence criterion to the EKF. (b) Comparison of the gMAD values for the standard convergence criterion to the EKF, separated into the low ($g \leq 75\,\mathrm{mg/dl}$) and high ($g > 75\,\mathrm{mg/dl}$) glucose range, respectively.



(a) (b)

Figure 3.15: Clarke's Error Grid Analysis of Set E for (a) the standard and (b) the EKF convergence criteria.

Figure 3.14 presents a comparison of $CV_{\hat{r}}$ values and gMAD values for the standard convergence criterion and the EKF convergence method. Note, that these validation criteria were presented in Sec. 2.4.3. The EKF convergence always leads to improved $CV_{\hat{r}}$ values. Similarly, for the high glucose range the gMAD results show the same outcome. For low glucose ranges the standard convergence compares favourably. The

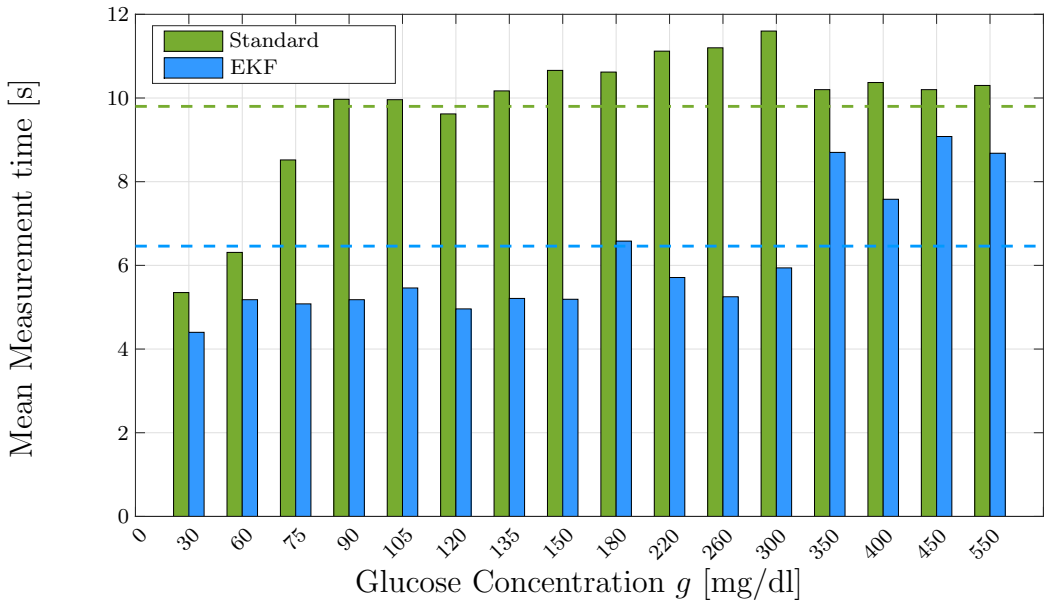


Figure 3.16: Comparison of the measurement times resulting from the standard convergence criterion and the EKF convergence criterion of Set C for different underlying glucose concentrations. The dashed lines indicate the overall mean convergence times of the corresponding algorithms.

reason for this is that the EKF method is quite sensitive to the model used. As can be seen in Fig. 4.16, the kinetic curves of low glucose measurements exhibit a dip after the drop followed by a steady rise. This is not embodied in the model in Eq. (3.23).

The Clarke Error Grid plot of Set E in Fig. 3.15 (a) shows that 93 % of the points lie in region A, 6 % lie in region B and 1% lie in region D using the standard convergence criterion. Thereby, the results do not conform with the requirements given by the most recent ISO standards, as presented in Sec. 2.4.3. As indicated by the results in Fig. 3.15 (b), using EKF state estimation to predict the convergence values leads to a significant improvement of the results. Now 100% of the estimates of Set E lie in the A-region.

Finally, Fig. 3.16 presents the measurement times obtained for Set C detailed according to the different underlying glucose concentrations. The figure sheds light on the superiority of the EKF approach w.r.t to measurement time, reducing the mean over all measurements from around 9.8 s to around 6.5 s. The largest gain in measurement time is achieved for the mid-level glucose concentrations between 100 mg/dl to 300 mg/dl.

Convergence Detection using the RBPF approach

The region-based particle filter approach is evaluated here using Set C from Table 2.1. The parameters of the experiment are given in Table 3.7. The RBPF converges as soon as the changes in both state estimates are sufficiently small. In this way the measurement time is determined.

Table 3.7: Simulation parameters of the RBPF approach.

$\hat{r}_C^{\text{Foreground}}(0)$	$\hat{r}_C^{\text{Background}}(0)$	N_p	$p(\hat{r}_C^{(j)}(0)\|\hat{r}_C^{(j)}(-1))$	σ_w^2	ϵ
$\hat{r}(n_D)$	100	500	$\mathcal{N}(\hat{r}_C(0), 2)$	3	10^{-3}

Using the RBPF for image-based photometric glucose measurement, the average testing time is reduced by approximately 50% at a comparable accuracy and precision of results as for the state-of-the-art method given by Eq. (3.20). Fig. 3.17 reveals the average measurement times for different underlying glucose concentrations. On average, the RBPF predicts the required image statistics by 4.8 s compared to 9.8 s for the reference. For specific measurement groups the testing speed is even increased by 65%.

The quality of glucose measurements is again assessed by $\text{CV}_{\hat{r}}$ and gMAD. The corresponding averaged values, which should both be small, are given in Table 3.8. The overall quality of results is slightly worse than the reference method. Table 3.8(b) shows that only for low glucose concentrations, the RBPF leads to clinically more accurate results compared to the reference. This is an important case, as in these cases hypoglycemia is present and overestimation of the glucose concentration can be fatal to the patient.

Additionally, the MATLAB simulations reveal that on average the proposed algorithm requires less than a third of the computational complexity of the reference.

Table 3.8: Comparison of glucose-specific evaluation methods for the standard convergence criterion and the RBPF for different underlying glucose concentrations C in mg/dl.

	Standard	RBPF
$g \leq 100\,\text{mg/dl}$	0.46	0.75
$g > 100\,\text{mg/dl}$	1.00	1.74
Overall	**0.91**	**1.27**

(a) Coefficient of Variation $\text{CV}_{\hat{r}}$.

	Standard	RBPF
$g \leq 75\,\text{mg/dl}$	11.41	7.56
$g > 75\,\text{mg/dl}$	8.03	11.55
Overall	**8.64**	**10.83**

(b) gMAD in mg/dl.

This is mainly due to the fact that using the RBPF the typically computationally costly segmentation procedures can be omitted here.

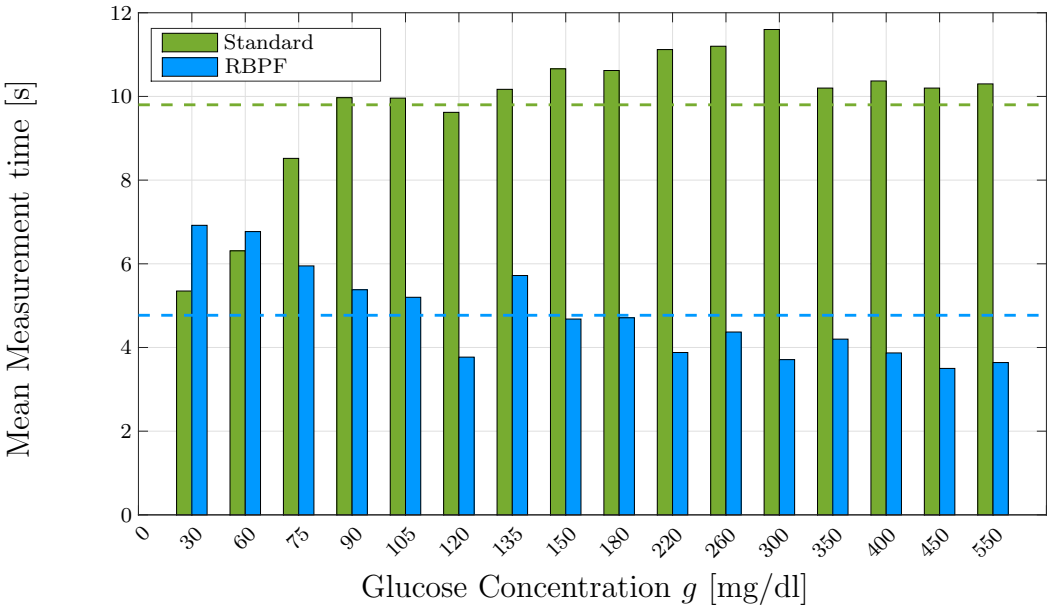


Figure 3.17: Comparison of the measurement times resulting from the standard convergence criterion and the RBPF convergence criterion of Set C for different underlying glucose concentrations. The dashed lines indicate the overall mean convergence times of the corresponding algorithms.

3.4 Summary and Conclusions

This chapter has presented the framework developed to measure the glucose concentration from photometric images, describing the chemical reaction taking place on the test strip.

It has been shown that pre-processing the frames is particularly beneficial as it eases consequent processing. This includes cropping the images, normalising them using calibration frames and binning the resulting frames. Binning the frames by a factor of $B = 5$ has proven to give the best results, allowing for noise reduction while preserving the image structure. Next, a Neyman-Pearson hypothesis test has been presented to detect the drop time n_D of the chemical reaction.

This has been followed by the development of a texture analysis approach to detect faulty glucose measurements. The approach presented is based on GLCM

texture features, more specifically the features energy, correlation and maximum probability. Both a sequential approach as well as a batch processing approach have been introduced. A perfect correct detection rate has been attained in all cases, while a false alarm rate of $0.63\% - 2.5\%$ has been achieved, depending on the feature used. The feature maximum probability has proved to be the most suitable for identifying the texture of faulty measurements. Using the sequential version of the developed approach, false measurements can be detected about 2.5 s earlier than when waiting for the measurements to saturate, hereby saving measurement time and unnecessary power consumption.

Finally, the problem of convergence detection has been highlighted. To this end, a model for the chemical kinetics has been derived, and a linear relation between the reaction rate and the convergence value has been established. This model has been used in two state estimation frameworks: the EKF and the RBPF. The EKF framework is able to predict the convergence values of the relative remission ahead of time. It has been elucidated that this approach clearly outperforms the state of the art in convergence detection both in accuracy and measurement time. A gain in time of up to 40 % has been achieved and a gain in precision of about 2% in relative remission. The RBPF outperforms both the EKF and the state of the art in reducing measurement time, attaining a reduction of up to 65%. The accuracy of the measurements however, does not improve in comparison to the state of the art. In fact, a slight deterioration in accuracy w.r.t. the state of the art can be determined. Therefore, the EKF approach is considered more appropriate for the application in a real photometric glucose measurement system.

Chapter 4

The Mean-Shift Approach for Segmentation of Glucose Images

This chapter presents the first segmentation approach developed in this thesis: segmentation based on the mean-shift concept. The aim is, given incoming glucose images, to segment the region of interest (ROI), i.e. the region where the chemical reaction is taking place, and accurately estimate the saturated relative remission value $\hat{r}_C$ of this region. This value is, then, mapped to an estimate of the underlying glucose concentration $\hat{g}$ as presented in Chap. 3.

The main contributions of this chapter are given by: (i) the adaption of the mean-shift approach to the glucose image segmentation problem. Section 4.1 motivates the choice of the mean-shift approach for the segmentation of blood glucose images. In Sec. 4.2.1 the standard mean-shift algorithm is reviewed and an alternative formulation of the mean-shift algorithm in terms of the reproducing kernel Hilbert space is given. Furthermore, the problem of kernel and bandwidth selection is highlighted. (ii) the development of a novel robust mean-shift variation in Sec. 4.2.2 that is able to deal with real world data, which does not follow the nominal assumptions and may be contaminated by outliers. (iii) the development of a novel sparse mean-shift variation that tackles the problem of high computational load of the standard mean-shift algorithm in Sec. 4.2.3. (iv) the introduction of the medoid-shift and the derivation of its robust and sparse variations in Sec. 4.2.4. (v) the analysis of the convergence properties of the developed algorithms in Sec. 4.3. (vi) the proposal of a scheme to identify the mode estimate corresponding to the ROI in Sec. 4.4. (vii) the validation of the developed approach using real data experiments in Sec. 4.5. Finally, Sec. 4.6 draws a conclusion.

The material presented in this chapter is partly taken from [DZ13,DZ14b,DZ15,DZ16].

4.1 State of the Art and Motivation

As presented in Sec. 2.4, recently an approach has been proposed to measure blood glucose concentration in invasive, hand-held glucometers, while drastically reducing the blood sample volume [AHKP11]. It relies on replacing the photodiode in the

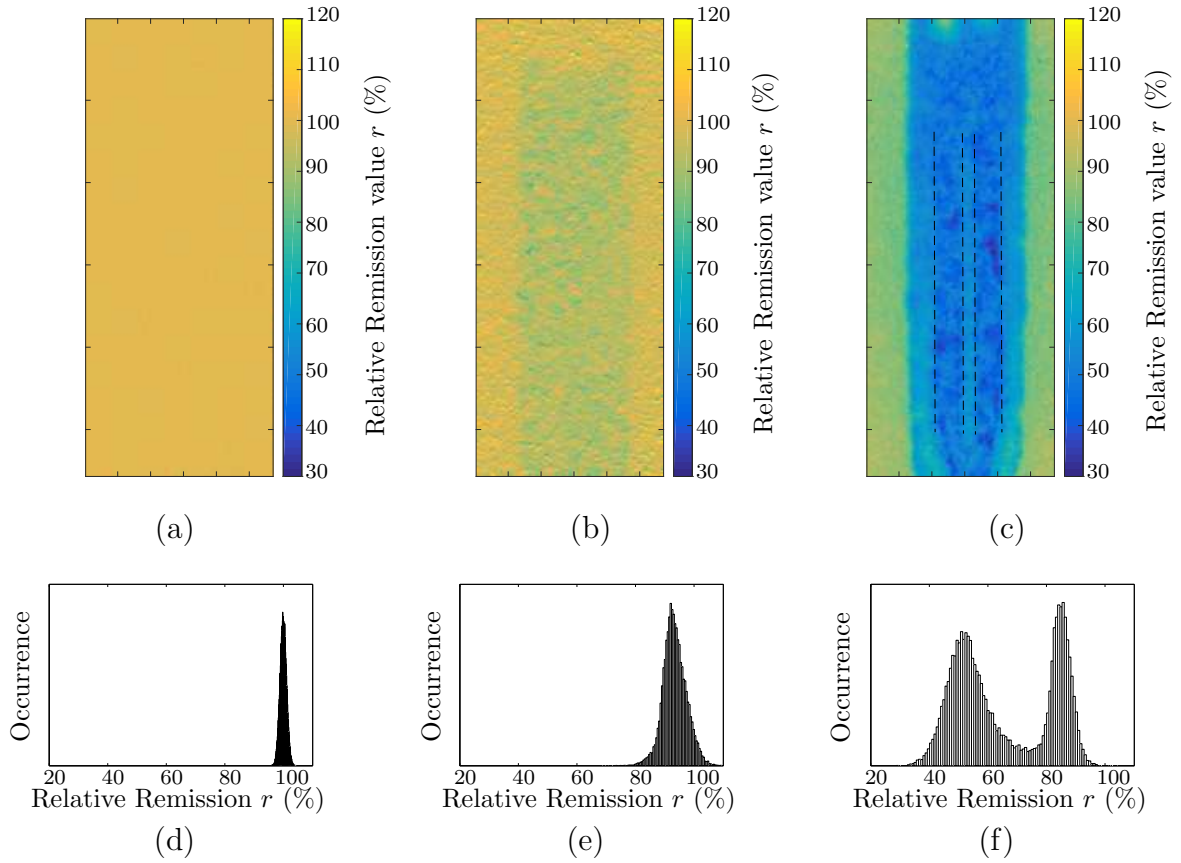


Figure 4.1: Examples of observed images and their respective histograms. (a) and (d) show an observation at $n < n_D$, (b) and (e) show an observation of a low glucose measurement at $n > n_C$, (c) and (f) show an observation of a high glucose measurement at $n > n_C$. The y-axis labels of the histogram have been omitted, as the absolute values of occurrence are irrelevant at this point.

traditional photometric measurement setup by a camera that observes the chemical reaction . In this setup, the colour change in the ROI does not cover the whole test strip but is only confined to small areas where the blood drop flows. By processing the observed videos of the chemical reaction the ROI can be delineated in each frame, and its underlying relative remission value $\hat{r}$ can be estimated. The segmentation of the frames and subsequent estimation of the relative remission $\hat{r}(n)$ of each frame constitute the objective of this chapter.

To the best of the author's knowledge, the problem of measuring blood glucose using the setup in Fig. 2.3 and an image processing-based approach has only been tackled in a limited number of studies. In [AHKP11] the authors propose to use a histogram-based approach to estimate the intensity of the region of interest. To this end, two clusters are assumed, one corresponding to the ROI and the other to the

background, with the respective mode locations (local maxima) indicating the position of the cluster centres. The authors propose to measure the displacement of the ROI cluster w.r.t. the background cluster. The underlying assumptions made are that just two clusters exist in the image, and that the background cluster exhibits a constant relative remission value over time. These are rather strong assumptions, which are hardly fulfilled in practice. A rigorous analysis of the real data gives evidence to a variable number of clusters in the images, as shown in Fig. 2.5 and Fig. 2.6. The variability is glucose level-specific, temporally dependent, and caused by the variability of the chemical reaction for different blood samples of different patients. Figure 4.1 presents examples of high glucose frames at different time instances, as well as their respective histograms. Evidently, the onset of the chemical reaction is characterised by a unimodal histogram. As time proceeds, more modes become visible indicating the existence of several clusters in the image. In Fig. 4.1 (f) it is not very clear whether two or three areas are present. On the one hand, one could argue that two regions are present, as the histogram shows two clear modes. On the other hand, the cluster on the left hand side is slightly skewed to the right, suggesting the presence of a third cluster. This assumption is underlined by the image itself depicted in Fig. 4.1 (c) that seems to contain three regions.

The second assumption concerning the constant behaviour of the background cluster is also not affirmed either when analysing the real data. Figure 4.1 clearly shows that the background cluster does not experience a constant intensity but changes over time, with its corresponding mode location moving from $r \approx 100\%$ in Fig. 4.1 (d) to $r \approx 85\%$ in Fig. 4.1 (f). The colour change of the background region can be attributed to temperature and humidity issues, as well as leakage of small amounts of blood to the background cluster.

This thesis assumes that the number of clusters in the images is unknown and changes for different underlying glucose concentrations as well as over time. Therefore, it has to be continuously estimated with each incoming frame. Furthermore, the assumption is made that the different image regions are all susceptible to some degree of the chemical reaction. Hence, they are susceptible to a change in relative remission. Therefore, the actual remission values of the regions should be estimated rather than the displacement between different relative remission values. Additionally, this thesis assumes that the glucose images contain a ROI, where the blood sample is distributed and, thus, the colour change of the ROI represents the reflectivity caused by the underlying glucose concentration. This area is not necessarily be completely spatially connected as it may contain artefacts where the reaction has not taken place, or granularities in the chemical that do not contribute to a proper reaction. Other

areas in the image do not correspond to the ROI, either, because the blood sample is unevenly distributed in these areas, or because the areas are dry, i.e., not covered with blood. The different areas contained in the image are interpreted as clusters with specific cluster centres overlaid by noise. The data points converging to a certain cluster center are assumed to belong to the corresponding area. The noise here is assumed to be Gaussian. This assumption is based on the calibration frames where no reaction has started. For the aforementioned reasons, traditional image segmentation approaches that directly work in the image domain, such as edge detection or intensity thresholding methods, fail at solving the given problem accurately. The approaches based on mean-shift proposed in this chapter work in the pdf (probability density function)-domain, and are able to systematically deal with these assumptions and segment the glucose images. Furthermore, they simultaneously estimate the relative remission value of the ROI as the estimated mode location of the corresponding cluster.

The mean-shift algorithm is an unsupervised, nonparametric clustering approach that is based on finding the modes of an underlying distribution and assigning each data point to its nearest mode. It does not require *a priori* information on the shape or the number of clusters, as this is rather an output of the algorithm. Each data point in the feature space is moved by the so-called mean-shift vector in the direction of the steepest ascent. Basically, the mean-shift algorithm is a gradient ascent approach with the advantage of an adaptive step-size that is determined by the magnitude of the mean-shift vector. The mean-shift vector reaches convergence when its magnitude becomes zero, indicating a stationary point. The mean-shift algorithm was proposed in 1975 by Fukanaga and Hostetler [FH75]. Later, it was readopted and generalised by Cheng [Che95] as a gradient ascent method. Both Cheng and Carreira-Perpinàn [CP00] proved its convergence for different kernel functions under mild conditions. Comaniciu *et al.*, [CRM00, CR00, CRM01, CM02, Com03], applied it to low-level vision problems such as segmentation and tracking. The mean-shift algorithm has recently become a popular tool for segmentation of biomedical images [WPS$^+$15, LMJM13, JCS$^+$12, SV10, MG09], where neither the number of clusters nor their shape are known in advance. In the sequel the original mean-shift approach will be reviewed, and its developed variations will be derived and their convergence proven. For a more detailed treatment of the standard mean-shift algorithm, the reader is referred to [Com00, CP15].

4.2 The Mean-Shift Algorithm and its Variations

4.2.1 The Standard Mean-Shift Algorithm

Assume a set of L pixels in a vectorised grey-scale image $\mathbf{x}^{(n)}$ at frame n as in Eq. (3.5) that can be expressed as $x_l^{(n)} \in \mathbb{R}$, $l = 1, \ldots, L$. Including spatial information for grey-scale images will increase the dimensions by two, leading to a three-dimensional data vector $\mathbf{x}_l^{(n)} = [x_l^{\text{intensity}}, m_{x,l}, m_{y,l}]^T \in \mathbb{R}^d$, where in this case $d = 3$. For better readability, the frame indicator, i.e. the superscript (n) in $\mathbf{x}_l^{(n)}$, will be omitted.

A consistent estimator of the density is given by the kernel density estimator (KDE) [Par62, Sil86]

$$\hat{f}_{\mathcal{K}}(\mathbf{x}) = \frac{1}{L} \sum_{l=1}^{L} \mathcal{K}_{\mathbf{H}}(\mathbf{x} - \mathbf{x}_l) \ , \tag{4.1}$$

where

$$\mathcal{K}_{\mathbf{H}}(\mathbf{x}) = \frac{1}{\det(\mathbf{H})} \mathcal{K}\big(\mathbf{H}^{-1}\mathbf{x}\big) \tag{4.2}$$

is a radially symmetric kernel function with a strictly decreasing profile for $x \geq 0$. The relation between the kernel function $\mathcal{K}(\cdot)$ and its univariate profile $K(\cdot)$ is given by

$$\mathcal{K}(\mathbf{x}) = c \cdot K\big(\|\mathbf{x}\|^2\big) \ , \tag{4.3}$$

with $\| \cdot \|^2$ denoting the Euclidean norm and c being a constant that ensures that $\mathcal{K}(\mathbf{x})$ integrates to one. The bandwidth matrix $\mathbf{H}$ is a symmetric positive definite $d \times d$ matrix. It is the main parameter that has to be tuned and is crucial to the performance of the mean-shift algorithm, as it affects the number of resulting clusters. It can be chosen to allow for equal bandwidths in all dimensions as: $\mathbf{H} = h \cdot \mathbf{I}_d$, $\mathbf{I}_d$ being the $d \times d$ identity matrix; or it can be chosen to allow for different bandwidths as: $\mathbf{H} = \text{diag}(h_1, \ldots, h_d)$. Extensive work has dealt with the choice of the bandwidth parameter, e.g., [CP15, CRM01, Sil86]. A discussion hereof will be given shortly.

Generally, the mean-shift algorithm is derived by taking the zeros of the density gradient estimator, which is equivalent to taking the zeros of the gradient of the KDE [FH75, CM02] [1]

$$\hat{\nabla} f(\mathbf{x}) \triangleq \nabla \hat{f}_{\mathcal{K}}(\mathbf{x}) \stackrel{!}{=} 0. \tag{4.4}$$

[1] Asymptotic unbiasedness, consistency and uniform consistency of the gradient estimate according to Eq. (4.4) was proven in [FH75].

Assuming equal bandwidths h in all dimensions, the gradient of the KDE reads

$$\begin{aligned}\nabla \hat{f}_{\mathcal{K}}(\mathbf{x}) &= \frac{2 \cdot c}{Lh^{d+2}} \sum_{l=1}^{L} (\mathbf{x} - \mathbf{x}_l) K'\big(\|\mathbf{H}^{-1}(\mathbf{x} - \mathbf{x}_l)\|^2\big) \\ &= \frac{2 \cdot c}{Lh^{d+2}} \sum_{l=1}^{L} K'\big(\|\mathbf{H}^{-1}(\mathbf{x} - \mathbf{x}_l)\|^2\big) \cdot \mathbf{m}_{\mathbf{H},\mathcal{K}}^{(\mathrm{MS})}(\mathbf{x}), \end{aligned} \tag{4.5}$$

where $K' = \frac{dK}{d\mathbf{x}}$ and $\mathbf{m}_{\mathbf{H},\mathcal{K}}^{(\mathrm{MS})}(\mathbf{x})$ is the mean-shift vector of $\mathbf{x}$

$$\mathbf{m}_{\mathbf{H},\mathcal{K}}^{(\mathrm{MS})}(\mathbf{x}) = \left[\frac{\sum_{l=1}^{L} \mathbf{x}_l K'(\|\mathbf{H}^{-1}(\mathbf{x} - \mathbf{x}_l)\|^2)}{\sum_{l=1}^{L} K'(\|\mathbf{H}^{-1}(\mathbf{x} - \mathbf{x}_l)\|^2)} - \mathbf{x} \right]. \tag{4.6}$$

The mean-shift vector can be intuitively understood as the difference between a data point and a weighted mean of all the other points. It always moves in the direction of the steepest ascent, with the step-size of the movement determined by the magnitude of the mean-shift vector. It is large for points with low local density and small for points with high local density. The mean-shift vector reaches convergence when its magnitude becomes zero, indicating a stationary point. Each data point is shifted according to Eq. (4.6) in the direction of steepest ascent until convergence, such that the sequence of successive locations of a data vector

$$\mathbf{x}^{(j+1)} = \left[\frac{\sum_{l=1}^{L} \mathbf{x}_l K'\big(\|\mathbf{H}^{-1}(\mathbf{x}^{(j)} - \mathbf{x}_l)\|^2\big)}{\sum_{l=1}^{L} K'(\|\mathbf{H}^{-1}(\mathbf{x}^{(j)} - \mathbf{x}_l)\|^2)} \right], \quad j = 1, 2, ..., \tag{4.7}$$

where $\mathbf{x}^{(j)}$ denotes the data vector at the j-th iteration step.

The standard mean-shift (MS) algorithm is presented in matrix form in the sequel. It requires as its inputs the data vectors $\mathbf{x}_l \in \mathbb{R}^d$, as well as the bandwidth parameter $\mathbf{H} \in \mathbb{R}^{d \times d}$ and the kernel function $\mathcal{K}_H(\cdot)$.

- **Step 1.** Initialise $\mathbf{x}^{(0)} = \mathbf{x}$.
- **Step 2.** While $\|\mathbf{x}^{(j+1)} - \mathbf{x}^{(j)}\|^2 > \epsilon$
  - $j := j + 1$.
  - Calculate the kernel matrix $\mathbf{C} = \Big(K'\big(\|\mathbf{H}^{-1}(\mathbf{x}^{(j)} - \mathbf{x}_l)\|^2\big)\Big)_{\mathbb{R}^{L \times L}}$, $l, k = 1, \ldots, L$.
  - Sum over the columns of the kernel matrix, obtaining $\mathbf{D} = \mathrm{diag}\Big(\sum_{l=1}^{L} c_{lk}\Big)$.
  - Compute next iterate of data vector $\mathbf{x}^{(j+1)} = (\mathbf{x}^{(j)} \cdot \mathbf{C}) \cdot \mathbf{D}^{-1}$.
- **Step 3.** Stop when the difference in iterates is small $\|\mathbf{x}^{(j+1)} - \mathbf{x}^{(j)}\|^2 < \epsilon$, obtaining a set $\tilde{k}$ of preliminary estimated mode locations $\tilde{\mathbf{y}} = [\tilde{\mathbf{y}}_1, \ldots, \tilde{\mathbf{y}}_{\tilde{k}}] \in \mathbb{R}^d$.

The mean-shift algorithm results in a set of estimated mode locations $\tilde{\mathbf{y}} = [\tilde{\mathbf{y}}_1, \ldots, \tilde{\mathbf{y}}_{\tilde{k}}] \in \mathbb{R}^d$, i.e. the locations of the shifted data vectors after convergence.

It has been shown in [CM02] that, under mild conditions, the mean-shift algorithm converges to the modes of the underlying pdf from any initial starting point, and that the convergence is to local maxima, as long as they are unique stationary points within some small open sphere. Carreira-Perpinàn [CP00] showed that the Gaussian mean-shift can be understood as an Expectation-Maximisation (EM) algorithm [Moo96] with linear convergence.

The advantages of the mean-shift algorithm are that it makes little assumptions on the data and only requires the choice of a suitable kernel function and a bandwidth parameter. These have a physical meaning and can be chosen mindfully.

The disadvantages of the mean-shift algorithm are mainly given by its slow computation, as the standard mean-shift has a complexity of $\mathcal{O}(JL^2d)$, J being the number of iterations which in practice does not exceed $J_{\max} = 50$. This thesis proposes the sparse mean-shift approach to tackle this problem, presented in Sec. 4.2.3.

Another disadvantage is that the mean-shift algorithm, typically, tends to overestimate the number of modes. This is due to some points converging to saddle points or local minima, when the initialisation falls on these points, forming small separate clusters. Another reason is that the mean-shift iterations are stopped when the error between iterations becomes small. This results in some points being close to the mode and hereby, converging to several neighbouring clusters. It is, therefore, typical to post-process the mean-shift results with a simple clustering method to group nearby modes together, e.g., using the so-called mode pruning step [DZ13, CM02]. The mode pruning process used in this thesis is summarised in the sequel, where $\tilde{\mathbf{y}} = [\tilde{\mathbf{y}}_1, \ldots, \tilde{\mathbf{y}}_{\tilde{k}}] \in \mathbb{R}^d$ denotes the vector containing the estimated mode locations from mean-shift and $\tilde{k}$ is the preliminary number of estimated modes.

- **Step 1.** Initialise $\tilde{\mathbf{y}}_1$ as the center of mass of the first cluster.
- **Step 2.** Go to the next mode location $\tilde{\mathbf{y}}_i$ and calculate its Euclidean distance to the existing clusters.
- **Step 3.** If the distance is larger than h (from Eq. (4.13)), assign the current mode location to be the centre of mass of a new cluster. Otherwise, assign it to an existing cluster.
- **Step 4.** Update the clusters.
- **Step 5.** Go to step 2.

The mode pruning process ends after iterating through all estimated mode locations. Mode pruning results in a smaller set of updated mode locations $\hat{\mathbf{y}} = [\hat{\mathbf{y}}_1, \ldots, \hat{\mathbf{y}}_{\hat{k}}]$, $\hat{k}$ being the final estimated number of modes.

To solve the problem of overestimation, the medoid-shift algorithm, as well as its variations, are developed and presented in Sec 4.2.4. They do not need a mode pruning step.

Kernel Function and Bandwidth Selection

The selection of the kernel function and the bandwidth parameter is crucial to the performance of the mean-shift algorithm. Many works have dealt with this in the literature [CP15, CRM01], therefore, this section will be kept short and will only comprise the important points concerning segmentation of photometric images.

Generally, the kernel function should satisfy following conditions [CM02]

$$\begin{aligned} &\int_{\mathbb{R}^d} \mathcal{K}(\mathbf{x}) d\mathbf{x} = 1, \\ &\lim_{\|\mathbf{x}\| \to \infty} \|\mathbf{x}\|^d \mathcal{K}(\mathbf{x}) = 0 \\ &\int_{\mathbb{R}^d} \mathbf{x} \cdot \mathcal{K}(\mathbf{x}) d\mathbf{x} = 0 \\ &\int_{\mathbb{R}^d} \mathbf{x}\mathbf{x}^T \cdot \mathcal{K}(\mathbf{x}) d\mathbf{x} = c_{\mathcal{K}} \mathbf{I}, \end{aligned} \tag{4.8}$$

$c_{\mathcal{K}}$ being a constant.

The two most common kernels used in the mean-shift framework are the Epanechnikov

and the Gaussian kernel [CM02]. While the Epanechnikov kernel is computationally more efficient, the Gaussian kernel produces better results [CM02], and thus is used in this thesis. The adopted Gaussian kernel has the following profile

$$K(x) = e^{-\frac{x^2}{2}}, \quad x \geq 0. \tag{4.9}$$

When the data vector is $\mathbf{x}^{(n)} = [x^{\text{intensity}}, m_x, m_y]^T \in \mathbb{R}^3$, the multivariate kernel function becomes a product of two radially symmetric kernels, with a separate bandwidth parameter for each domain

$$\mathcal{K}_{h_r,h_p}(\mathbf{x}) = \frac{c}{h_r h_p^2} K\left(\left\|\frac{x^r}{h_r}\right\|^2\right) \cdot K\left(\left\|\frac{\mathbf{x}^p}{h_p}\right\|^2\right), \tag{4.10}$$

h_r being the intensity or remission bandwidth, and h_p the position bandwidth parameter. The vector $\mathbf{x}^p = [m_x, m_y]$ denotes the spatial component and $x^r = x^{\text{intensity}}$ the intensity component of the data vector. Intuitively, this means that including the spatial components gives data vectors in the proximity of a data vector $\mathbf{x}$ higher weights than data vectors that are further away.
The inclusion of the spatial weighting kernel function leads to the following bandwidth matrix $\mathbf{H}$

$$\mathbf{H} = \begin{pmatrix} h_r & 0 & 0 \\ 0 & h_p & 0 \\ 0 & 0 & h_p \end{pmatrix}, \tag{4.11}$$

which allows finding separate bandwidths for intensity and position.

The basis for the choice of the bandwidth parameter can be either statistical or task-driven [PM90]. In this thesis the statistical approach to compute h_r is used, which relies on maximising the quality of the KDE by minimising the error between the actual density $f(\mathbf{x})$ and its estimator $\hat{f}(\mathbf{x})$. This is, typically, done by finding the best compromise between bias and variance over all $\mathbf{x} \in \mathbb{R}^d$. This requires minimising the asymptotic mean integrated squared error (AMISE) [CRM01, Sil86, WJ94] given for the univariate case by

$$\text{AMISE}(\hat{f}) = \frac{1}{Lh}\|K\|^2 + \frac{h^2}{4}\left(\mu_2(K)\right)^2 \|f''\|_2^2, \tag{4.12}$$

where $\mu_2(K) = \int u^2 K(u) du$ is the second-order moment function and $f'' = \frac{d^2 f}{d\mathbf{x}^2}$. Differentiating the AMISE w.r.t h and choosing the Gaussian kernel for K, the following optimal bandwidth parameter is achieved [Sil86, HWMS04]

$$\hat{h}^{\text{rot}} \approx 1.06 \cdot \hat{\sigma} L^{-\frac{1}{5}}, \tag{4.13}$$

where the robust scale estimator is chosen to estimate the standard deviation

$$\hat{\sigma}(\mathbf{x}) = 1.483 \cdot \text{median}(|\mathbf{x} - \text{median}(\mathbf{x})|). \tag{4.14}$$

The bandwidth parameter in Eq. (4.13) can be very sensitive to outliers and, therefore, a more robust estimator can be derived using the interquartile range Q

$$\hat{h}^{\text{rrot}} \approx 0.79 \cdot \hat{Q} L^{-\frac{1}{5}}. \tag{4.15}$$

Often both estimators are combined to obtain the following form

$$\hat{h}^{\text{crot}} \approx 1.06 \cdot \min\left(\hat{\sigma}, \frac{\hat{Q}}{1.34}\right) L^{-\frac{1}{5}}. \tag{4.16}$$

The bandwidth estimators in Eq. (4.13)-(4.16) all result in optimal AMISE for the unimodal Gaussian case. However, the photometric frames often show multiple modes and sometimes deviate from the Gaussian assumption. One way to tackle this is by using the so-called adaptive kernel density estimator (AKDE) [Sil86], which uses a data-driven approach to adapt the bandwidth to the data samples. Instead of using a fixed bandwidth h for all samples, a bandwidth parameter $\mathbf{h}(\mathbf{x}_l) \triangleq \mathbf{h}_l$ according to Abramson [Abr82] is calculated. The intuition behind this choice of bandwidth is that it should become large for points with low local density and small for points with high local density. It is given by

$$\mathbf{h}(\mathbf{x}) = h_0 \left[\frac{\lambda}{\hat{f}_0(\mathbf{x})}\right]^{\alpha}, \tag{4.17}$$

where h_0 is an initial fixed bandwidth for which Eq. (4.13)-(4.16) can be used, α is a sensitivity parameter satisfying $0 \leq \alpha \leq 1$. Here, $\alpha = \frac{1}{2}$ is chosen as suggested in [Sil86]. λ is a proportionality constant that splits the data into high and low densities. A good choice is to take it as the geometric mean of $\{\hat{f}_0(\mathbf{x}_l)\}_{l=1,\ldots,L}$, attaining $\lambda = \left(\prod_{l=1}^{L} \hat{f}_0(\mathbf{x}_l)\right)^{\frac{1}{L}}$. An initial estimate of the probability density $\hat{f}_0(\mathbf{x})$, the so-called pilot estimator, is also needed to calculate $\mathbf{h}(\mathbf{x})$ and is calculated using the KDE for the fixed bandwidth in Eq. (4.13) [Sil86]. From now on, the kernel function will be denoted by $\mathcal{K}_{\mathbf{h}}(\cdot)$ to indicate the adaptive bandwidth $\mathbf{h}$ as in Eq. (4.17) and by $\mathcal{K}_h(\cdot)$ to indicate the fixed bandwidth h as in Eq. (4.13)-(4.16).

Setting the spatial bandwidth parameter h_p is a more straightforward matter, as it only depends on the frame size.

The KDE in the Reproducing Kernel Hilbert Space

An alternative formulation for the KDE, which allows for extensions of the mean-shift, is to consider the kernel to be an inner product in the reproducing kernel Hilbert space (RKHS) $\mathcal{H}$, such that [SS02]

$$\mathcal{K}_h\left(\mathbf{x} - \mathbf{x}_l\right) = \langle \Phi(\mathbf{x}), \Phi(\mathbf{x}_l) \rangle, \tag{4.18}$$

where $\Phi : \mathbb{R} \rightarrow \mathcal{H}$ is a mapping function and $\langle \cdot \rangle$ denotes the inner product. In the machine learning community, this transformation is termed *the kernel trick* [SS02, SC08]. Every positive semidefinite kernel function can be associated to a unique space of functions called the Hilbert space $\mathcal{H}$. The important property here is the *reproducing property* which states that for every $g \in \mathcal{H}$ and all $\mathbf{x} \in \mathbb{R}^d$, $g(\mathbf{x}) = \langle \Phi(\mathbf{x}), g \rangle$ lies in $\mathcal{H}$. Hence, the KDE can be formulated as

$$\begin{aligned} \hat{f}_{\mathcal{K}}(\mathbf{x}) &= \frac{1}{L} \sum_{l=1}^{L} \langle \Phi(\mathbf{x}), \Phi(\mathbf{x}_l) \rangle \\ &= \langle \Phi(\mathbf{x}), \frac{1}{L} \sum_{l=1}^{L} \Phi(\mathbf{x}_l) \rangle, \end{aligned} \tag{4.19}$$

and $\hat{f}_{\mathcal{K}} \in \mathcal{H}$. By the reproducing property

$$\hat{f}_{\mathcal{K}}(\cdot) = \frac{1}{L} \sum_{l=1}^{L} \Phi(\mathbf{x}_l). \tag{4.20}$$

the KDE can be interpreted as the sample mean of $\Phi(\mathbf{x}_l) \in \mathcal{H}$, $\forall l = \{1, \ldots, L\}$. Hence, the KDE can be considered a solution to a least-squares problem

$$\hat{f}_{\mathcal{K}}(\cdot) = \arg\min_{g \in \mathcal{H}} \sum_{l=1}^{L} \| \Phi(\mathbf{x}_l) - g \|_{\mathcal{H}}^2, \tag{4.21}$$

where $\|g\|_{\mathcal{H}}^2 = \langle g, g \rangle$ denotes the norm in the Hilbert space $\mathcal{H}$.

The Hilbert space formulation will be used in the sequel to derive the proposed variants of the mean-shift algorithm. For a detailed study of the RKHS, the reader is referred to [SC08].

4.2.2 The Robust Mean-Shift Algorithm

Real data is often contaminated by outliers and rarely follows the nominal distributional assumptions. In the glucose segmentation application, the frames can be contaminated due to lighting effects or artefacts in the scene of interest, such as air bubbles, granularities in the chemical substance or small contaminating particles. Examples for such cases are presented in Fig 4.2. For these cases that deviate from the Gaussian assumption, the standard mean-shift algorithm may produce a bias. Using a standard mean-shift algorithm with an adaptive bandwidth as in Eq. (4.17) is one approach to tackle this problem. Another approach is to use the medoid-shift [SAS09, SKK07] that replaces the mean by the median, hereby

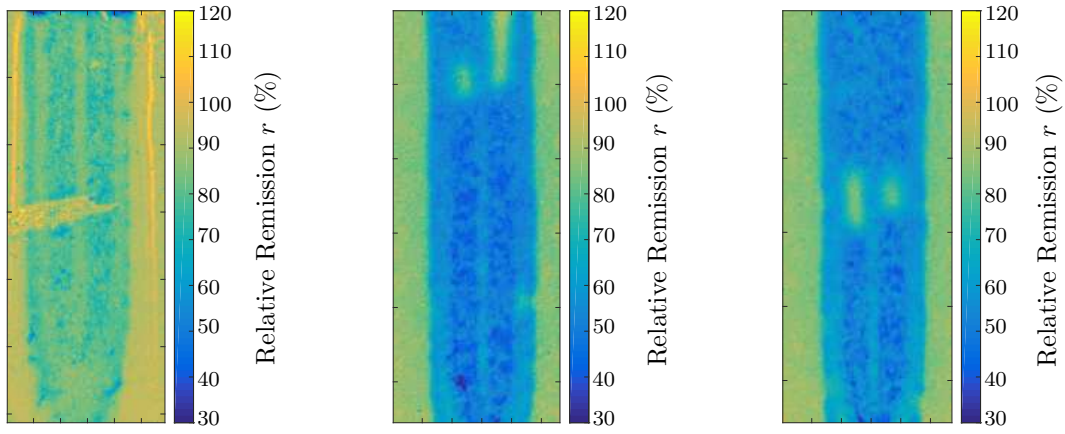


Figure 4.2: Examples of glucose images contaminated with artefacts. (a) shows lighting effects and a dust particle, (b) and (c) are examples of air bubble contaminations.

gaining robustness. Nevertheless, it tends to under-segment the data [JA13] and, therefore, does not always correctly assign each data point to the corresponding cluster.

In [DZ14b, DZ16], the robust mean-shift (R-MS) algorithm is derived as an alternative to account for heavy-tailed noise in the data. It is based on the robust KDE (RKDE) of Kim and Scott [KS08, KS11, KS12]. Recalling Eq. (4.21), the KDE is interpreted as the sample mean of $\Phi(\mathbf{x}_l) \in \mathcal{H}, \forall l \in \{1, \ldots, L\}$ and the solution to a least-squares problem. However, the least-squares (LS) estimator is non-robust in the presence of outliers or heavy-tailed noise. Robust statistics present a good solution to this problem [Hub11, ZKCM12]. For location estimation the so-called *M-estimator* provides a robust solution [Hub11] when the data is contaminated. The underlying contaminated data model assumed by the M-estimator is given as [Hub11]

$$f_{\mathbf{X}} = (1 - \varepsilon) f_N + \varepsilon f_C, \tag{4.22}$$

for contaminated data $\mathbf{x}$. Here, f_N is the nominal distribution, f_C is the contaminating distribution and ε is the degree of contamination.

Replacing the LS estimator by the M-estimator, Eq. (4.21) is replaced by

$$\hat{f}_{\mathcal{K}}^{\mathrm{RKDE}}(\cdot) = \arg\min_{g \in \mathcal{H}} \sum_{l=1}^{L} \rho\left(\left\| \frac{\Phi(\mathbf{x}_l) - g}{\hat{\sigma}} \right\|_{\mathcal{H}}^2 \right), \tag{4.23}$$

where $\rho(\cdot)$ is a monotone, differentiable loss function, such as Huber's loss function [Hub64] that gives smaller weights to larger distances $\left\| \frac{\Phi(\mathbf{x}_l) - g}{\hat{\sigma}} \right\|_{\mathcal{H}}^2$. The scale $\hat{\sigma}$ is

initialised with a robust estimate [ZKCM12]

$$\hat{\sigma}(\mathbf{x}) = 1.483 \cdot \text{median}(|\mathbf{x} - \text{median}(\mathbf{x})|).$$

Different loss functions exist in the literature [BBW89]. Following assumptions on $\rho(\cdot)$ and $\psi = d\rho/dx$ are made to find a solution for Eq. (4.23):

- $\rho(\cdot)$ is non-decreasing,
- $\rho(0) = 0$,
- $\rho(x)/x \to 0$ as $x \to 0$
- $\varphi(0) \triangleq \lim_{x \to 0} \frac{\psi(x)}{x}$ exists and is finite.

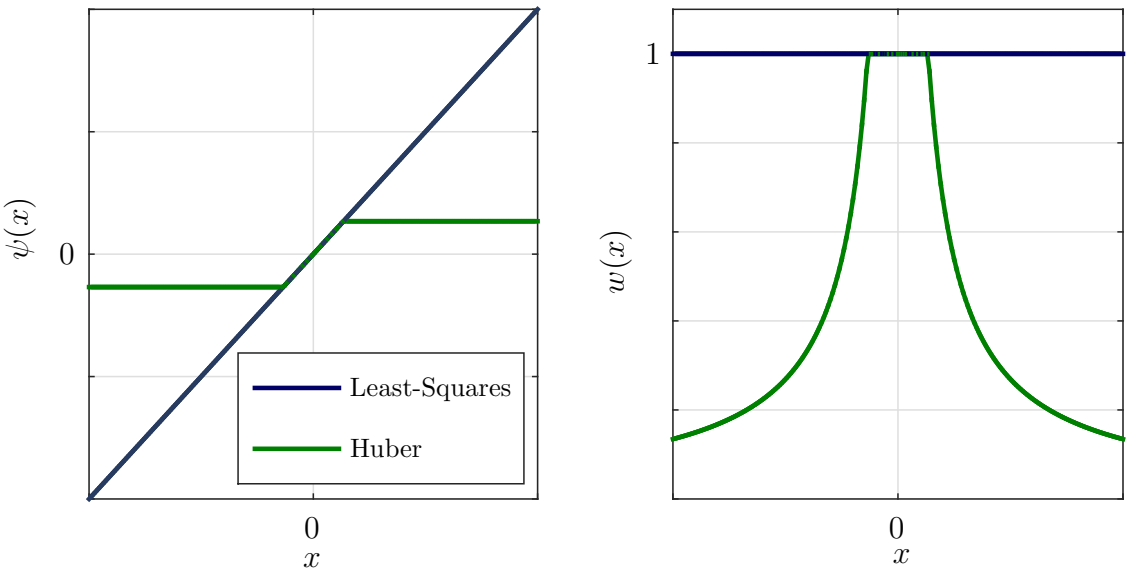


Figure 4.3: Huber's (a) score $\psi(x)$ and (b) weight $w(x)$ functions. The least-squares solution is shown for comparison.

In this thesis Huber's loss function [Hub64] and Hampel's loss function [Ham74] are used. They are given by

$$\rho^{\text{Huber}}(x) = \begin{cases} \frac{x^2}{2} & 0 \leq x \leq a \\ ax - \frac{a^2}{2} & a < x \end{cases}, \tag{4.24}$$

$$\psi^{\text{Huber}}(x) = \begin{cases} x & 0 \leq x \leq a \\ a & a < x \end{cases}, \tag{4.25}$$

$$\rho^{\text{Hampel}}(x) = \begin{cases} \frac{x^2}{2} & 0 \le x \le a \\ ax - \frac{a^2}{2} & a \le x \le b \\ \frac{a(x-c)^2}{(c-b)} + \frac{a(b+c-a)}{2} & b \le x \le c \\ \frac{a(b+c-a)}{2} & c \le x \end{cases}, \tag{4.26}$$

$$\psi^{\text{Hampel}}(x) = \begin{cases} x & 0 \le x \le a \\ a & a \le x \le b \\ a \cdot \frac{c-x}{c-b} & b \le x \le c \\ 0 & c \le x \end{cases}. \tag{4.27}$$

Fig. 4.3 illustrates Huber's score function $\psi(x)$, the resulting weights $w(x)$ from Eq. (4.30), as well as the score and weight functions for the LS solution for comparison.

According to the representer theorem in [KS12], the Gateaux differential [Lue97] of Eq. (4.23) is solved by the RKDE

$$\begin{aligned} \hat{f}_{\mathcal{K}}^{\text{RKDE}}(\mathbf{x}) &= \sum_{l=1}^{L} w_l \mathcal{K}_h(\mathbf{x} - \mathbf{x}_l) \\ &= \sum_{l=1}^{L} w_l \langle \Phi(\mathbf{x}), \Phi(\mathbf{x}_l) \rangle, \end{aligned} \tag{4.28}$$

where

$$w_l \propto \varphi\left(\left\| \frac{\Phi(\mathbf{x}_l) - \hat{f}_{\mathcal{K}}^{\text{RKDE}}}{\hat{\sigma}} \right\|_{\mathcal{H}} \right). \tag{4.29}$$

Intuitively, the weights w_l are small if $\hat{f}_{\mathcal{K}}^{\text{RKDE}}(\mathbf{x}_l)$ is small, hereby down-weighting points with low local densities.

Generally, Eq. (4.23) does not have a closed form solution and the weights w_l have to be found iteratively. They can be obtained using a kernalised version of Iteratively ReWeighted Least Squares (IRWLS) [MMY06, KS12] as

$$w_l^{\text{R-MS}} = \begin{cases} \frac{\hat{\sigma} \cdot \psi\left(\frac{\| \Phi(\mathbf{x}_l) - \hat{f}_{\mathcal{K}}^{\text{RKDE}} \|}{\hat{\sigma}} \right)}{\left\| \Phi(\mathbf{x}_l) - \hat{f}_{\mathcal{K}}^{\text{RKDE}} \right\|} & \text{if} \quad \frac{\left\| \Phi(\mathbf{x}_l) - \hat{f}_{\mathcal{K}}^{\text{RKDE}} \right\|}{\hat{\sigma}} \neq 0 \\ \psi(0) & \text{if} \quad \frac{\left\| \Phi(\mathbf{x}_l) - \hat{f}_{\mathcal{K}}^{\text{RKDE}} \right\|}{\hat{\sigma}} = 0. \end{cases} \tag{4.30}$$

The kernalised IRWLS (KIRWLS) proceeds as follows:

- **Step 1.** Initialise $\mathbf{w}^{(0)}$ with uniform weights $w_l^{(0)} = \frac{1}{L}, l = 1, \ldots, L$.
- **Step 2.** While the algorithm has not converged, i.e., $|\hat{f}_{\mathcal{K}}^{\mathrm{RKDE},(j)} - \hat{f}_{\mathcal{K}}^{\mathrm{RKDE},(j-1)}| > \hat{\sigma} \cdot \epsilon$, perform the following steps
  - $j := j + 1$
  - Compute $\|\Phi(\mathbf{x}_l) - \hat{f}_{\mathcal{K}}^{\mathrm{RKDE},(j)}\|$ as in Eq. (4.31).
  - Calculate $\hat{\sigma}(\|\Phi(\mathbf{x}_l) - \hat{f}_{\mathcal{K}}^{\mathrm{RKDE},(j)}\|)$ robustly using Eq. (4.14).
  - Update $\tilde{\mathbf{w}}^{(j)} = \frac{\psi(\|\Phi(\mathbf{x}_l) - \hat{f}_{\mathcal{K}}^{\mathrm{RKDE},(j)}\|/\hat{\sigma})}{\|\Phi(\mathbf{x}_l) - \hat{f}_{\mathcal{K}}^{\mathrm{RKDE},(j)}\|/\hat{\sigma}}$ and normalize to $\mathbf{w}^{(j)}$.
- **Step 3.** Stop when the difference in iterates is small $|\hat{f}_{\mathcal{K}}^{\mathrm{RKDE},(j)} - \hat{f}_{\mathcal{K}}^{\mathrm{RKDE},(j-1)}| < \hat{\sigma} \cdot \epsilon$, obtaining the robust weights $\mathbf{w}^{\text{R-MS}}$, and the robust KDE $\hat{f}_{\mathcal{K}}^{\mathrm{RKDE}}$.

Using $\hat{f}_{\mathcal{K}}^{\mathrm{RKDE}} = \sum_{l=1}^{L} w_l \Phi(\mathbf{x}_l)$, $\|\Phi(\mathbf{x}_l) - \hat{f}_{\mathcal{K}}^{\mathrm{RKDE}}\|$ is obtained as

$$\begin{aligned}
\|\Phi(\mathbf{x}_l) - \hat{f}_{\mathcal{K}}^{\mathrm{RKDE}}\|_{\mathcal{H}}^2 &= \langle \Phi(\mathbf{x}_l) - \hat{f}_{\mathcal{K}}^{\mathrm{RKDE}}, \Phi(\mathbf{x}_l) - \hat{f}_{\mathcal{K}}^{\mathrm{RKDE}} \rangle \\
&= \langle \Phi(\mathbf{x}_l), \Phi(\mathbf{x}_l) \rangle - 2\langle \Phi(\mathbf{x}_l), \hat{f}_{\mathcal{K}}^{\mathrm{RKDE}} \rangle + \langle \hat{f}_{\mathcal{K}}^{\mathrm{RKDE}}, \hat{f}_{\mathcal{K}}^{\mathrm{RKDE}} \rangle \\
&= \mathcal{K}_h(\mathbf{x}_l - \mathbf{x}_l) - 2\sum_{j=1}^{L} w_j \mathcal{K}_h(\mathbf{x}_j - \mathbf{x}_l) \\
&\quad + \sum_{i=1}^{L}\sum_{j=1}^{L} w_i w_j \mathcal{K}_h(\mathbf{x}_i - \mathbf{x}_j). \qquad (4.31)
\end{aligned}$$

Finally, the KDE reads

$$\hat{f}_{\mathcal{K}}^{\mathrm{RKDE}}(\mathbf{x}) = \sum_{l=1}^{L} w_l^{\mathrm{R-MS}} \mathcal{K}_h(\mathbf{x} - \mathbf{x}_l),$$

representing a weighted version of the standard KDE. Here, the weights $w_l^{\mathrm{R-MS}}$ are derived from robust statistics and calculated according to the KIRWLS algorithm. The consistency of the RKDE for $L \to \infty$ and $h \to 0$ is derived in [VS13], showing that the RKDE converges to the true density in the asymptotic case and exhibits robustness for small finite sample sizes. Additionally, in [KS11, KS12] the influence function is derived for the KDE and the RKDE. The influence function is a qualitative measure of robustness, indicating the bias impact of infinitesimal contamination on the estimator. The authors show that the RKDE is less affected by outliers than the KDE.

Deriving the robust mean-shift algorithm from $\hat{f}_{\mathcal{K}}^{\mathrm{RKDE}}(\mathbf{x})$ comprises taking the gradient of the RKDE and setting it to zero

$$\hat{\nabla} f(\mathbf{x}) \triangleq \nabla \hat{f}_{\mathcal{K}}^{\mathrm{RKDE}}(\mathbf{x}) \stackrel{!}{=} 0.$$

As for the MS algorithm, this results in

$$\begin{aligned}\nabla \hat{f}_{\mathcal{K}}(\mathbf{x}) &= \frac{2 \cdot c}{h^{d+2}} \sum_{l=1}^{L} (\mathbf{x} - \mathbf{x}_l) w_l^{\text{R-MS}} K'\big(\|\mathbf{H}^{-1}(\mathbf{x} - \mathbf{x}_l)\|^2\big) \\ &= \frac{2 \cdot c}{h^{d+2}} \sum_{l=1}^{L} w_l^{\text{R-MS}} K'\big(\|\mathbf{H}^{-1}(\mathbf{x} - \mathbf{x}_l)\|^2\big) \cdot \mathbf{m}_{h,\mathcal{K}}^{(\text{R-MS})}(\mathbf{x}),\end{aligned} \tag{4.32}$$

obtaining the robust mean-shift vector as

$$\mathbf{m}_{h,\mathcal{K}}^{(\text{R-MS})}(\mathbf{x}) = \left[\frac{\sum_{l=1}^{L} w_l^{\text{R-MS}} \mathbf{x}_l K'(\|h^{-1}(\mathbf{x} - \mathbf{x}_l)\|^2)}{\sum_{l=1}^{L} w_l^{\text{R-MS}} K'(\|h^{-1}(\mathbf{x} - \mathbf{x}_l)\|^2)} - \mathbf{x} \right]. \tag{4.33}$$

The robust mean-shift vector is derived for a fixed bandwidth parameter h, as presented in Eq. (4.13)-Eq. (4.16). If the spatial information is to be included, the kernel $\mathcal{K}_h$ can be replaced by Eq. (4.10).

Note that the difference between Eq. (4.6) and Eq. (4.33) is the weights $w_l^{\text{R-MS}}$. The formulation of the robust mean-shift vector, thus, results in a separation of the problem, where the KDE is, first, expressed robustly. Then, the modes of the RKDE are found using the mean-shift concept. The movement towards the modes follows with a step-size that down-weights outliers in the data. The R-MS requires as its inputs the data vectors $\mathbf{x}_l \in \mathbb{R}^d$, a specific loss function $\rho(\cdot)$, as well as the bandwidth parameter h and the kernel function $\mathcal{K}_h(\cdot)$. It proceeds as described in the sequel.

- **Step 1.** Initialise $\mathbf{x}^{(0)} = \mathbf{x}$.
- **Step 2.** While $\|\mathbf{x}^{(j+1)} - \mathbf{x}^{(j)}\|^2 > \epsilon$
  - $j := j + 1$.
  - Compute a matrix of robust weights using KIRWLS (Eq. (4.30)) $\mathbf{W} = \text{diag}(\mathbf{w})$, where $\mathbf{w} \in \mathbb{R}^{1\times L}$.
  - Calculate the kernel matrix $\mathbf{C} = \left(K'\left(\|h^{-1}(\mathbf{x} - \mathbf{x}_l)\|^2\right)\right)_{\mathbb{R}^{L\times L}}, l, k = 1, \ldots, L$.
  - Calculate the weighted kernel matrix $\mathbf{C_W} = \mathbf{C} \cdot \mathbf{W}$.
  - Sum over the columns of the weighted kernel matrix, obtaining $\mathbf{D} = \text{diag}\left(\sum_{l=1}^{L} c_{lk}\right)$, where c_{lk} is the lk-th element of $\mathbf{C_W}$.
  - Compute next iterate of data vector $\mathbf{x}^{(j+1)} = (\mathbf{x}^{(j)} \cdot \mathbf{C_W}) \cdot \mathbf{D}^{-1}$.
- **Step 3.** Stop when the difference in iterates is small $\|\mathbf{x}^{(j+1)} - \mathbf{x}^{(j)}\|^2 < \epsilon$, obtaining a set of $\tilde{k}$ preliminary estimated mode locations $\tilde{\mathbf{y}} = [\tilde{\mathbf{y}}_1, \ldots, \tilde{\mathbf{y}}_{\tilde{k}}] \in \mathbb{R}^d$ and a set of robust weights $w_l^{\text{R-MS}}$.

The sequence of successive locations of $\mathbf{x}$ is given by

$$\mathbf{x}^{(j+1)} = \left[\frac{\sum_{l=1}^{L} w_l^{\text{R-MS}} \mathbf{x}_l K'\left(\|h^{-1}(\mathbf{x}^{(j)} - \mathbf{x}_l)\|^2\right)}{\sum_{l=1}^{L} w_l^{\text{R-MS}} K'(\|h^{-1}(\mathbf{x}^{(j)} - \mathbf{x}_l)\|^2)}\right], \quad j = 1, 2, ... \tag{4.34}$$

The convergence of the sequence of locations of the data vector Eq. (4.34) is guaranteed for $w_l^{\text{R-MS}} > 0$. Section 4.3 provides the convergence proof.

The robust mean-shift has a complexity of $\mathcal{O}(J^{\text{MS}} L^2 d + J^{\text{KIRWLS}} L^2)$, J^{KIRWLS} being the number of iterations of the KIWRLS, which is, typically, below $J_{\max}^{\text{KIWRLS}} = 50$.

4.2.3 The Sparse Mean-Shift Algorithm

A big disadvantage of the mean-shift algorithm is its computational complexity of $\mathcal{O}(jL^2 d)$ as stated in Sec. 4.2.1. For image segmentation problems, L usually presents the number of pixels in the image, which can be very large. In the glucose segmentation application, e.g. the original number of pixels is $L \approx 30,000$ making the problem prohibitive. Binning the images, as described in Chap. 3 reduces the number of pixels to $L \approx 1,200$, making it feasible, however still computationally heavy.

Some recent studies have addressed the computational complexity of the mean-shift. Scalable hardware architectures are proposed in [CWG+13, AMM09] to

accelerate the mean-shift. In [CP06], acceleration methods are presented for the Gaussian mean-shift and evaluated with respect to the segmentation performance. The first is based on spatial discretisation. The image is divided into a number of cells. The set of pixels in a common cell is assumed to converge to the same mode. To this end, when a point is visited in an already visited cell the travelled iterations are inherited instead of travelling them twice. The second method reduces the data set to a smaller subset based on neighbourhoods. All pixels within a certain distance are represented by one pixel, hereby reducing the data set. This is similar to the binning operation described in Chap. 3, which is performed for the glucose images as a pre-processing step. The third method proposed by [CP06] uses a sparse EM approach to reduce the data set. The fourth method performs mean-shift to quickly move away from the starting points. As soon as the hill-climbing procedure becomes slow Newton's method is adopted to converge to the final solution in a few iterations.

In [DZ15, DZ16], a method is proposed to reduce the computational effort required for the mean-shift, regardless of the kernel function employed. As the number of pixels L in an image is one major bottleneck for computational complexity, the proposed approach is based on choosing a subset of data points to represent the whole set and only use these, hereby reducing the number of pixels from L to L_{red}, where $L_{\text{red}} \ll L$. Recently published work by [CS14] to approximate the sample mean in a scalable, sparse manner constitutes the basis of the developed framework.

The notion of representing the KDE as a sample mean in the RKHS $\mathcal{H}$ provides the basis for the developed approach. The gist here lies in substituting the sample mean in Eq. (4.32) by a sparse approximation [CS14, CS15] that only uses a subset of all data vectors. The subset of data vectors is chosen according to an incoherence criterion, which will be described shortly. The subset of data vectors is be denoted by $\{\mathbf{x}_i\}, i \in \mathcal{I}^*$, where $1 \leq L_{\text{red}} \ll L$ and $\mathcal{I} \subseteq \{1, \ldots, L\}$ of cardinality $|\mathcal{I}| = L_{\text{red}}$ is defined to be a subset of the full set of indices.

The sparse approximation of the mean can be formulated as an unconstrained, quadratic optimisation problem

$$\min_{\mathcal{I}_{|\mathcal{I}|=L_{\text{red}}}} \min_{\alpha_i, i \in \mathcal{I}} \left\| \sum_{l=1}^{L} w_l \Phi(\mathbf{x}_l) - \sum_{i \in \mathcal{I}} \alpha_i \Phi(\mathbf{x}_i) \right\|^2 , \tag{4.35}$$

where the weights $\boldsymbol{\alpha} \in \mathbb{R}^{L_{\text{red}}}$. This is slightly different than the derivation of [CS14, CS15], as it is extended to include non-uniform weights w_l, such as the robust weights $w_l^{\text{R-MS}}$. This will allow the derivation of a sparse, robust mean-shift algorithm.

According to Eq. (4.35), the problem of finding a sparse mean consists of 1) finding a solution to the inner optimisation problem, i.e., solving for $\alpha_{\mathcal{I}^*}$, and 2) finding an optimal subset of indices $\mathcal{I}^*$. Once these are known, the sparse KDE (SKDE) reads

$$\hat{f}_{\mathcal{K}}^{\mathrm{SKDE}}(\mathbf{x}) = \sum_{i \in \mathcal{I}^*} \alpha_{\mathcal{I}^*,i} \mathcal{K}_h \left(\mathbf{x} - \mathbf{x}_i\right), \tag{4.36}$$

where $\alpha_{\mathcal{I}^*,i}$ denotes the i-th weight obtained using the optimal set $\mathcal{I}^*$.

First, a solution for $\boldsymbol{\alpha}_{\mathcal{I}^*}$ has to be found. To this end, a fixed optimal $\mathcal{I}^*$ is assumed, obtaining the solution to the inner optimisation problem of Eq. (4.35) as

$$\boldsymbol{\alpha}_{\mathcal{I}^*} = \boldsymbol{\Xi}_{\mathcal{I}^*}^{-1} \boldsymbol{\xi}_{\mathcal{I}^*}. \tag{4.37}$$

Here

$$\boldsymbol{\Xi}_{\mathcal{I}^*} = \left(\langle \Phi(x_i), \Phi(x_j) \rangle\right)_{i,j \in \mathcal{I}^*} \tag{4.38}$$

$$\boldsymbol{\xi}_{\mathcal{I}^*} = \sum_{j=1}^{L} w_j \langle \Phi(x_m), \Phi(x_j) \rangle, m \in \mathcal{I}^*, \tag{4.39}$$

where $\boldsymbol{\Xi}_{\mathcal{I}^*}$ is the *Gram matrix*. The weights $\boldsymbol{\alpha}_{\mathcal{I}^*}$ are, thus, calculated as the distance between all data vectors inside the chosen subset, normalised by the distances of the chosen data vectors to the set of all data vectors. To ensure convergence of the mean-shift algorithm, $\alpha_{\mathcal{I}^*,i} > 0, \forall i \in \mathcal{I}^*$ is required. Generally, $\alpha_{\mathcal{I}^*,i}$ will not always fulfill this property. Therefore, the weights are normalised by setting to zero all weights $\alpha_{\mathcal{I}^*,i} > 0$ and re-normalising such that $\sum_i \alpha_{\mathcal{I}^*,i} = 1$.

Next, the attention is turned to finding the optimal index subset $\mathcal{I}^*$. Accordingly, following assumptions are made: the cardinality L_{red} of the subset of data points is assumed to be given, and the optimal $\boldsymbol{\alpha}_{\mathcal{I}^*}$ has been found using Eq. (4.37). The problem in Eq. (4.35) reduces to

$$\min_{\mathcal{I}_{|\mathcal{I}|=L_{\mathrm{red}}}} \left\| \sum_{l=1}^{L} w_l \Phi(\mathbf{x}_l) - \sum_{i \in \mathcal{I}} \alpha_{\mathcal{I}^*,i} \Phi(\mathbf{x}_i) \right\|^2. \tag{4.40}$$

In [CS14,CS15], the authors propose to solve this problem by setting an upper bound on $\left\| \sum_{l=1}^{L} w_l \Phi(\mathbf{x}_l) - \sum_{i \in \mathcal{I}} \alpha_{\mathcal{I}^*,i} \Phi(\mathbf{x}_i) \right\|$ that depends on $\mathcal{I}$ and finding an $\mathcal{I}$ that minimises this bound. To this end, an incoherence function $\nu_{\mathcal{I}}$ is defined as

$$\nu_{\mathcal{I}} = \min_{j \notin \mathcal{I}} \max_{i \in \mathcal{I}} \langle \Phi(\mathbf{x}_i), \Phi(\mathbf{x}_j) \rangle. \tag{4.41}$$

The incoherence function $\nu_{\mathcal{I}}$ describes how disparate two data vectors in a set are. In the literature, coherence functions are a common tool to measure the closeness of

atoms in a dictionary in the context of optimisation problems [Tro04, NHR12, KS12]. The derived bound is given by Theorem 1 in [CS14] as

$$\left\| \sum_{l=1}^{L} w_l \Phi(\mathbf{x}_l) - \sum_{i \in \mathcal{I}} \alpha_{\mathcal{I}^*,i} \Phi(\mathbf{x}_i) \right\| \leq \left(1 - \frac{L_{\text{red}}}{L} \sqrt{\frac{1}{\langle \Phi(\mathbf{x}_i), \Phi(\mathbf{x}_i) \rangle} \left((\langle \Phi(\mathbf{x}_i), \Phi(\mathbf{x}_i) \rangle)^2 - \nu_{\mathcal{I}}^2 \right)} \right), \tag{4.42}$$

where $\langle \Phi(\mathbf{x}_i), \Phi(\mathbf{x}_i) \rangle > 0$ is constant $\forall i \in \mathcal{I}^*$. Since all components of the bound are constants except for $\nu_{\mathcal{I}}$, minimisation of the bound results in maximisation of $\nu_{\mathcal{I}}$

$$\mathcal{I}^* = \max_{\mathcal{I} \subseteq \{1,\ldots,L\}} \nu_{\mathcal{I}}. \tag{4.43}$$

Equation (4.43) is intuitive in the sense that to find the most representative subset of data vectors, the ones that are most incoherent to each other are chosen.

When $\mathcal{K}_h(\cdot)$ is the Gaussian kernel, the maximisation of $\nu_{\mathcal{I}}$ becomes

$$\mathcal{I}^* = \arg\min_{\mathcal{I} \subseteq \{1,\ldots,L\}} \max_{j \notin \mathcal{I}} \min_{i \in \mathcal{I}} \|\mathbf{x}_i - \mathbf{x}_j\|. \tag{4.44}$$

This is the well-known k-center problem [V.V01]. It takes as its inputs the data vectors $\mathbf{x}_l \in \mathbb{R}^d$ and the cardinality $L_{\text{red}} = |\mathcal{I}|$ of the reduced subset. The choice of L_{red} will be discussed shortly. First the k-center algorithm is presented [CS14,V.V01].

- **Step 1.** Create two sets: one containing the original set of data vectors $\mathcal{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_L\}$ and the second constituting the reduced subset of data vectors $\mathcal{X}^{\mathcal{I}} = \emptyset$, which is initialised as an empty set.
- **Step 2.** Randomly choose the first index $u \in \{1, \ldots, L\}$.
- **Step 3.** Remove $\mathbf{x}_u$ from the original set $\mathcal{X} = \mathcal{X}/\mathbf{x}_u$ and place it into the reduced set $\mathcal{X}^{\mathcal{I}} = \mathcal{X}^{\mathcal{I}} \cup \mathbf{x}_u$.
- **Step 4.** While $|\mathcal{X}^{\mathcal{I}}| < L_{\text{red}}$
  - Choose $\mathbf{x}_i \in \mathcal{X}$ for which $\min_{j \in \mathcal{I}} \|\mathbf{x}_i - \mathbf{x}_j\|$ is maximised.
  - Remove $\mathbf{x}_i$ from the original set $\mathcal{X} = \mathcal{X}/\mathbf{x}_i$.
  - Place $\mathbf{x}_i$ into the reduced set $\mathcal{X}^{\mathcal{I}} = \mathcal{X}^{\mathcal{I}} \cup \mathbf{x}_i$.
- **Step 3.** Stop when the reduced data set $\mathcal{X}^{\mathcal{I}}$ has the cardinality L_{red}, obtaining a set of optimal indices $\mathcal{I}^* = \{i \in \{1, \ldots, L\} | \mathbf{x}_i \in \mathcal{X}^{\mathcal{I}}\}$.

Until now, it has been assumed that the cardinality L_{red} of the subset of indices $\mathcal{I}$ is given. It remains to discuss how to choose a suitable value for L_{red}. Clearly, in some cases this can be given by the application at hand knowing a certain $L_{\text{red}}^{\max}$, or

by using test data for cross-validation, as in [DZ15], where L_{red} is set to be sufficiently high to incorporate the worst-case scenario. In [DZ16], a data-driven selection of L_{red} is introduced, which results in a smaller L_{red} than cross-validation for most cases. Figure 4.4 shows typical examples of the progression of a normalised version of $\nu_{\mathcal{I}}$

$$\widetilde{\nu}_{\mathcal{I}} = \frac{\nu_{\mathcal{I},|\mathcal{I}|=L_{\text{red}}}}{\max \nu_{\mathcal{I}}} \tag{4.45}$$

for the images given in Fig. 4.1, using different values of $L_{\text{red}} = 1, \ldots, L$. It can be observed that the value of $\widetilde{\nu}_{\mathcal{I}}$ drops quickly after a certain value of L_{red} and that the behaviour of the curve is similar for different images. This signifies that the first $L_{\text{red},\nu}$ samples, chosen by the algorithm, contribute highly to the incoherence in the image and it is sufficient to use a subset $\mathcal{I}$ of cardinality $L_{\text{red},\nu}$ for the sparse representation.

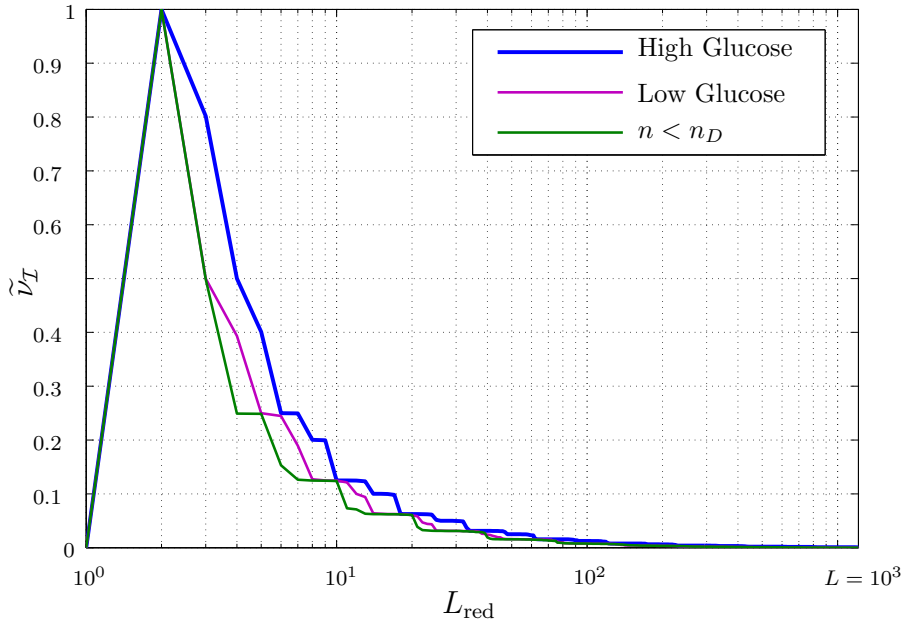


Figure 4.4: An example of the progression of normalised ν with increasing cardinality L_{red}.

Following criterion is developed to choose $L_{\text{red},\nu}$ based on the gradient of $\widetilde{\nu}_{\mathcal{I}}$ in Fig. 4.4 and a threshold T_ν

$$\Delta\widetilde{\nu}^{(L_{\text{red},\nu})} = \frac{\nu_{\mathcal{I},|\mathcal{I}|=L_{\text{red},\nu}}}{\nu_{\max}} - \frac{\nu_{\mathcal{I},|\mathcal{I}|=L_{\text{red},\nu}-1}}{\nu_{\max}} \leq T_\nu. \tag{4.46}$$

The choice of T_ν is chosen specifically for the data at hand by observing the trade-off between accuracy and computation.

So far, the optimal weights $\alpha_{\mathcal{I}^*,i}$ have been determined, the optimal set of indices $\mathcal{I}^*$ has been calculated and the cardinality $L_{\text{red}} = |\mathcal{I}|$ has been chosen using a data-driven method. Now, the sparse KDE reads

$$\hat{f}_{\mathcal{K}}^{\text{SKDE}}(\mathbf{x}) = \sum_{i \in \mathcal{I}^*} \alpha_{\mathcal{I}^*,i} \mathcal{K}_h \left(\mathbf{x} - \mathbf{x}_i\right).$$

and the mean-shift vector can be derived by taking the zeros of the gradient of the SKDE

$$\hat{\nabla} f(\mathbf{x}) \triangleq \nabla \hat{f}_{\mathcal{K}}^{\text{SKDE}}(\mathbf{x}) \stackrel{!}{=} 0,$$

where

$$\begin{aligned} \nabla \hat{f}_{\mathcal{K}}(\mathbf{x}) &= \frac{2 \cdot c}{h^{d+2}} \sum_{l=1}^{L} (\mathbf{x} - \mathbf{x}_l) \alpha_{\mathcal{I}^*,i} K'\big(\|\mathbf{H}^{-1}(\mathbf{x} - \mathbf{x}_i)\|^2\big) \\ &= \frac{2 \cdot c}{h^{d+2}} \sum_{l=1}^{L} \alpha_{\mathcal{I}^*,i} K'\big(\|\mathbf{H}^{-1}(\mathbf{x} - \mathbf{x}_i)\|^2\big) \cdot \mathbf{m}_{h,\mathcal{K}}^{\text{(S-MS)}}(\mathbf{x}). \end{aligned} \tag{4.47}$$

The resulting mean-shift vector has the form

$$\mathbf{m}_{h,\mathcal{K}}^{\text{(S-MS)}}(\mathbf{x}) = \left[\frac{\sum_{i \in \mathcal{I}^*} \alpha_{\mathcal{I}^*,i} \mathbf{x}_i K'(\|h^{-1}(\mathbf{x} - \mathbf{x}_i)\|^2)}{\sum_{i \in \mathcal{I}^*} \alpha_{\mathcal{I}^*,i} K'(\|h^{-1}(\mathbf{x} - \mathbf{x}_i)\|^2)} - \mathbf{x} \right]. \tag{4.48}$$

This formulation of the mean-shift vector includes the robust, sparse mean-shift (RS-MS) vector, substituting the weights in Eq. (4.39) by the robust weights $w_l^{\text{R-MS}}$ calculated in Eq. (4.30).

The sequence of successive locations of $\mathbf{x}$ is given by

$$\mathbf{x}^{(j+1)} = \left[\frac{\sum_{i \in \mathcal{I}^*} \alpha_{\mathcal{I}^*,i} \mathbf{x}_i K'\big(\|h^{-1}(\mathbf{x}^{(j)} - \mathbf{x}_i)\|^2\big)}{\sum_{\in \mathcal{I}^*} \alpha_{\mathcal{I}^*,i} K'(\|h^{-1}(\mathbf{x}^{(j)} - \mathbf{x}_i)\|^2)} \right], \quad j = 1, 2, ... \tag{4.49}$$

Section 4.3 shows that this sequence converges.

The S-MS requires as its inputs the data vectors $\mathbf{x}_l \in \mathbb{R}^d$, a specific loss function $\rho(\cdot)$ if the RS-MS is being calculated, as well as the bandwidth parameter h and the kernel function $\mathcal{K}_h(\cdot)$. It proceeds as described in the sequel.

- **Step 1.** Initialise $\mathbf{x}^{(0)} = \mathbf{x}$.
- **Step 2.** Determine the optimal set of subindices $\mathcal{I}^*$ using the k-center algorithm previously presented, hereby calculating L_{red} using Eq. (4.46).
- **Step 3.** Compute the matrix of sparse weights $\mathbf{W} = \text{diag}(\boldsymbol{\alpha}_{\mathcal{I}^*})$, where $\alpha_{\mathcal{I}^*,i}, i \in \mathcal{I}^*$ as in (4.37).
- **Step 4.** Determine the reduced set of data vectors $\tilde{\mathbf{x}} = \{\mathbf{x}_i | i \in \mathcal{I}^*\}$.
- **Step 5.** While $\|\tilde{\mathbf{x}}^{(j+1)} - \tilde{\mathbf{x}}^{(j)}\|^2 > \epsilon$
  - $j := j + 1$.
  - Calculate the kernel matrix $\mathbf{C} = (K'(\|h^{-1}(\mathbf{x} - \mathbf{x}_i)\|^2))_{\mathbb{R}^{L_{\text{red}} \times L_{\text{red}}}}$, $l, k = 1, \ldots, L_{\text{red}}$.
  - Calculate the weighted kernel matrix $\mathbf{C}_\mathbf{W} = \mathbf{C} \cdot \mathbf{W}$.
  - Sum over the columns of the weighted kernel matrix, obtaining $\mathbf{D} = \text{diag}\left(\sum_{l=1}^{L_{\text{red}}} c_{lk}\right)$, where c_{lk} is the lk-th entry of $\mathbf{C}_\mathbf{W}$.
  - Compute next iterate of data vector $\tilde{\mathbf{x}}^{(j+1)} = (\tilde{\mathbf{x}}^{(j)} \cdot \mathbf{C}_\mathbf{W}) \cdot \mathbf{D}^{-1}$.
- **Step 6.** Stop when the difference in iterates is small $\|\tilde{\mathbf{x}}^{(j+1)} - \tilde{\mathbf{x}}^{(j)}\|^2 < \epsilon$, obtaining a set of $\tilde{k}$ preliminary estimated mode locations $\tilde{\mathbf{y}} = [\tilde{\mathbf{y}}_1, \ldots, \tilde{\mathbf{y}}_{\tilde{k}}] \in \mathbb{R}^d$ and a set of sparse weights $\boldsymbol{\alpha}_{\mathcal{I}^*}$.
- **Step 7.** For each data vector not contained in the reduced set, i.e. $\overline{\mathbf{x}} = \{\mathbf{x}_i | i \notin \mathcal{I}^*\}$, determine the index of its closest neighbour from the set of points $\tilde{\mathbf{x}}$. All points $\overline{\mathbf{x}}$ are assigned to the mode locations of their closest neighbour from the set of points $\tilde{\mathbf{x}}$ to perform the segmentation.

For all other mean-shift variants, the segmentation of the data vector results automatically from the cluster membership, i.e. each data vector converging to a certain cluster is considered part of one segment. For sparse mean-shift, however, an additional step has to be carried out (Step 7. in the presented algorithm), as the clustering is only performed with a subset of all data vectors. To this end, each point not contained in the reduced dataset obtains the mode location of its closest neighbour. Herewith, the final segmentation result is obtained.

Coming back to the computational complexity of the derived algorithm, it can be shown that the S-MS outperforms the MS w.r.t. computation. Remember that the standard mean-shift algorithm has a computational complexity of $\mathcal{O}(J^{\text{MS}} L^2 d)$, where J^{MS} is the average number of iterations. The proposed approach reduces this due to the reduction of the data points, as well as the number of iterations

needed. However, the overhead caused by the calculation of $\mathcal{I}^*$ and $\alpha_{\mathcal{I}^*,i}$ has to be taken into consideration. This results in a computational complexity of $\mathcal{O}(J^{\text{S-MS}}L_{\text{red}}^2 d + L \cdot L_{\text{red}} d + L_{\text{red}}^3)$, where $J^{\text{S-MS}} < J^{\text{MS}}$ is the average number of iterations of S-MS. Thus, for the sparse mean-shift to result in an improvement in computation $J^{\text{MS}}L^2 d > J^{\text{S-MS}}L_{\text{red}}^2 d + L \cdot L_{\text{red}} d + L_{\text{red}}^3$. The experiments in Sec. 4.5 will show that this is the case for values of $L_{\text{red}}/L < 0.3$.

4.2.4 The Medoid-Shift Algorithm

The medoid-shift was derived by Sheikh *et al.* in [SKK07] as an alternative to the mean-shift that uses the medoid instead of the mean. The medoid is defined as the point contained in the data set that has the smallest distance to all other points. In the mean-shift formulation the point $\mathbf{x}^{(j+1)}$ is calculated from $\mathbf{x}^{(j)}$ using Eq. (4.47) as

$$\mathbf{x}^{(j+1)} = \arg\min_{\mathbf{x}} \frac{1}{L} \sum_{l=1}^{L} \|\mathbf{x} - \mathbf{x}_l\|^2 K'\big(\|h^{-1}(\mathbf{x}^{(j)} - \mathbf{x}_l)\|^2\big) \ . \tag{4.50}$$

Setting the derivative to zero and reformulating yields the mean-shift vector as in Eq. (4.6). The medoid-shift (MedS) slightly changes Eq. (4.50), searching for a point contained in the data set $\{\mathbf{x}_1, \ldots, \mathbf{x}_L\}$ that minimises this argument as

$$\mathbf{x}^{(j+1)} = \arg\min_{\mathbf{x}\in\{\mathbf{x}_l\}} \frac{1}{L} \sum_{l=1}^{L} \|\mathbf{x} - \mathbf{x}_l\|^2 K'\big(\|h^{-1}(\mathbf{x}^{(j)} - \mathbf{x}_l)\|^2\big) \ . \tag{4.51}$$

Note that, once one medoid-shift is calculated for all points $\mathbf{x}^{(j)}$, the next shifts for $\mathbf{x}^{(j+1)}$ will already exist, as $\mathbf{x}^{(j)}$ will have moved to a point in the data set. The algorithm is terminated when $\mathbf{x}^{(j)} = \mathbf{x}^{(j+1)}$. The advantages of medoid-shift are that it is quicker than mean-shift and that it will not need a mode pruning step, because the number of attained modes is limited by the data set. Furthermore, [VS08] shows that its computational complexity is $\mathcal{O}(L^2)$, and hereby smaller than the mean-shift. The disadvantage of medoid-shift, however, is that it will not always detect all the modes of $\hat{f}(\mathbf{x})$.

In [GR11] a medoid-shift with a data-driven bandwidth as in Eq. (4.17) is introduced, to robustify the medoid-shift for real-data applications. To this end, this thesis develops a robust medoid-shift, using the same rationale as when deriving the R-MS. The starting point is constituted by the RKDE of Eq. (4.32). The robust medoid-shift (R-MedS) can then be derived from

$$\mathbf{x}^{(j+1),\text{R-MedS}} = \arg\min_{\mathbf{x}\in\{\mathbf{x}_l\}} \sum_{l=1}^{L} \|\mathbf{x} - \mathbf{x}_l\|^2 w_l^{\text{R-MS}} K'\big(\|h^{-1}(\mathbf{x}^{(j)} - \mathbf{x}_l)\|^2\big) \ , \tag{4.52}$$

where $w_l^{\text{R-MS}}$ represents the weights from Eq. (4.30). The robust medoid-shift converges to a weighted medoid of the set of data vectors, down-weighting data vectors of low density.

Finally, a sparse medoid-shift is introduced by using the SKDE and deriving

$$\mathbf{x}^{(j+1),\text{S-MedS}} = \underset{\mathbf{x}\in\{\mathbf{x}_i\}}{\arg\min} \sum_{i\in\mathcal{I}^*} \|\mathbf{x}-\mathbf{x}_i\|^2 \times \alpha_{\mathcal{I}^*,i} K'\big(\|h^{-1}(\mathbf{x}^{(j)}-\mathbf{x}_i)\|^2\big) \ , \tag{4.53}$$

where $\alpha_{\mathcal{I}^*,i}$ stems from Eq. (4.37).

The medoid-shift algorithm proceeds as described in the sequel.

- **Step 1.** Initialise $\mathbf{x}^{(0)} = \mathbf{x}$.
- **Step 2.** If the sparse version is being computed, otherwise proceed without the following steps:
  - Determine the optimal set of subindices $\mathcal{I}^*$ using the k-center algorithm previously presented, hereby calculating L_{red} using Eq. (4.46).
  - Compute the matrix of sparse weights $\mathbf{W} = \text{diag}(\boldsymbol{\alpha}_{\mathcal{I}^*})$, where $\alpha_{\mathcal{I}^*,i}$, $i = 1, \cdots, L_{\text{red}}$as in Eq. (4.37).
  - Determine the reduced set of data vectors $\tilde{\mathbf{x}} = \{\mathbf{x}_i | i \in \mathcal{I}^*\}$.
- **Step 3.** While $\|\mathbf{x}^{(j+1)} - \mathbf{x}^{(j)}\|^2 > \epsilon$
  - $j := j + 1$.
  - For MedS initialise the weight matrix as $\mathbf{W} = \text{diag}(\mathbf{1})$, where $\mathbf{1} \in \mathbb{R}^{1\times L}$, for R-MedS initialise it as $\mathbf{W} = \text{diag}(\mathbf{w})$,where $\mathbf{w} \in \mathbb{R}^{1\times L}$ using KIRWLS.
  - Calculate the kernel matrix $\mathbf{C} = \big(K'\big(\|h^{-1}(\mathbf{x}^{(j)}-\mathbf{x}_i)\|^2\big)\big)_{\mathbb{R}^{L\times L}}$, $l, k = 1, \ldots, L$, if full set of indices are being used and $\mathbf{C} = \big(K'\big(\|h^{-1}(\mathbf{x}^{(j)}-\mathbf{x}_i)\|^2\big)\big)_{\mathbb{R}^{L_{\text{red}}\times L_{\text{red}}}}$, $l, k \in \mathcal{I}^*$, if the sparse version is being used.
  - Calculate the weighted kernel matrix $\mathbf{C_W} = \mathbf{C} \cdot \mathbf{W}$.
  - Compute the distance matrix $\mathbf{D} = \Big(\|\mathbf{x}_i - \mathbf{x}_j\|^2\Big)_{\mathbb{R}^{L\times L}}$ for the non-sparse versions and $\mathbf{D} = \Big(\|\tilde{\mathbf{x}}_i - \tilde{\mathbf{x}}_j\|^2\Big)_{\mathbb{R}^{L_{\text{red}}\times L_{\text{red}}}}$ for the sparse versions.
  - Calculate $\mathbf{S} = \mathbf{C_W} \cdot \mathbf{D}$.
  - Select the data vectors that produce minimum value over the columns of $\mathbf{S}$ as $\mathbf{x}^{(j+1)} = \{\mathbf{x}_k | k \in$ indices of minima across columns of $\mathbf{S}\}$.
- **Step 3.** Stop when the difference in iterates is small $\|\mathbf{x}^{(j+1)} - \mathbf{x}^{(j)}\|^2 < \epsilon$, obtaining a set of $\hat{k}$ estimated mode locations $\hat{\mathbf{y}} = [\hat{\mathbf{y}}_1, \ldots, \hat{\mathbf{y}}_{\hat{k}}] \in \mathbb{R}^d$.

The medoid-shift (MedS) algorithm requires as its inputs the data vectors $\mathbf{x}_l \in \mathbb{R}^d$, a specific $\rho(\cdot)$-function if the robust version is being calculated, as well as the bandwidth parameter h and the kernel function $\mathcal{K}_h(\cdot)$. Note that MedS estimates the final mode locations $\hat{\mathbf{y}}$ without mode pruning.

The convergence of the sequence of trajectory points in the medoid-shift is guaranteed [SKK07] and the proof is analogous to that of the mean-shift. The only difference is that for the medoid-shift there should be no cycles in the sequence of trajectory points, i.e., $\mathbf{x}^{(j)} \neq \mathbf{x}^{(j+c)}$, for all $c > 0$. The proof thereof is given in Sec. 4.3.

The sparse medoid-shift reduces the computational complexity of the standard medoid-shift further to $\mathcal{O}(L_{\text{red}}^2 + L_{\text{red}}L + L_{\text{red}}^3)$

4.3 Convergence Properties of the Derived Algorithms

It remains to prove that the mean-shift algorithm in all its variations will converge. In [CM02], the convergence proof of the standard MS is given. It is extended in the sequel for the cases of the R-MS and the S-MS.

Theorem 1. *If $w_l^{R\text{-}MS} > 0$ holds for the weights from Eq. (4.30) and $\alpha_{\mathcal{I}^*,i}$ holds for the weights from Eq. (4.37) and the kernel function $K(\cdot)$ has a convex and monotonically decreasing profile, the sequence of trajectory points $\{\mathbf{x}^{(j)}\}$, $j = 1, 2, \ldots$ from Eq. (4.34) and Eq. (4.49) will converge and the sequence $\{\hat{f}_{\mathcal{K}}(\mathbf{x}^{(j)})\}$, $j = 1, 2, \ldots$ is monotonically increasing.*

The proof given in the sequel is a generalised version of Theorem 1 in [CM02]. It needs to be ensured that the weights w_l will fulfill $w_l^{\text{R-MS}} > 0$ for R-MS and $\alpha_{\mathcal{I}^*,i} > 0$ for S-MS. $w_l^{\text{R-MS}}$ are derived using the KIRWLS and are, thus, certain to follow the required property for monotone loss functions $\rho(\cdot)$. The derivation of $\alpha_{\mathcal{I}^*,i}$ includes a normalisation step to ensure that the weights follow this property, as presented in Sec. 4.2.3.

Proof of convergence of the mean-shift algorithm with weights w_l

Proof. Since L is finite, the sequence $\{\hat{f}_{\mathcal{K}}(\mathbf{x}^{(j)})\}$, $j = 1, 2, \ldots$ is bounded. It will be shown that for $\mathbf{x}^{(j)} \neq \mathbf{x}^{(j+1)}$: $\hat{f}_{\mathcal{K}}(\mathbf{x}^{(j)}) < \hat{f}_{\mathcal{K}}(\mathbf{x}^{(j+1)})$. Without loss of generality $\mathbf{x}^{(j)} = 0$ is assumed, K being the profile of the kernel $\mathcal{K}(\cdot)$ [CM02]

$$\hat{f}_{\mathcal{K}}\big(\mathbf{x}^{(j+1)}\big) - \hat{f}_{\mathcal{K}}\big(\mathbf{x}^{(j)}\big) = \frac{1}{h}\sum_{l=1}^{L} w_l \times \big[K\big(h^{-1}\|\mathbf{x}^{(j+1)} - \mathbf{x}_l\|\big) - K\big(h^{-1}\|\mathbf{x}_l\|\big)\big] \tag{4.54}$$

For convex profiles and $x_2 \neq x_1, x_1, x_2 \in [0,\infty)$ it follows $k(x_2) \geq k(x_1) + k'(x_1)\,(x_2 - x_1)$, such that

$$\begin{aligned}\hat{f}_{\mathcal{K}}\big(\mathbf{x}^{(j+1)}\big) - \hat{f}_{\mathcal{K}}\big(\mathbf{x}^{(j)}\big) &\geq -\frac{1}{h}\sum_{l=1}^{L} w_l \cdot K'\big(h^{-1}\|\mathbf{x}_l\|\big) \\ &\quad \times \Big[||\mathbf{x}_l||^2 - \big|\big|\mathbf{x}^{(j+1)} - \mathbf{x}_l\big|\big|^2\Big] \\ &= -2\frac{1}{h}\cdot \mathbf{x}^{T(j+1)} \sum_{l=1}^{L} w_l \mathbf{x}_l \cdot K'\big(h^{-1}\|\mathbf{x}_l\|\big) \\ &\quad + \frac{1}{h}\sum_{l=1}^{L} w_l \big|\big|\mathbf{x}^{(j+1)}\big|\big|^2 \cdot K'\big(h^{-1}\|\mathbf{x}_l\|\big) \\ &= -\sum_{l=1}^{L} w_l \big|\big|\mathbf{x}^{(j+1)}\big|\big|^2 \cdot K'\big(h^{-1}\|\mathbf{x}_l\|\big)\ . \end{aligned} \tag{4.55}$$

Since $K(x)$ is monotonically decreasing, $-K'(x) \geq 0$ for $x \in [0,\infty)$. $\sum_{l=1}^{L} -K'(h^{-1}\|\mathbf{x}_l\|)$ is strictly positive, as are the weights w_l. Hence $\hat{f}_{\mathcal{K}}\big(\mathbf{x}^{(j+1)}\big) - \hat{f}_{\mathcal{K}}\big(\mathbf{x}^{(j)}\big) > 0$ and the sequence $\{\hat{f}_{\mathcal{K}}\big(\mathbf{x}^{(j)}\big)\}$ is convergent.

Without assuming $\mathbf{x}^{(j)} = 0$ and reformulating Eq. (4.54)

$$\begin{aligned}\hat{f}_{\mathcal{K}}\big(\mathbf{x}^{(j+1)}\big) - \hat{f}_{\mathcal{K}}\big(\mathbf{x}^{(j)}\big) &\geq \\ &= -\frac{1}{h}\sum_{l=1}^{L} w_l \big|\big|\mathbf{x}^{(j+1)} - \mathbf{x}^{(j)}\big|\big|^2 \cdot K'\big(\|h^{-1}(\mathbf{x}^{(j)} - \mathbf{x}_l)\|^2\big)\ . \end{aligned} \tag{4.56}$$

Since $\hat{f}_{\mathcal{K}}\big(\mathbf{x}^{(j+1)}\big) - \hat{f}_{\mathcal{K}}\big(\mathbf{x}^{(j)}\big)$ converges to zero, then $\big|\big|\mathbf{x}^{(j+1)} - \mathbf{x}^{(j)}\big|\big|^2$ also converges to zero and $\mathbf{x}^{(j)}$,$j = 1, 2, \ldots$, is a Cauchy sequence [2]. □

[2] A sequence is called a Cauchy sequence if the terms of the sequence eventually all become arbitrarily close to one another. That is, given a positive real number $\epsilon > 0$ there exists an integer N such that for all positive integers $m, n > N$ the distance $d(x_m, x_n) < \epsilon$.

Extension of the proof of convergence for the medoid-shift with weights w_l

Proof. The choice of successive points in the medoid-shift algorithm, as given in Eq. (4.52) and Eq. (4.53), is carried out according to

$$\begin{aligned}&\sum_{l=1}^{L}\frac{1}{h}\left|\left|\mathbf{x}^{(j)}-\mathbf{x}_l\right|\right|^2 w_l K'\big(h^{-1}\|\mathbf{x}^{(j)}-\mathbf{x}_l\|\big) \\ &>\sum_{l=1}^{L}\frac{1}{h}\left|\left|\mathbf{x}^{(j+1)}-\mathbf{x}_l\right|\right|^2 w_l K'\big(h^{-1}\|\mathbf{x}^{(j)}-\mathbf{x}_l\|\big)\ ,\end{aligned} \tag{4.57}$$

as equality of these two terms indicates convergence. This can be reformulated as

$$\begin{aligned}&\sum_{l=1}^{L}\frac{1}{h}w_l K'\big(h^{-1}\|\mathbf{x}^{(j)}-\mathbf{x}_l\|\big) \\ &\times\left[\left|\left|\mathbf{x}^{(j)}-\mathbf{x}_l\right|\right|^2-\left|\left|\mathbf{x}^{(j+1)}-\mathbf{x}_l\right|\right|^2\right] \\ &=-\sum_{l=1}^{L}\frac{1}{h}w_l\left|\left|\mathbf{x}^{(j+1)}-\mathbf{x}^{(j)}\right|\right|^2 K'\big(h^{-1}\|\mathbf{x}^{(j)}-\mathbf{x}_l\|\big)\ >0,\end{aligned} \tag{4.58}$$

□

proving that $\hat{f}_{\mathcal{K}}\big(\mathbf{x}^{(j+1)}\big) > \hat{f}_{\mathcal{K}}\big(\mathbf{x}^{(j)}\big)$. Hence the sequence $\{\hat{f}_{\mathcal{K}}\big(\mathbf{x}^{(j)}\big)\}, j = 1, 2, \ldots$ is strictly positive and thus $\mathbf{x}^{(j)} \neq \mathbf{x}^{(j+c)}$, for all $c > 0$. This proves that there are no cycles and medoid-shift will converge.

4.4 From Mean-Shift Segmentation to Relative Remission Estimation

Section 4.2 presented the different variants of the mean-shift and medoid-shift algorithm. What is common between all algorithms is that they all produce a set of $\hat{k}$ estimated mode locations $\hat{\mathbf{y}}$. Typically, $\hat{k}$ lies between 2 and 6, depending on the degree of contamination of the images, the stage of the chemical reaction and the underlying glucose concentration. In all cases, mode locations describing the background and the ROI are obtained. Often a transition region between both is detected as a third cluster. Other clusters can be attributed to artefacts or smaller image areas, or clustering errors. The question remaining is: which mode location corresponds to the ROI? To this end, the following approach is used driven by given knowledge of the data.

- **Step 1.** Compute $\mathbf{H}$ and $\mathcal{K}_{\mathbf{H}}$ as described in Sec. 4.2.1.
- **Step 2.** Select a mean-shift variation to perform the segmentation and mode estimation on $\mathbf{x}^{(n)}$ as described in Sec. 4.2. If medoid-shift is used the final set of estimated mode locations $\hat{\mathbf{y}}$ is obtained. Otherwise, if mean-shift is used
  - a set of preliminary modes $\tilde{\mathbf{y}}$ is obtained.
  - Mode pruning as described in Sec. 4.2 is performed to obtain the final set of estimated mode locations $\hat{\mathbf{y}}$.
- **Step 3.** Calculate cluster weights w_c for each cluster by computing the ratio of data points assigned to each cluster to the total number of data points.
- **Step 4.** Discard clusters with $w_c < 0.1$.
- **Step 5.** Select the mode location estimate with the smallest value to be the relative remission estimate as $\hat{r}(n) = \hat{y}_{\text{ROI}} = \{\hat{y}_i | \arg\min_y \hat{\mathbf{y}}\}$.

In Step 4, clusters with weights smaller than 10% are discarded, because the ROI typically takes up at least 40%. If not, the ROI is too small to accurately estimate the underlying glucose concentration and the measurement is classified as faulty and is dealt with as presented in Sec. 3.2.3. After discarding insignificant regions, three clusters typically remain corresponding to the background, the ROI and the transition region. Due to the chemical reaction process described in Sec. 2.4, the ROI exhibits the highest colour change and hereby, the smallest intensity value. This motivates Step 5 of the described approach. The approach is repeated for all incoming frames, gradually building the relative remission kinetic curve $\hat{r}(n)$ and it is terminated when either the last frame of the measurement is obtained or when saturation of the chemical reaction is detected as portrayed in Sec. 3.2.5.

4.5 Experimental Results and Discussion

This section validates the methods developed in this chapter. First, as a proof of concept the standard mean-shift (MS) algorithm is used to perform segmentation and relative remission estimation of the glucose images. Hereby, the bandwidth selection problem is highlighted. Next, the robust mean-shift (R-MS) algorithm is validated both on real and contaminated simulation data proving its robustness compared to the MS.

The sparse mean-shift algorithm (S-MS) is tested on real data, elucidating the computation gain resulting from it. Next, the performance of the medoid-shift (MedS) and its variations, the robust medoid-shift (R-MedS) and the sparse medoid-shift (S-MedS), is demonstrated. Finally, a comparison of all methods is presented w.r.t. precision of the relative remission estimates $\hat{r}$ and accuracy of the glucose concentration estimates $\hat{g}$ using all sets from Table 2.1. An extensive study using the real data sets allows making inference on the effect of the blood drop volume and the resolution Υ used in the setup.

The validation methods used are highlighted in Sec 2.4.3. Recall that according to the most recent ISO standards [Int13], the maximal permissible error is $\pm 15\,\mathrm{mg/dl}$ within a reference range of $0\,\mathrm{mg/dl}$ to $75\,\mathrm{mg/dl}$ and 20% for a reference range above $75\,\mathrm{mg/dl}$.

4.5.1 The Standard Mean-Shift Algorithm for the Glucose Estimation Problem

Table 4.1 highlights the parameters used for this experiment. The dimension of the data d indicates wether the spatial components of the data vectors have been taken into account or not. If $d = 1$, only the intensity has been considered, and if $d = 3$, the spatial components have been taken into account and the kernel has been designed according to Eq. (4.10).

Table 4.1: Simulation parameters for the MS experiment.

$\mathcal{K}_{\mathbf{H}}(\mathbf{x})$	d	$\mathbf{H}$	ϵ
Gaussian (Eq. (4.9))	1	Eq. (4.16)	0.05

The MS algorithm is first validated by applying it to a simulated data set for 100 runs/glucose value. The simulation data set is developed by creating a mask from the real data set using a frame at $n = n_D$. The area of the test strip is modulated with the model of the chemical reaction given in Eq. (3.23). To this end, $N_g = 20$ different glucose concentrations are used, resulting in synthetic measurement videos of the same size as the real data videos. Additive zero-mean Gaussian noise with a variance σ^2_{syn} is estimated from the real data samples. The aim of using a synthetic data set is to validate the performance with an available ground truth. Figure 4.5 presents the resulting relative remission estimates $\hat{r}$ compared to the true values r for different glucose values. The total mean-squared error (MSE) is equal to MSE = 0.001.

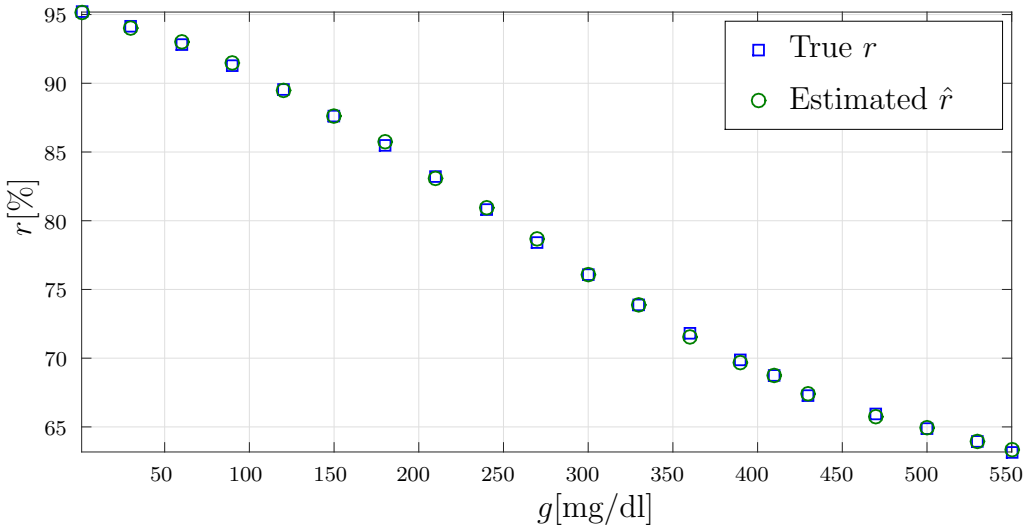
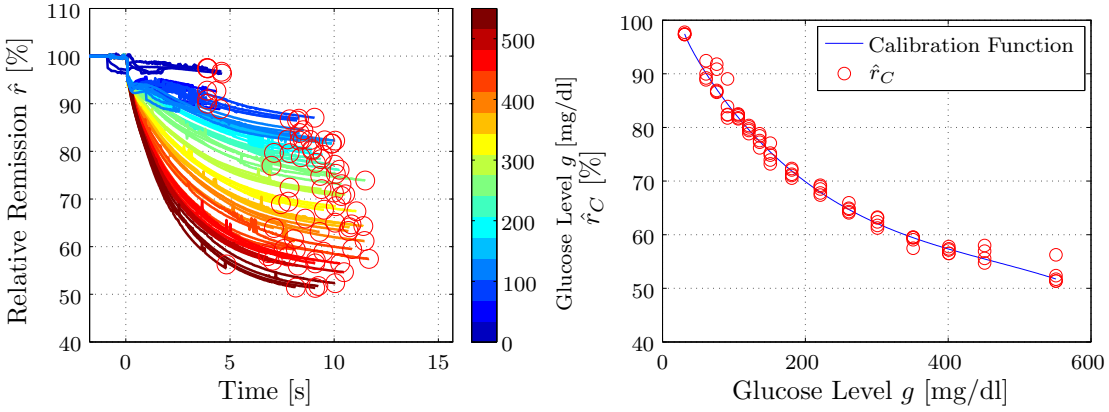


Figure 4.5: Relative remission estimates $\hat{r}$ for 20 different glucose concentrations, performed using simulated data and compared to the ground truth.

Next, the real data Set C from Table 2.1 is used to validate the MS algorithm in a real photometric measurement scenario, using the parameter setup from Table 4.1. Figure 4.6 (a) illustrates the kinetic curves produced using MS, and the relative remission convergence estimates $\hat{r}_C^{(\text{Stand. Conv})}$ using the standard criterion from Sec. 3.2.5. Overall, the expected progression of the kinetic curve is established with its three typical stages as in Sec. 2.4.1. Measurements of the same underlying glucose concentration are, mostly, bundled together. Convergence is reached faster for low glucose concentrations than for high glucose concentrations. Moreover, the assumption made in Sec. 3.2.5 that the decay rate τ of the reaction is correlated with the underlying glucose concentration is confirmed. The higher the underlying glucose concentration, the steeper the decay. The calibration curve and the converged estimates are presented in Figure 4.6 (b), illustrating the variance of measurements of the same underlying glucose concentration. Evidently, the measurements for glucose concentrations between 60 mg/dl to 90 mg/dl are the most problematic. The reason for this is that measurements in this glucose region do not show high contrast and are quite difficult to segment.

To quantify the results, the coefficient of variation $\text{CV}_{\hat{r}}$ (from Sec. 2.4.3) is presented in Table 4.2, comparing MS to the application of Gaussian Mixture Models (GMM) to segment the frames and estimate the mean of the ROI cluster. GMM is solved using the EM algorithm, initialising the number of clusters k with a fixed number of clusters, as well as with an estimate of the cluster number obtained using



(a) (b)

Figure 4.6: (a) Kinetic curves of all measurements in Set C estimated using the MS algorithm. The red circles signify convergence points $\hat{r}_C^{(\text{Stand. Conv})}$. (b) The converged relative remission estimates $\hat{r}_C$ and the corresponding fitted calibration curve for Set C estimated using the MS algorithm.

Table 4.2: Comparison of the $\text{CV}_{\hat{r}}$ for MS and GMM using a fixed model order of $k = 3$ (GMM-3), a fixed model order of $k = 4$ (GMM-4) and BIC to estimate the model order (GMM-BIC). The optimal value is indicated by the bold font.

	MS	GMM-3	GMM-4	GMM-BIC
$\text{CV}_{\hat{r}}$	**1.15**	1.27	1.36	1.4

the Bayesian Information Criterion (BIC). Clearly, MS outperforms GMM.

Figure 4.7 compares Clarke's Error Grid Analysis (from Sec. 2.4.3) for Set C using MS for the fixed and adaptive bandwidth case, and for including the spatial component. Contrarily to the expectations, including the spatial weighting as proposed in Eq. (4.10) impairs the performance of MS for the glucose problem especially for low- and mid-level glucose concentrations.

This can be seen when comparing the plots on the right side (including spatial information) to the ones on the left side (excluding spatial information). The reason for this, is that the ROI in the glucose images is not necessarily spatially connected. Including a spatial weighting in this case, gives higher weights to data points in the near proximity, which may be misleading as they may not belong to the ROI.

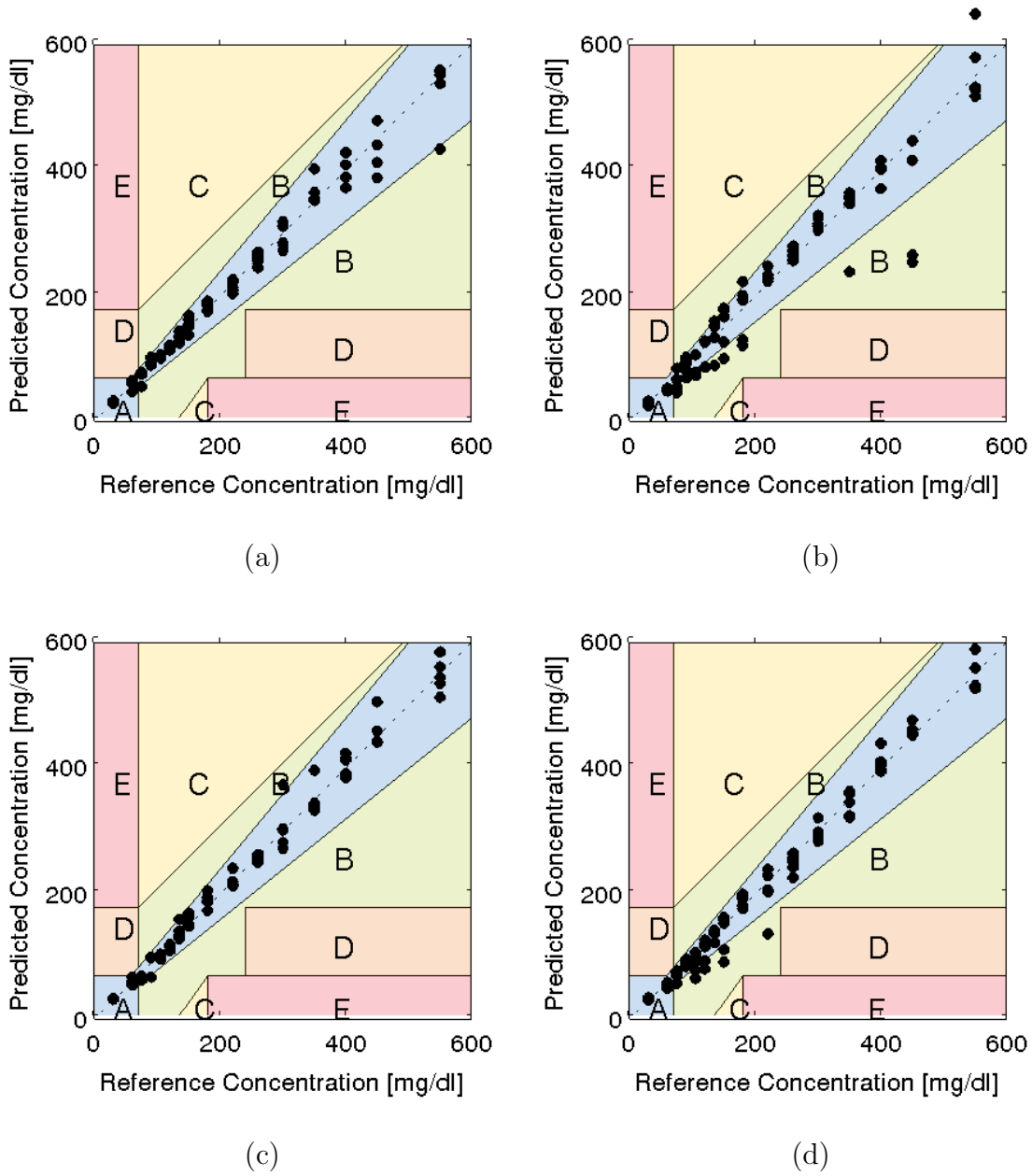


Figure 4.7: Clarke's Error Grid analysis for Set C using MS (a) with a fixed bandwidth h_0 and excluding spatial information ($d = 1$), (b) with a fixed bandwidth h_0 including spatial information ($d = 3$), (c) with an adaptive bandwidth $\mathbf{h}(\mathbf{x})$ and excluding spatial information and (d) with an adaptive bandwidth $\mathbf{h}(\mathbf{x})$ and including spatial information ($d = 3$).

This is specially prominent in lower glucose concentrations more than higher ones, because in the higher concentrations the intensity difference between pixels inside the ROI and those in other areas is large and the effect of spatial weighting is negligible. The adaptive bandwidth for the intensity manages to improve the results of the spatial

bandwidth, particularly for the mid-level glucose range.

Comparing the upper plots to the lower ones, the effect of using an adaptive bandwidth for the intensity component can be seen. Evidently, the adaptive bandwidth seems to improve the accuracy of high glucose measurements slightly and has not much influence on mid-level and low glucose measurements. The reason for this is that high glucose measurements are characterised by a multi-modal pdf, and the adaptive bandwidth is able to deal with this. Due to these results spatial weighting will not be considered for the remainder of the experiments. Furthermore, a fixed bandwidth will be used as the difference in accuracy is minimal, however the adaptive bandwidth requires higher computation due to the calculation of the pilot estimator of the pdf in Eq. (4.17).

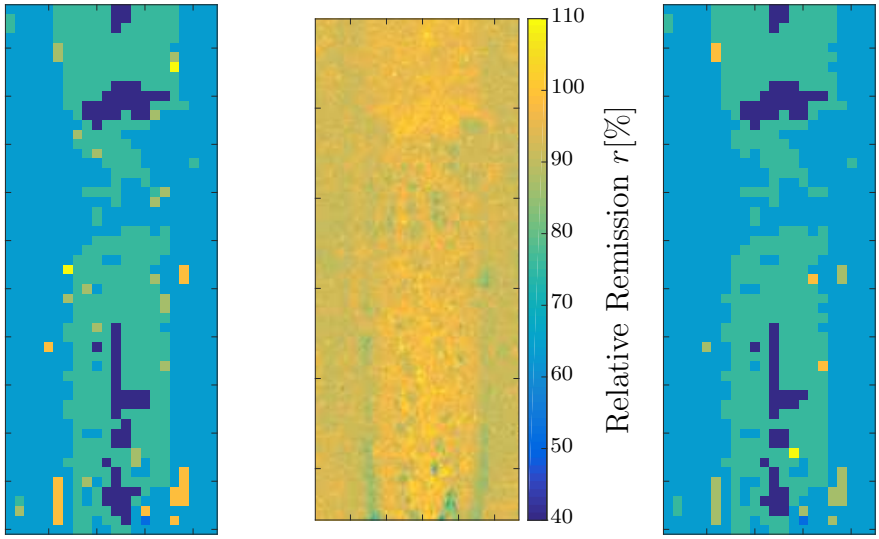
Segmentation results of MS are portrayed in Fig. 4.8 for a low glucose frame and a high glucose frame. Evidently, the algorithm succeeds at delineating the different image regions. In the low glucose case, the blood flow is not homogenous over the whole test strip and this behaviour is caught well in the segmentation. In the high glucose case, the bridge is well segmented as well as the transition region from the ROI to the background. The artefact in the lower left part of the frame is also identified and segmented.

4.5.2 The Robust Mean-Shift Algorithm

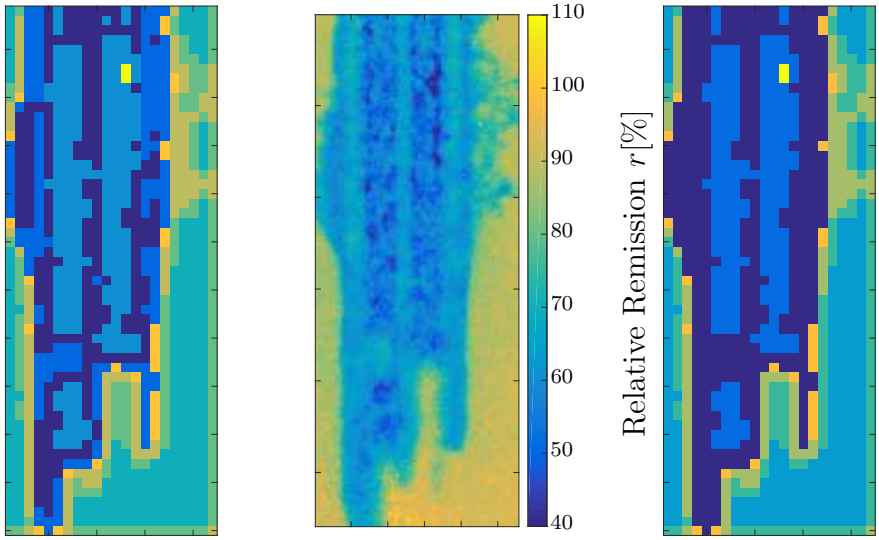
Table 4.3: Simulation parameters for the R-MS experiment.

$\mathcal{K}_{\mathbf{H}}(\mathbf{x})$	d	$\mathbf{H}$	$\rho(\cdot)$	a_{Huber}	ϵ
Gaussian (Eq. (4.9))	1	Eq. (4.16)	Huber (Eq. (4.24))	0.5	0.01

First, the robustness of the R-MS algorithm is validated by means of 1-dimensional unimodal and bimodal contaminated data modelled as in Eq. (4.22). The parameters used for the experiments in this section are given in Table 4.3. For the unimodal case, the nominal distribution f_N follows $\mathcal{N}(\mu_1, \sigma^2)$. The contaminating distribution f_C follows $\mathcal{N}(\mu_{C1}, \kappa\sigma^2)$, where κ is the strength of the impulsive component.



(a) Glucose frame with $g = 30\,\mathrm{mg/dl}$.



(b) Glucose frame with $g = 350\,\mathrm{mg/dl}$.

Figure 4.8: From left to right: The segmentation result of MS, the original glucose frame and the segmentation result of R-MS for (a) 30 mg/dl glucose concentration and (b) 350 mg/dl glucose concentration.

For the bimodal data, f_N follows $\frac{1}{2}\mathcal{N}(\mu_1, \sigma^2) + \frac{1}{2}\mathcal{N}(\mu_2, \sigma^2)$, while f_C follows

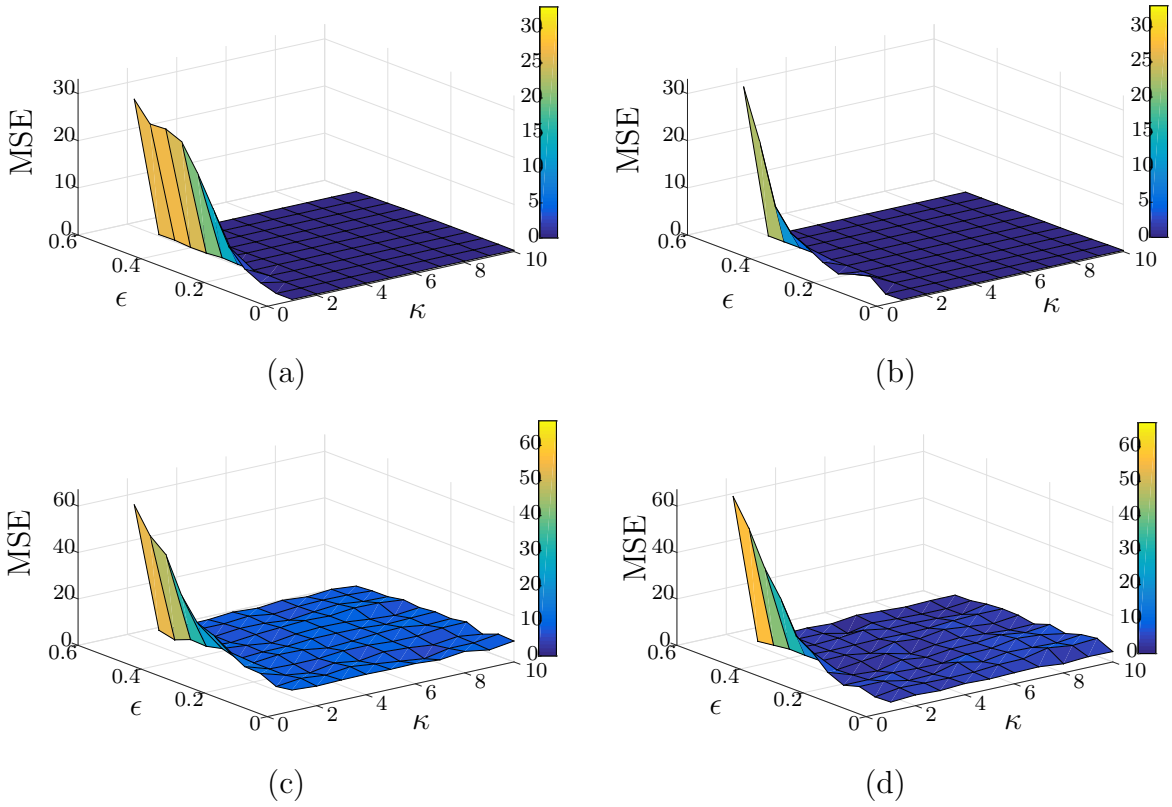


Figure 4.9: Results of (a) MS, (b) R-MS for the unimodal distribution and (c) MS, (d) R-MS for the bimodal distribution.

$\mathcal{N}(\mu_{C2}, \kappa\sigma^2)$. The degree of contamination ε is varied between $0 \leq \varepsilon \leq 0.5$ and κ is varied between $1 \leq \kappa \leq 10$. The error between the actual mode locations (μ_1, μ_2) and the mode location estimates $(\hat{\mu}_1, \hat{\mu}_2)$ are evaluated over 500 Monte Carlo runs for different levels of ε and κ using the mean squared error (MSE). The results of R-MS and the comparison to MS can be seen in Fig. 4.9. It can be observed that the overall performance of R-MS is better than MS. For $\varepsilon = 0$, both R-MS and MS perform in a similar manner, with R-MS slightly outperforming MS. The higher the contamination level the better the performance of R-MS, compared to MS.

Coming to the real data experiments, first, an example of the performance of R-MS is given for one specific contaminated glucose frame. Figure 4.10 presents the histogram of the first glucose frame from Fig. 4.2, which shows contamination in the low intensity region. The performance of the MS is compared to the R-MS w.r.t. mode location estimation. The MS produces a mode location estimate that is biased towards the low intensity region. The R-MS algorithm is able to down-weight the data vectors associated with the contamination and estimate the mode location with higher accuracy.

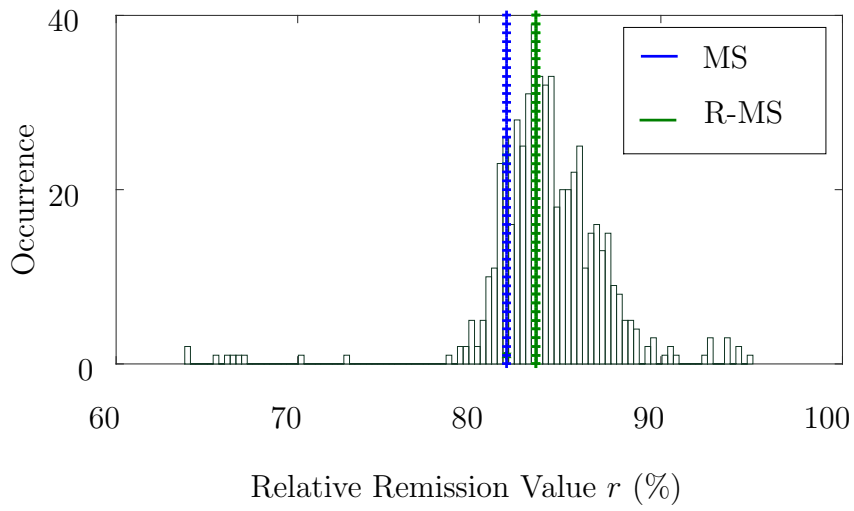


Figure 4.10: Example of the histogram of a contaminated image for a blood glucose sample of 150 mg/dl, as well as mode location estimates for MS (blue dotted line) and R-MS (green dotted line).

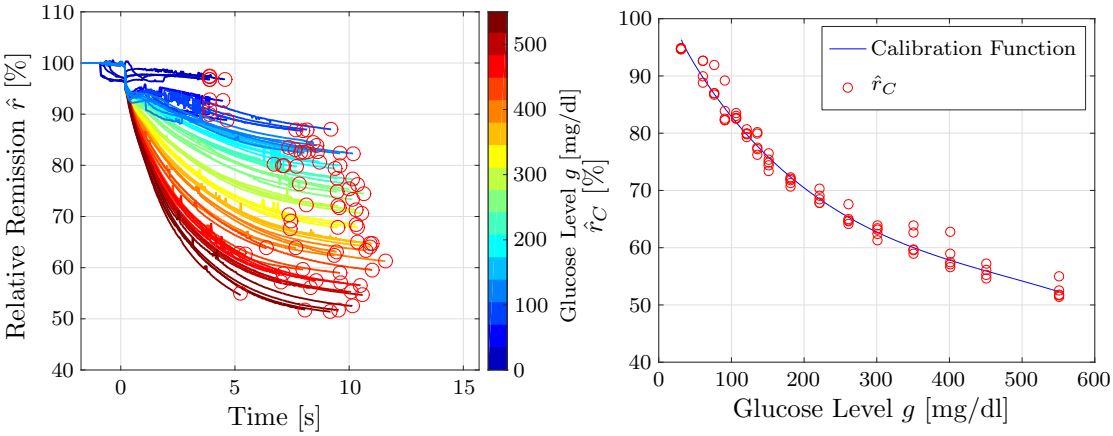
Figure 4.11 (a) presents the kinetic curves resulting from R-MS for Set C from Table 2.1. Here too, the expected progression of the chemical reaction is evident. Figure 4.11 (b) presents the converged relative remission estimates and the fitted calibration curve for Set C.

The segmentation results of RMS for a low and a high glucose case are depicted in Fig. 4.8. The results are similar to the MS case, however it is noticeable that RMS produces less clutter in both cases and a segmentation that makes more sense physically. This behaviour is more evident in the high glucose case where the transition region between the ROI and the background is mainly segmented as one cluster, indicated by the dark blue segment.

4.5.3 The Sparse Mean-Shift Algorithm

The S-MS aims at producing accurate relative remission estimates, using only a fraction of the available data vectors, hereby reducing computation. Table 4.4 lists the parameters used in this experiment.

To this end, first the relative remission estimate $\hat{r}$ of the S-MS is evaluated in comparison with the MS estimate, which is considered the ground truth in this experiment. The objective is to evaluate the effect of using a smaller subset of data



(a) (b)

Figure 4.11: (a) Kinetic curves of all measurements in Set C estimated using the R-MS algorithm. The red circles signify convergence points $\hat{r}_C^{(\text{Stand. Conv})}$. (b) The converged relative remission estimates $\hat{r}_C$, and the corresponding fitted calibration curve for Set C estimated using the R-MS algorithm.

Table 4.4: Simulation parameters for the S-MS experiment.

$\mathcal{K}_{\mathbf{H}}(\mathbf{x})$	d	$\mathbf{H}$	$\rho(\cdot)$	a_{Huber}	ϵ	T_ν
Gaussian (Eq. (4.9))	1	Eq. (4.16)	Huber (Eq. (4.24))	0.5	0.01	0.002

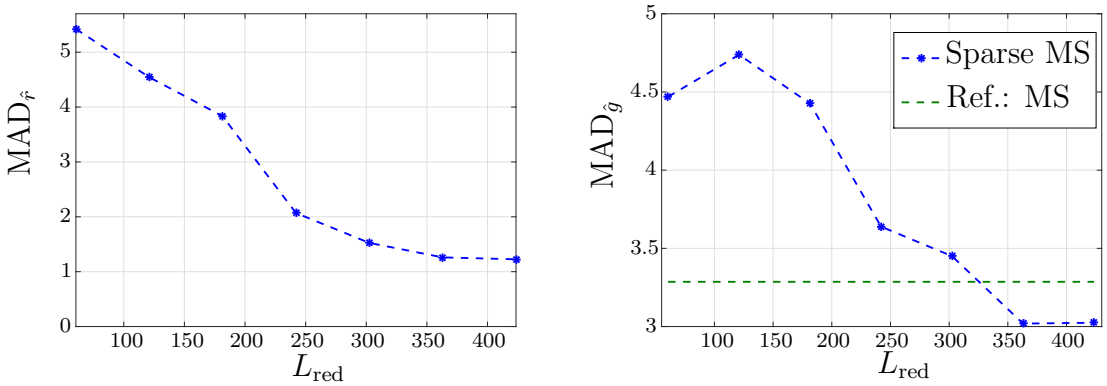
vectors in S-MS, compared to using the full set of data in MS. Following criterion is used to evaluate the error

$$\text{MAD}_{\hat{r}} = \frac{1}{N_M \cdot N} \sum_{k=1}^{N_M} \sum_{n=1}^{N} |\hat{r}_{k,n}^{(\text{MS})} - \hat{r}_{k,n}^{(\text{S-MS})}|, \tag{4.59}$$

where N_M denotes the number of measurements in the data set and N is the number of frames in a measurement.

Figure 4.12 (a) presents the $\text{MAD}_{\hat{r}}$ for different values of L_{red}. Note that the original number of data vectors is $L = 1,210$. The remission values typically variate between $r = 50\,\%$ to $100\,\%$. Observing Fig. 4.12 (a), a deviation of $\approx 10\%$ in remission compared to the MS case can be determined, when using only $L_{\text{red}} = 0.05 \cdot L$ of the data vectors. This deviation sinks to just 2% when using $L_{\text{red}} = 0.3 \cdot L$ of the data.

Next, the accuracy of the glucose estimates $\hat{g}$ using S-MS is compared to the



(a) (b)

Figure 4.12: (a) MAD$_{\hat{r}}$ analysis for different subset cardinalities L_{red}. Recall that the original number of data vectors is $L = 1,210$. (b) MAD$_{\hat{G}}$ analysis for different subset pixel sizes, the MS value is plotted for comparison.

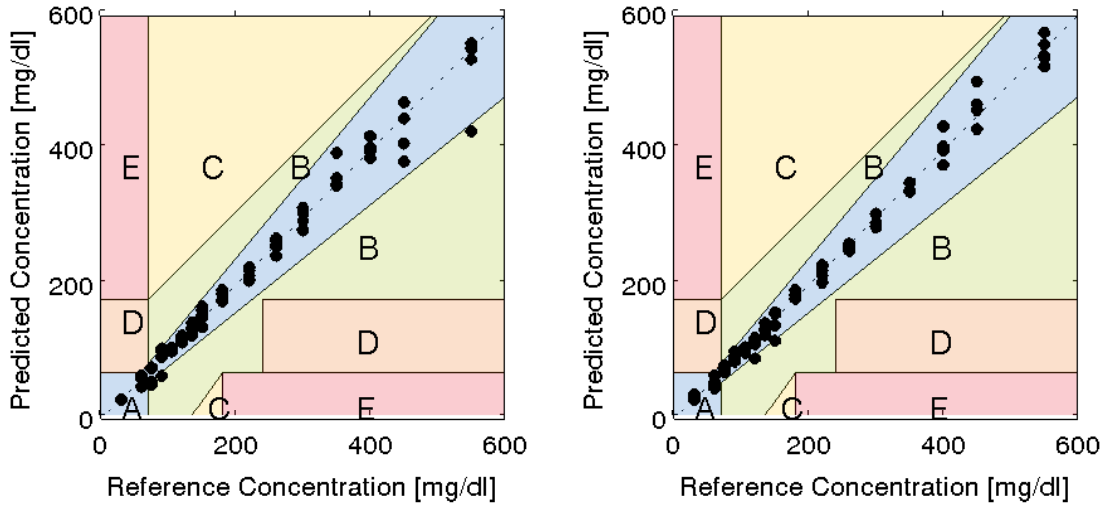
MS case w.r.t. the following criterion

$$\text{MAD}_{\hat{g}} = \frac{1}{N_M} \sum_{k=1}^{N_M} |g_k - \hat{g}_k|. \tag{4.60}$$

The results are presented in Fig 4.12 (b). The MS produces an error of MAD$_{\hat{g}}$ = 3.3 mg/dl, while S-MS achieves errors that lie below MAD$_{\hat{g}}$ = 5 mg/dl for all subset cardinalities L_{red}. Hence, both methods lie within the permissible error range according to the most recent ISO standards [Int13]. Furthermore, it is interesting to observe that for subset cardinalities $L_{\text{red}} > 0.3L \approx 360$, S-MS outperforms MS. This can be explained by the fact that if the subset of data vectors is large enough, the sparse selection of data vectors acts as a noise reduction filter.

The CEG analysis in Fig 4.13 underlines these findings. The results for MS mostly lie in the A-region, with just two measurements lying in the B-region. For S-MS with $L_{\text{red}} = 0.35 \cdot L$ two measurements lie directly on the border between the A-region and the B-region, while all other measurements lie in the A-region and show a lower variance than MS.

Figure 4.14 presents the evolution of the number of iterations j, as well as the computation time t_{comp}, over L_{red} for MS and S-MS. Up to $L_{\text{red}} = 0.5 \cdot L$ the number of iterations, as well as the computation time, for S-MS is lower than for the MS. After that, the overhead needed to calculate $\mathcal{I}$ as well as $\boldsymbol{\alpha}^*_{\mathcal{I}}$ surpasses the benefit of using less data vectors. For $L_{\text{red}} = 0.35 \cdot L$, t_{comp} sinks by 81% and j by 46%.



(a) (b)

Figure 4.13: Clarke's Error Grid analysis for MS, and S-MS for $L_{\text{red}} = 0.35 \cdot L$.

As portrayed in Sec. 4.2.3, a data-driven approach is proposed to select $L_{\text{red},\nu}$ uniquely for every frame. Figure 4.15 shows the choice of $L_{\text{red},\nu}$ for Set C, with $L_{\text{red},\nu}$ averaged over all frames for each measurement. The tendency matches the expectations: very low glucose concentrations are characterised by low contrast and therefore higher coherence. The choice of $L_{\text{red},\nu}$ is higher than for high glucose concentrations, i.e., more data points are needed to reliably represent the image. As the concentration increases and the images become more distinct in the different regions, the variance in the choice of $L_{\text{red},\nu}$ for the same glucose concentrations decreases. The outlier at $g = 60\,\text{mg/dl}$ is related to a faulty measurement, that shows a homogeneous behaviour. To this end, only a small number of data points is chosen as the the homogeneity implies high coherence. Choosing a fixed L_{red} according the worst-case scenario would have resulted in $L_{\text{red}} > 370$ data points, which is much higher than the data-driven choice of L_{red}.

Finally, the quality of the S-MS is analysed when using a data-driven choice of $L_{\text{red},\nu}$ versus a fixed $L_{\text{red},\nu} = 0.37L$. The results are given in Table. 4.5. Evidently, the accuracy of both approaches is comparable. This underlines that a data-driven $L_{\text{red},\nu}$ maintains the accuracy needed.

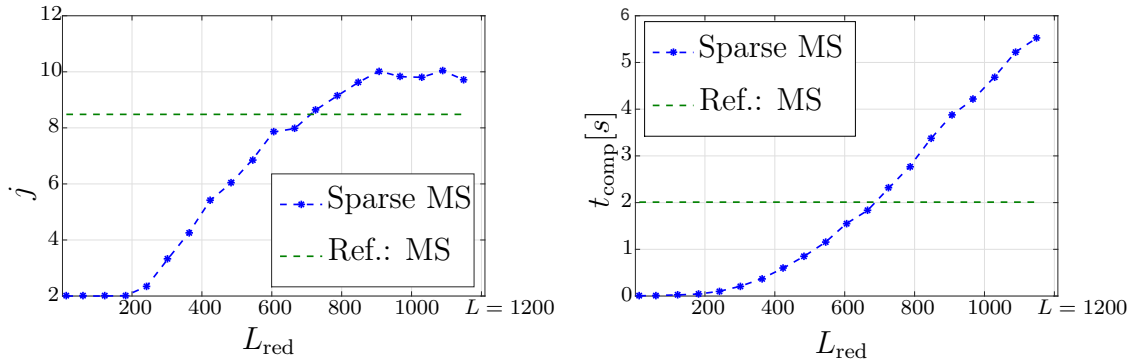


Figure 4.14: (a) Number of iterations j for different values of L_{red} for MS and S-MS. (b) Computation time t_{comp} for different values of L_{red} for MS and S-MS.

Table 4.5: Comparison of the $\text{CV}_{\hat{r}}$ values for a data-driven choice of $L_{\text{red},\nu}$ vs. a fixed $L_{\text{red},\nu} = 0.37L$ using RS-MedS.

Choice of $L_{\text{red},\nu}$	A	B	C	D	E	F
Data-driven	1.03	1.29	0.81	1.90	4.60	1.59
Fixed	0.95	1.38	1.27	1.97	4.35	1.64

4.5.4 The Medoid-Shift Algorithm

The results of the MedS algorithm and its variations, R-MedS and S-MedS, are given in Fig. 4.19. Evidently, the medoid-shift algorithms result in much smoother kinetic curves. As the MedS always chooses points contained in the set of data vectors, the remission estimates do not fluctuate as much as in the mean-shift case. This allows the standard convergence criterion, to detect more reasonable convergence times n_C and remission values $\hat{r}_C$.

The difference between MedS and R-MedS is not so evident from these images, and a quantitative evaluation, as is given in Table 4.6, is necessary. The S-MedS result in Fig. 4.19 (e) shows much stronger fluctuation which is expected in the sparse case, as k-center algorithm is calculated again for each frame, always choosing a different subset of data vectors. Evidently, the variance of estimates for the S-MedS in Fig. 4.19 (f) case increases compared to MedS and R-MedS. The quantitative results will show, that it is still in the permissible error range.

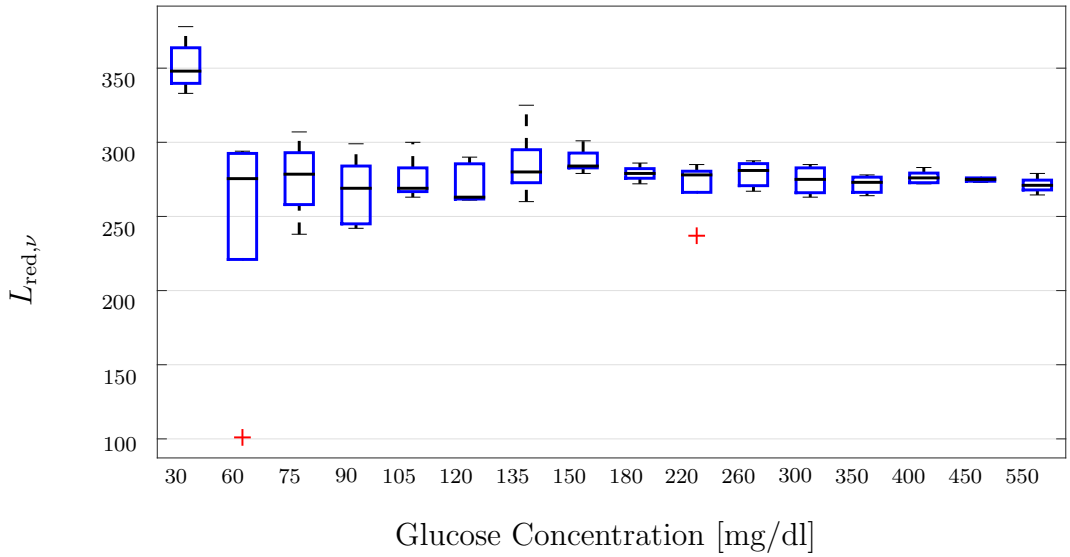
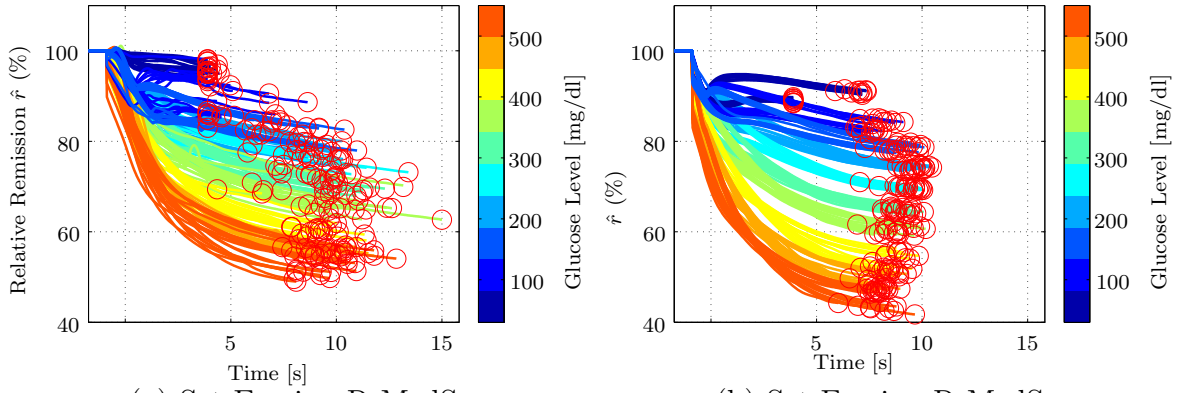


Figure 4.15: The choice of $L_{\text{red},\nu}$ for different glucose concentrations for Set C. The red crosses indicate outliers.

4.5.5 Comparison of All Mean-Shift Variants

First, the quality of the remission results is analysed. Herein, a comparison of the segmentation methods is given, as well as an analysis of the different data sets. Then, the precision of the glucose estimates after the mapping operation is studied. Figure 4.16 presents the estimated kinetic curves using R-MedS, for data sets E and



(a) Set E using R-MedS (b) Set F using R-MedS

Figure 4.16: Kinetic curves for Set E and Set F using R-MedS. The red circles signify convergence points $\hat{r}_C^{(\text{Stand. Conv})}$.

F. These are obtained using the exact same setup and blood samples, however using a

blood sample volume of ≈ 1 nl for Set E and a blood sample volume $\gg 100$ nl for Set F. A significant performance decrease can be observed when comparing Set E with Set F. While Set F shows smooth kinetic curves, with same underlying glucose measurements bundled together, Set E shows an overlap of measurements with different underlying glucose concentrations.

Table 4.6: $CV_{\hat{r}}$ values for the different data sets using the standard convergence criterion.

Set	MS	R-MS	MedS	R-MedS	S-MS	RS-MS	S-MedS	RS-MedS
A	1.59	1.29	**1.17**	1.22	1.87	1.81	**1.52**	**1.52**
B	1.95	1.83	1.82	**1.71**	2.4	2.1	**1.95**	**1.95**
C	1.15	1.51	0.90	**0.99**	1.50	1.41	**1.39**	**1.39**
D	1.87	1.7	1.58	**1.39**	1.87	2.90	1.10	**1.09**
E	3.98	4.17	3.39	**3.25**	5.9	5.6	4.3	**4.2**
F	1.48	1.37	1.10	**1.04**	2.44	2.25	1.99	**1.78**

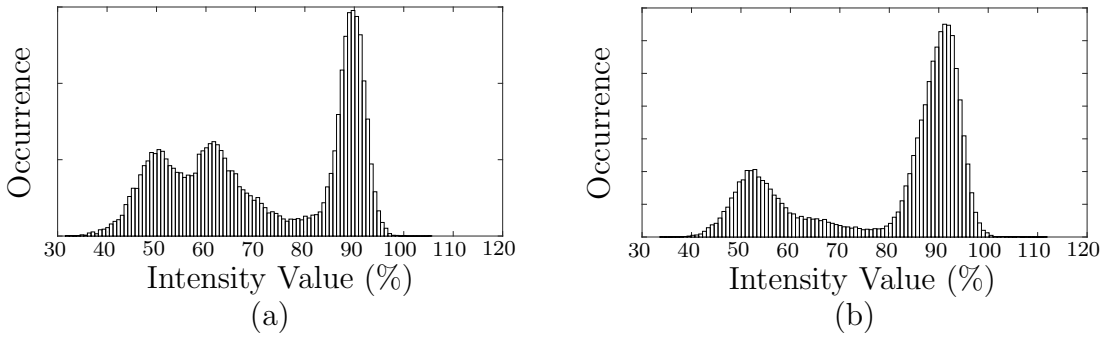


Figure 4.17: Histograms of converged images of high glucose range measurements from (a) Set B and (b) Set C.

These results are underlined when analysing the resulting $CV_{\hat{r}}$ values for these two sets in the left part of Table 4.6. This indicates that blood sample volumes in the range of 1 nl could be problematic. The volume range from 10 nl to 100 nl, in contrast, seems to perform well.

Table 4.6 shows that the medoid-shift versions are significantly better than the mean-shift versions for all sets. The difference in performance between MS and MedS is not as high for Set B as for other sets. The reason can be understood from

Fig. 4.17. The more distinct the modes are, the more separable the clusters, the better mean-shift performs. By choosing the medoid instead of the mean and converging to actual points in the data set, the medoid-shift doesn't show this bias in estimators.

The results for the sparse mean-shift are given in the right part of Table 4.6. The resulting $CV_{\hat{r}}$ values are only slightly worse than their non-sparse counterparts. This degradation, however, is not too severe and the loss in accuracy can be accepted to ensure lower computation. Only for Set E can a severe degradation in performance be noticed, when using the S-MS algorithm. This can be traced back to the fact that due to the very small blood volume, the ROI is much smaller than for the other measurements. Furthermore, including robust weights almost always improves the performance of the sparse algorithm versions. This is intuitive in the sense that when choosing a smaller subset of the data using the k-center algorithm, outliers are easily included in the data set. Subsequent down-weighting of these using the robust methods improves the performance.

Finally, the results of the glucose estimates $\hat{g}$ are evaluated. Table 4.7 presents

Table 4.7: gMAD values using the standard convergence criterion.

.

	Low Glucose (≤ 75 mg/dl)				High Glucose (> 75 mg/dl)			
Set	MS	R-MS	MedS	R-MedS	MS	R-MS	MedS	R-MedS
B	14.4	13.0	**6.6**	**6.6**	13.4	**11.8**	17.0	17.2
C	6.9	6.9	4.6	**4.0**	12.8	14.4	**7.8**	8.9
E	22.6	29.4	24.4	**21.0**	30.8	37.4	25.5	**23.6**
F	9.3	9.4	8.5	**8.1**	17.5	17.0	**16.6**	18.4

the mean gMAD values for $\hat{g}$ of each set separated according to the glucose range. This analysis is omitted for Sets A and D, as they do not contain enough different glucose concentrations to construct the mapping function. The results for most sets lie below the ISO requirements, i.e., low glucose ranges show errors smaller than gMAD $= \pm 15$ mg/dl and high glucose ranges show an error larger than 20%. However, in Set E the limit is often exceeded, confirming again the fact that blood sample volumes around 1 nl are problematic.

The previously made observation concerning medoid-shift outperforming mean-shift is confirmed again here. The robust version seems to improve the performance in low glucose ranges more than in high glucose ranges. In most cases the robust

formulations of the algorithms perform better than their non-robust counterparts. Even when the contamination is not as strong as in the example of Fig. 4.10, the robust formulations outperform. This is due to the finite data size, in which the nominal assumptions do not always fully hold.

Finally, Fig. 4.18 portrays the computation time needed to segment one glucose frame, using the different mean-shift variations. In a real system, an efficient hardware implementation will lead to much smaller computation times. The computation times presented here are only meant to give an indicator on the computational complexity of the different methods. Evidently, mean-shift outperforms medoid-shift in computation. Furthermore, the sparse variations lead to a drastic decrease in computation time.

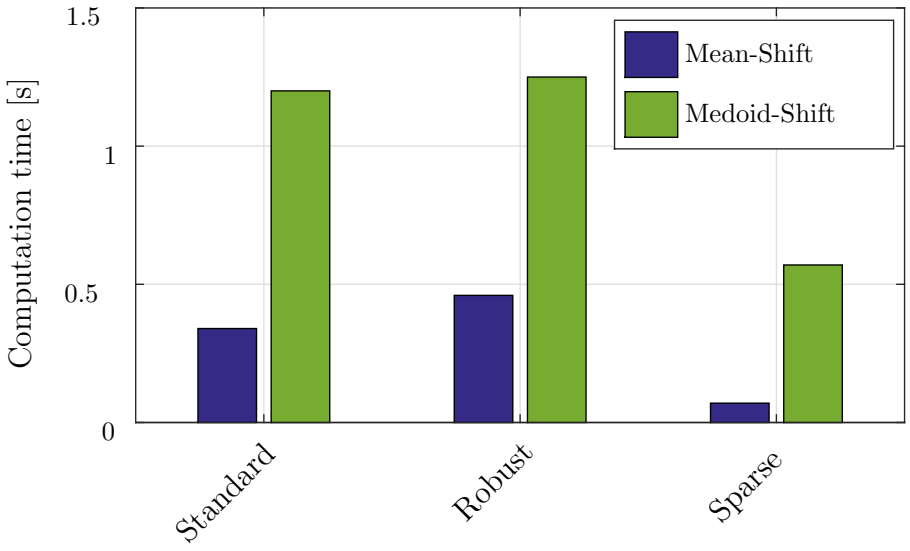


Figure 4.18: Computation time for the segmentation of one image, for the different mean-shift variations.

4.6 Summary and Conclusions

This chapter has presented the mean-shift and the medoid-shift algorithm for segmentation of glucose images and subsequent estimation of the relative remission value, specifically highlighting the bandwidth selection problem. A robust formulation and a sparse formulation of each algorithm have been derived using the Reproducing Kernel Hilbert Space (RKHS) to reformulate the kernel density estimator, which is

the basis of the mean-shift concept. For the sparse formulation, a criterion has been developed to choose an appropriate subset cardinality. Convergence proofs for all derived algorithms have been presented. An approach has been presented to identify the ROI from the segmented frame and the corresponding relative remission estimate.

The algorithms have been validated using both simulated and real data, as well as compared to each other. The results elucidate that mean-shift succeeds at segmenting the glucose frames and estimating the underlying relative remission value of the region of interest, asserting the capability of the mean-shift concept in problems where the number and shape of clusters are unknown. Robust mean-shift outperforms standard mean-shift in the presence of heavy-tailed noise. Sparse mean-shift is able to reduce the computation by selecting a subset of data vectors, while only compromising slightly in accuracy. In some cases, when the cardinality of the subset is large enough, the results show that the sparse mean-shift can even lead to a gain in accuracy as it acts as a noise reduction method. It has been shown that the combination of the robust and sparse approaches improves the performance of the sparse methods. The reason for this is that the choice of a subset of data points on the basis of incoherence leads to choosing points at the tails of the distribution. Combining the sparse approach with the robust approach leads to down-weighting these points. The medoid-shift approach and its variations almost always outperform the mean-shift approaches. A study of computational complexity has shown that mean-shift is more efficient than medoid-shift. The sparse versions evidently show a drastic gain in computational efficiency.

Finally, results using different data sets with different blood sample volumes have established that the use of very small blood samples around 1 nl seem to be at the limit of what is acceptable in accuracy.

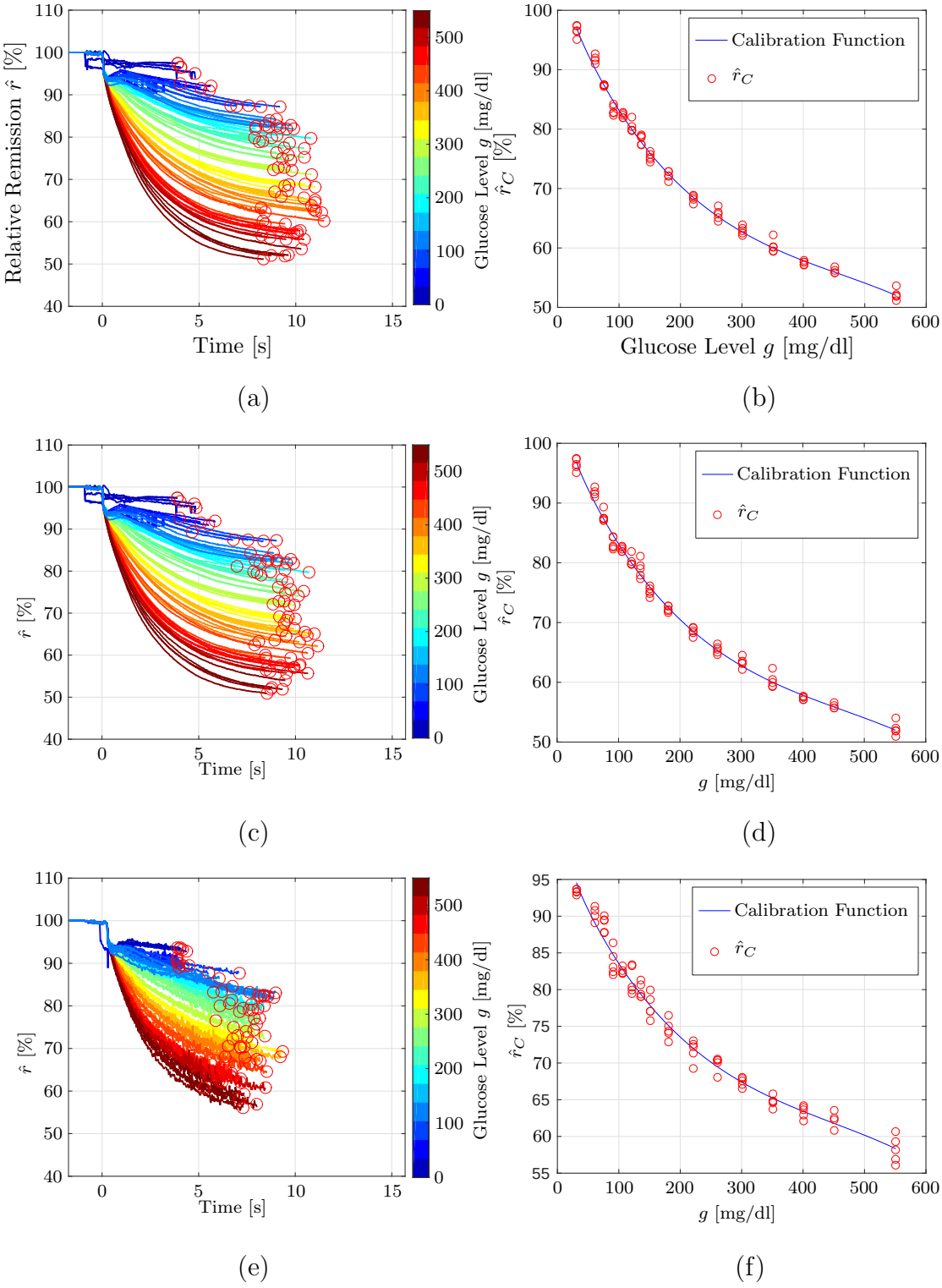


Figure 4.19: Kinetic curves of all measurements in Set C estimated using (a) MedS, (c) R-MedS, and (e) S-MedS. The red circles signify convergence points $\hat{r}_C^{(\text{Stand. Conv})}$. The converged relative remission estimates $\hat{r}_C$, and the corresponding fitted calibration curve for Set C estimated using (b) MedS, (d) R-MedS and (f) S-MedS.

Chapter 5

Segmentation Using the Union of Subspaces Model

This chapter presents the second segmentation approach developed in this thesis: segmentation based on the union of subspaces (UoS) model. As opposed to the mean-shift methods presented in Chapter 4, the approaches presented in this chapter take into account the temporal evolution of the frames to perform the segmentation.

A short review on the state of the art in UoS methods is presented in Sec. 5.2. The main contributions of this chapter are given as follows: (i) the adoption of sparse subspace clustering and low-rank representation for the glucose segmentation problem. (ii) the development of two novel sparse subspace clustering approaches in Sec. 5.3.2 that are specifically suited to segment temporal video data such as the glucose data. The first proposed model, termed Ordered Robust Subspace Recovery (ORoSuRe), assumes that all data vectors may be corrupted by noise, which is a reasonable assumption when dealing with real data. The second approach, termed Ordered Temporal Sparse Subspace Clustering (OTSSC), assumes that only few data vectors will be corrupted. Furthermore, both approaches take into account the fact that neighbouring pixels in a frame will be similar to produce smoother segmentations, as well as consider the sequential nature of temporal data in the optimisation problem. (iii) the development of a criterion to estimate the number of underlying subspaces in the spectral clustering step, presented in Sec. 5.3.4. (iv) the derivation of a sequential approach for the UoS model in Sec. 5.4 that processes the data sequentially as it comes in, subtracting the estimated error from the previous frames in each step.
(v) the validation of the developed approaches on the glucose problem using real data experiments in Sec. 5.5.

The material presented in this chapter is partly taken from [DKZ16, WDZ16].

5.1 Introduction and Motivation

It has been established that high-dimensional data can often be shown to have a low-dimensional representation [EV09, Vid10, EV13, LLY10]. Methods such as Principal Component Analysis (PCA) [VMS05, CLMW11] assume that the data is

drawn from a single low-dimensional subspace. Recent research has shown, however, that in many cases, it is more realistic to assume that the data is drawn from a union of low-dimensional subspaces. The challenge is, then, to estimate the properties of the underlying subspaces, i.e., their number, their dimensions, their basis, and the labels describing subspace membership. Subspace segmentation methods have been studied in many works [Vid10, CLMW11] with low-rank methods [LLY10] and sparse subspace clustering [EV09, BK14] gaining special attention for applications in machine learning and motion tracking. Besides clustering the data into its different subspaces, subspace segmentation methods are able to separate the data from the outliers, as the outliers will lie in a separate subspace. In practical applications, however, the data is often contaminated by noise and the data vectors do not lie in their original subspaces, which is non-conforming to the UoS model. Several works have dealt with this problem [SC$^+$12, EV13, BK14] and propose methods to recover the data lying in the subspaces and separate it from the noise and outliers.

Chapter 4 presented mean-shift based approaches that are able to simultaneously segment the images and estimate the underlying relative remission value without requiring knowledge on the number of underlying clusters or their shape. The downside of mean-shift, however, is that it does not take into consideration the temporal nature of the video data, but segments each frame individually. The objective of this chapter is the segmentation of the glucose images obtained using the photometric measurement setup presented in Chapter 2.4, taking into account their temporal behaviour. Hereby, not only the intensity behaviour of the data vectors is considered a feature, but also the velocity with which the intensity changes over time. The incoming frames are segmented using the fact that the feature value of data vectors in the ROI temporally changes faster than that of the data vectors in other regions. This behaviour motivates the representation of the glucose video data as a union of low-dimensional subspaces.

5.2 State of the art

Let $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_L]$ in $\mathbb{R}^{N \times L}$ be a given set of L sample vectors that are drawn from an unknown union of k linear subspaces $\mathcal{S}_1 \cup \mathcal{S}_2 \cup \ldots \cup \mathcal{S}_k$ of dimensions $\{d_l\}_{l=1}^k$ in $\mathbb{R}^N$, with k and d_l being unknown. Furthermore, $\mathbf{X}$ is assumed to be self-representative [EV13], i.e.

$$\mathbf{X} = \mathbf{X}\mathbf{W}. \tag{5.1}$$

This means that each data vector in $\mathbf{X}$ can be efficiently reconstructed as a linear combination of the other vectors in the data set. Here $\mathbf{W}$ is an $L \times L$ coefficient

matrix with $\text{diag}(\mathbf{W}) = 0$, hence avoiding degenerate self-representation of individual vector components. A non-zero element w_{ij} of $\mathbf{W}$ indicates that $\mathbf{x}_i$ and $\mathbf{x}_j$, the i-th and respectively j-th column of $\mathbf{X}$, lie in the same subspace.

The last years have witnessed many works in the area of sparse modelling based on the UoS model. The aim is on the one hand to reduce the dimension of the large amounts of data being recorded, i.e., reduce the feature-space dimension, and on the other hand to cluster this data into meaningful segments, i.e., reduce the object-space dimension. A further result of this is finding outliers and errors in the data as they will not belong to the same subspaces as the data points.

The two most successful methods are low-rank representation (LRR) [LLY+13] and sparse subspace clustering (SSC) [EV09, EV10]. What unites these methods is the basic idea consisting of two steps:

1. Find a sparse representation for each data point by formulating an adequate optimization problem.
2. Employ this sparse representation to build an affinity matrix and use it to perform spectral clustering of the data.

The developed methods differ essentially in the first step.

SSC searches for the sparsest representation of each column in $\mathbf{X}$ as a linear combination of all other columns. The proposed optimisation problem reads [EV09, EV10]

$$\begin{aligned} \underset{\mathbf{W},\mathbf{E},\mathbf{Z}}{\text{minimize}} \quad & \|\mathbf{W}\|_1 + \zeta_1\|\mathbf{E}\|_1 + \zeta_2\|\mathbf{Z}\|_F^2 \\ \text{subject to} \quad & \mathbf{X} = \mathbf{X}\mathbf{W} + \mathbf{E} + \mathbf{Z},\ \text{diag}(\mathbf{W}) = 0, \end{aligned} \tag{5.2}$$

where $\mathbf{E}$ corresponds to sparse outlying entries, $\mathbf{Z}$ is a noise term, and ζ_1, ζ_2 are the regularisation parameters. Furthermore $\|\cdot\|_F^2$ denotes the Frobenius norm and $\|\cdot\|_1$ is the l_1-norm that promotes sparsity. Here, both the $\mathbf{E}$ and the $\mathbf{Z}$ terms can be left out if no outliers exist or no noise is present in the data.

This formulation is slightly adjusted for clustering applications in [ESV12]

$$\begin{aligned} \underset{\mathbf{W}}{\text{minimize}} \quad & \|\mathbf{W}\|_{1,2} + \zeta\|\mathbf{X} - \mathbf{X}\mathbf{W}\|_F^2 \\ \text{subject to} \quad & \mathbf{1}^T\mathbf{W} = \mathbf{1}^T. \end{aligned} \tag{5.3}$$

By applying the l_1/l_2-norm $\|\cdot\|_{1,2}$ on $\mathbf{W}$, the sparsity of $\mathbf{W}$ is forced to have common support and, hereby, few representatives are chosen to express the whole data set.

In [BK14, BK15], the authors consider a different noise model by assuming that all data samples are corrupted by sparse errors, which is more realistic for practical applications. This model, however, renders the UoS model damaged as no data vector is close to its original subspace. The authors propose a bi-sparsity model to deal with this assumption

$$\begin{aligned} \underset{\mathbf{W},\mathbf{E}}{\text{minimize}} \quad & \|\mathbf{W}\|_1 + \zeta\|\mathbf{E}\|_1, \\ & \text{subject to} \quad \mathbf{X} = \mathbf{X}_0 + \mathbf{E}, \\ & \mathbf{X}_0 = \mathbf{X}_0\mathbf{W},\ \operatorname{diag}(\mathbf{W}) = 0, \end{aligned} \tag{5.4}$$

where $\mathbf{X}_0$ represents the uncontaminated data vectors and $\mathbf{E}$ contains the sparse errors.

There have also been a few works that deal with sequential data [GGL13, TGG14, TGG15, WHG+15]. In [GGL13] the authors propose a spatial subspace clustering (SpatSC) method for the segmentation of hyperspectral drill hole data. Due to geological properties in the data they are able to assume that the difference between two neighbouring data vectors will not be large: $\mathbf{x}_i \approx \mathbf{x}_{i+1}$. They introduce a penalty term into the optimisation problem which enforces smoothness of the segmentation, leading to

$$\begin{aligned} \underset{\mathbf{W},\mathbf{E}}{\text{minimize}} \quad & \frac{1}{2}\|\mathbf{E}\|_F^2 + \zeta_1\|\mathbf{W}\|_1 + \zeta_2\|\mathbf{W}\mathbf{R}\|_1 \\ & \text{subject to} \quad \mathbf{X} = \mathbf{X}\mathbf{W} + \mathbf{E},\ \operatorname{diag}(\mathbf{W}) = 0, \end{aligned} \tag{5.5}$$

where $\mathbf{R}$ is given as in Eq. (5.11) and the term $\|\mathbf{W}\mathbf{R}\|_1$ forces consecutive columns of $\mathbf{X}$ to be similar.

Tierney *et al.* extend this in [TGG14, TGG15] by deriving an ordered subspace clustering (OSC) algorithm, which strictly enforces column similarity. They achieve this by replacing $\|\mathbf{W}\mathbf{R}\|_1$ by $\|\mathbf{W}\mathbf{R}\|_{1,2}$. The resulting optimisation problem reads

$$\begin{aligned} \underset{\mathbf{W},\mathbf{E}}{\text{minimize}} \quad & \frac{1}{2}\|\mathbf{E}\|_F^2 + \zeta_1\|\mathbf{W}\|_1 + \zeta_2\|\mathbf{W}\mathbf{R}\|_{1,2} \\ & \text{subject to} \quad \mathbf{X} = \mathbf{X}\mathbf{W} + \mathbf{E},\ \operatorname{diag}(\mathbf{W}) = 0. \end{aligned} \tag{5.6}$$

The method of low-rank representation (LRR) [LLY10, LLY+13] adopts a different strategy by attempting to find a joint representation of the data with the help

of the lowest rank representation. In order to make the problem computationally feasible, the lowest rank optimisation problem is replaced by the minimisation of the nuclear norm. The resulting optimisation problem reads

$$\begin{aligned} \underset{\mathbf{W},\mathbf{E}}{\text{minimize}} \ & \|\mathbf{W}\|_* + \zeta\|\mathbf{E}\|_{1,2}, \\ \text{subject to } & \mathbf{X} = \mathbf{X}\mathbf{W} + \mathbf{E}, \end{aligned} \tag{5.7}$$

where $\|\cdot\|_*$ is the nuclear norm and $\mathbf{E}$ is an error term representing data corruptions. If the error is assumed to be Gaussian, the $l_{1,2}$-norm can be replaced by the Frobenius norm $\|\cdot\|_F^2$.

As for the SSC formulation, a sequential low-rank subspace clustering method has been developed that takes into account the sequential nature of the data [GGL+15]. The corresponding optimisation problem is given by

$$\begin{aligned} \underset{\mathbf{W},\mathbf{E}}{\text{minimize}} \quad & \frac{1}{2}\|\mathbf{E}\|_F^2 + \zeta_1\|\mathbf{W}\|_* + \zeta_2\|\mathbf{W}\mathbf{R}\|_1 \\ \text{subject to} \quad & \mathbf{X} = \mathbf{X}\mathbf{W} + \mathbf{E}, \ \operatorname{diag}(\mathbf{W}) = 0. \end{aligned} \tag{5.8}$$

After obtaining the sparse representation matrix $\mathbf{W}$ using the presented methods, clustering can be performed. Typically this is done using spectral clustering techniques such as Normalised Cuts (NCuts) [SM00]. Basically, the idea of spectral clustering is to interpret the sparse representation matrix as a graph of L nodes, where each element w_{ij} describes the similarity between node i and node j. Spectral clustering techniques define so-called graph Laplacians $\mathbf{L}_C$ [Chu97] with the help of $\mathbf{W}$ and use them to perform the clustering. More details on how the Laplacian is computed are given in the sequel.

Spectral clustering requires knowledge of the number of underlying clusters, which is not always given in practice. There are many methods in the literature to estimate this parameter such as information-theoretic criteria [SS04]. A nice property of graph Laplacians, however, is that they allow estimating the unknown number of underlying clusters $\hat{k}$ in a very elegant manner. In an ideal case, where $\mathbf{W}$ is a strictly block diagonal matrix and, therefore, $\mathbf{L}_C$ is also strictly block diagonal, the number of underlying clusters is obtained by calculating the number of eigenvalues of $\mathbf{L}_C$ that are equal to zero [Moh97]. For real data, $\mathbf{L}_C$ is rarely strictly block diagonal and its first $\hat{k}$ eigenvalues are small but not equal to zero. To this end, eigengap methods are used to determine $\hat{k}$, which analyse the distance between subsequent eigenvalues.

5.3 Sparse Subspace Clustering for Temporal Video Segmentation

The contribution of this section is the proposal of two novel formulations for subspace clustering intended for segmentation of temporal video data [DKZ16]. As the data is quite noisy, the first approach adopts the signal model proposed by Bian *et al.* [BK14], as given in Eq. (5.4). That is, the assumption is made that all data vectors can be degraded by noise and not lie directly in their corresponding subspace. The underlying noise is assumed to be Gaussian and, therefore, the Frobenius norm is applied to the error matrix. The second approach assumes only few data vectors to be contaminated.

Inspired by [TGG14, TGG15], smoothness of segmentation is incorporated into the problem by including a term that forces neighbouring columns to behave similarly. This is well-grounded by the fact that neighbouring columns represent neighbouring pixels of an image, which typically do not change abruptly. Furthermore, to make the formulation more suitable for segmentation a mixed norm is applied on the coefficient matrix as in [ESV12]. This encourages the choice of only few representatives to express each data vector.

The proposed optimisation problems are solved with a linearized alternating direction method with parallel splitting and adaptive penalty (LADMPSAP). This was recently proposed by Lin *et al.* [LLL13] to solve multi-block convex optimisation problems. Provided that all the component objective problems have bounded subgradients, LADMPSAP is able to guarantee global convergence of the dual problem and achieves this under reasonable complexity. The proposed approaches are applied to the glucose measurement problem. To the best of knowledge, this is the first time that subspace clustering methods have been applied to segmentation of medical video data.

This section starts with presenting the data structure. Then, the Ordered Robust Subspace Recovery (ORoSuRe) algorithm is derived, assuming that all data points may be corrupted. Relaxing this assumption leads to a slightly different data model, where only a few data points are assumed to be corrupted. Using the relaxed model, the Ordered Temporal Sparse Subspace Clustering (OTSSC) is derived. Subsequently, the attention is turned to the spectral clustering step, where the contribution lies in proposing a slightly different criterion for the estimation of the number of underlying subspaces that is better suited to the glucose application.

5.3.1 Data Model and Assumptions

Each incoming frame obtained by the camera is denoted by $\mathbf{I}^{(n)} \in \mathbb{R}^{M_x \times M_y}, n = 1, ..., N$, where M_x, M_y are the number of rows and columns in the image and N is the total number of frames. By vectorising the frames and stacking the vectors into a matrix, the data matrix $\mathbf{X} = [\text{vec}(\mathbf{I}^{(1)}), ..., \text{vec}(\mathbf{I}^{(N)})]^T \in \mathbb{R}^{N \times L}$ is obtained, where $L = M_x \cdot M_y$. Figure 5.1 illustrates an example of a data matrix constructed using a glucose measurement, i.e., stacking vectorised versions of the glucose images over frames. The columns of the matrix represent the individual pixels in an image and the rows represent the time frames.

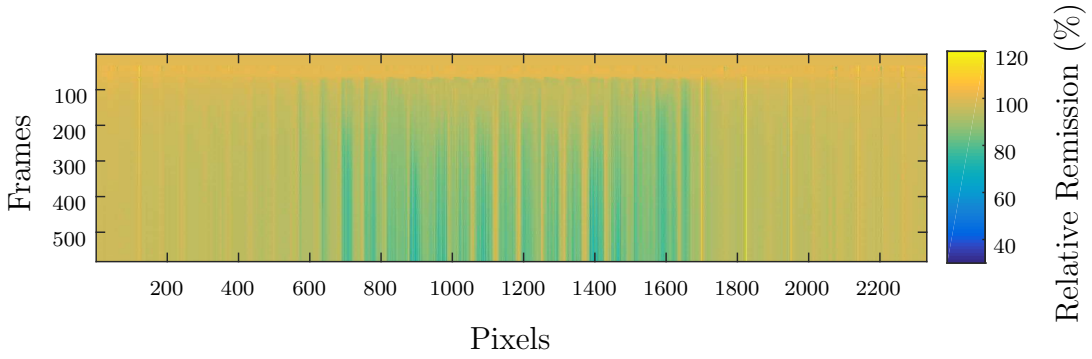


Figure 5.1: An example of the data matrix $\mathbf{X} \in \mathbb{R}^{N \times L}$.

Neighbouring columns in $\mathbf{X}$ represent the progression of neighbouring pixels in the images over frames and, therefore, they can be assumed to be similar with high probability, i.e., $\mathbf{x}_i \approx \mathbf{x}_{i+1}$, $\mathbf{x}_i$ being the i-th column of $\mathbf{X}$. Note that due to the vectorisation of the frames, only the smoothness of neighbouring pixels in the vertical direction is being considered. This is intended in the glucose application as the images are vertical in structure. Extending the model to include full pixel neighbourhoods is also possible. The term $\|\mathbf{WR}\|_{1,2}$ is included into the optimisation problem to enforce neighbourhood similarity, where $\mathbf{R}$ is as in Eq. (5.11). It compels the difference between neighbouring columns $[\mathbf{w}_2 - \mathbf{w}_1, \mathbf{w}_3 - \mathbf{w}_2, \ldots, \mathbf{w}_L - \mathbf{w}_{L-1}]$ to be zero, and hereby forces a smoothness of the coefficient matrix $\mathbf{W}$.

In order to make the problem suitable for segmentation, the coefficient matrix $\mathbf{W}$ should not only be sparse at the element-level, but also favour the selection of a small number of representatives to express the data, i.e., whole rows should go to

zero. The indices of the non-zero rows in $\mathbf{W}$ indicate columns of $\mathbf{X}$ that are chosen as representatives. To achieve this, the group sparse l_1/l_2-norm is chosen for the coefficient matrix. It forces the rows of $\mathbf{W}$ to be zero. Thereby, the choice of only a few representatives for each data vector is favoured, making it more suitable for segmentation applications.

5.3.2 Ordered Robust Subspace Recovery (ORoSuRe): All data vectors are contaminated

As previously asserted, the glucose frames can be very noisy and it is essential to assume that all data points may be corrupted. To this end, the model proposed in Eq. (5.4) [BK14] is adopted, which assumes that all original data points $\mathbf{X}_0$ may be corrupted by noise. This results in the following system model for the observed data matrix:

$$\mathbf{X} = \mathbf{X}_0 + \mathbf{E}. \tag{5.9}$$

A Gaussian noise model is assumed for the data. This assumption is based on calibration frames that are obtained prior to the onset of the chemical reaction.

These considerations yield the following formulation of the optimisation problem

$$\begin{aligned} \underset{\mathbf{W},\mathbf{E}}{\text{minimize}} \quad & \|\mathbf{W}\|_{1,2} + \frac{\zeta_1}{2}\|\mathbf{E}\|_F^2 + \zeta_2\|\mathbf{W}\mathbf{R}\|_{1,2} \\ & \text{subject to } \mathbf{X} = \mathbf{X}_0 + \mathbf{E}, \\ & \mathbf{X}_0 = \mathbf{X}_0\mathbf{W}, \end{aligned} \tag{5.10}$$

where

$$\mathbf{R} = \begin{bmatrix} -1 & & & & \\ 1 & -1 & & & \\ & 1 & -1 & & \\ & & \ddots & \ddots & \\ & & & 1 & -1 \end{bmatrix} \in \mathbb{R}^{L \times L-1}. \tag{5.11}$$

ζ_1, ζ_2 are the regularisation parameters, the factor $\frac{1}{2}$ is included to ease computation. The Frobenius norm is given by $\|\mathbf{A}\|_F = \sqrt{\sum_i \sum_j |a_{ij}|^2}$, and the l_1/l_2-norm is defined as $\|\mathbf{A}\|_{1,2} \triangleq \sum_i \|\mathbf{a}_j\|$, $\mathbf{a}_j$ being the j-th column of $\mathbf{A}$. The self-representative property, $\mathbf{X}_0 = \mathbf{X}_0\mathbf{W}$, is incorporated into the optimisation problem as a constraint. The penalty $\|\mathbf{W}\mathbf{R}\|_{1,2}$ enforces strong similarity of neighbouring columns, i.e., neighbouring pixels in an image, hereby leading to a smoother image segmentation. Finally, the l_1/l_2-norm over $\mathbf{W}$ counts the number of non-zero rows in $\mathbf{W}$, which indicate the

chosen representative columns of $\mathbf{X}_0$.

To efficiently solve this problem auxiliary variables are introduced, converting it to the following equivalent three-variable problem:

$$\begin{aligned} \underset{\mathbf{W},\mathbf{E},\mathbf{J}}{\text{minimize}} \quad & \|\mathbf{W}\|_{1,2} + \frac{\zeta_1}{2}\|\mathbf{E}\|_F^2 + \zeta_2\|\mathbf{J}\|_{1,2} \\ \text{subject to } & \mathbf{X} = \mathbf{X}_0\mathbf{W} + \mathbf{E}, \\ & \mathbf{J} = \mathbf{W}\mathbf{R}. \end{aligned} \tag{5.12}$$

The augmented Lagrange function can be written as

$$\begin{aligned} \mathcal{L}(\mathbf{E},\mathbf{W},\mathbf{J},\mathbf{Y}_1,\mathbf{Y}_2,\mu) &= \|\mathbf{W}\|_{1,2} + \frac{\zeta_1}{2}\|\mathbf{E}\|_F^2 + \zeta_2\|\mathbf{J}\|_{1,2} \\ &+ \langle \mathbf{Y}_1, (\mathbf{X}-\mathbf{E})\mathbf{W} - (\mathbf{X}-\mathbf{E})\rangle \\ &+ \frac{\mu}{2}\left\|(\mathbf{X}-\mathbf{E})\mathbf{W} - (\mathbf{X}-\mathbf{E})\right\|_F^2 \\ &+ \langle \mathbf{Y}_2, \mathbf{J} - \mathbf{W}\mathbf{R}\rangle + \frac{\mu}{2}\left\|\mathbf{J} - \mathbf{W}\mathbf{R}\right\|_F^2, \end{aligned} \tag{5.13}$$

where $\mathbf{Y}_1, \mathbf{Y}_2$ are the Lagrange multipliers and $\mu > 0$ is a penalty parameter.

Lin *et al.* [LLL13] recently proposed a linearized alternating direction method with parallel splitting and adaptive penalty (LADMPSAP) that extends the linearized alternating direction method (LADM) [LLS11] to solve multi-block separable convex problems. The LADMSAP provides global convergence guarantees of the dual problem [LLL13], and is computationally very efficient. It is used to solve Eq. (5.13) by alternating between $\mathbf{W}, \mathbf{E}$ and $\mathbf{J}$.

1. Updating $\mathbf{W}^{(j+1)}$:
   First, $\mathbf{X}_0$ is updated as $\mathbf{X}_0^{(j)} = \mathbf{X} - \mathbf{E}^{(j)}$, then $\mathbf{W}^{(j+1)}$ can be obtained as

$$\mathbf{W}^{(j+1)} = \arg\min_{\mathbf{W}} \|\mathbf{W}\|_{1,2} + F(\mathbf{W}), \tag{5.14}$$

   where

$$\begin{aligned} F(\mathbf{W}) &= \langle \mathbf{Y}_1^{(j)}, \mathbf{X}_0^{(j)}\mathbf{W} - \mathbf{X}_0^{(j)}\rangle + \frac{\mu^{(j)}}{2}\left\|\mathbf{X}_0\mathbf{W} - \mathbf{X}_0\right\|_F^2 \\ &+ \langle \mathbf{Y}_2^{(j)}, \mathbf{J}^{(j)} - \mathbf{W}\mathbf{R}\rangle + \frac{\mu^{(j)}}{2}\left\|\mathbf{J}^{(j)} - \mathbf{W}\mathbf{R}\right\|_F^2 \\ &= \frac{\mu^{(j)}}{2}\left\|\mathbf{X}_0^{(j)}\mathbf{W} - \mathbf{X}_0^{(j)} + \frac{Y_1^{(j)}}{\mu^{(j)}}\right\|_F^2 + \frac{\mu^{(j)}}{2}\left\|\mathbf{J}^{(j)} - \mathbf{W}\mathbf{R} + \frac{Y_2^{(j)}}{\mu^{(j)}}\right\|_F^2. \end{aligned} \tag{5.15}$$

The last equality is achieved by ignoring constant terms.
Linearising $F(\mathbf{W})$ Eq. (5.14) is solved by

$$\begin{aligned}\mathbf{W}^{(j+1)} \approx \arg\min_{\mathbf{W}} \|\mathbf{W}\|_{1,2} \\ + \frac{\mu^{(j)}}{2}\left\|\mathbf{W} - \left(\mathbf{W}^{(j)} - \frac{1}{\mu^{(j)}}\nabla F(\mathbf{W}^{(j)})\right)\right\|_F^2,\end{aligned} \tag{5.16}$$

where

$$\begin{aligned}\nabla F(\mathbf{W}^{(j)}) = \mu^{(j)}\mathbf{X}_0^{(j)T}\left(\mathbf{X}_0^{(j)}\mathbf{W}^{(j)} - \mathbf{X}_0^{(j)} + \frac{1}{\mu^{(j)}}\mathbf{Y}_1^{(j)}\right) \\ - \mu^{(j)}\left(\mathbf{J}^{(j)} - \mathbf{W}^{(j)}\mathbf{R} + \frac{1}{\mu^{(j)}}\mathbf{Y}_2^{(j)}\right)\mathbf{R}^T.\end{aligned} \tag{5.17}$$

If $\mathbf{V}^{(j)} = \left(\mathbf{W}^{(j)} - \frac{1}{\mu^{(j)}}\nabla F(\mathbf{W}^{(j)})\right)$, then the closed-form solution to Eq. (5.14) is [LLY10]

$$\mathbf{W}_l^{(j+1)} = \begin{cases}\frac{\|\mathbf{V}_l^{(j)}\| - \frac{1}{\mu^{(j)}}}{\|\mathbf{V}_l^{(j)}\|} \cdot \mathbf{V}_l^{(j)}, & \text{if } \|\mathbf{V}_l^{(j)}\| > \frac{1}{\mu^{(j)}} \\ 0, & \text{otherwise.}\end{cases} \tag{5.18}$$

2. Updating $\mathbf{E}^{(j+1)}$:

$$\begin{aligned}\mathbf{E}^{(j+1)} =& \arg\min_{\mathbf{E}} \frac{\zeta_1}{2}\|\mathbf{E}\|_F^2 \\ &+ \frac{\mu^{(j)}}{2}\left\|(\mathbf{X} - \mathbf{E})\mathbf{W}^{(j)} - (\mathbf{X} - \mathbf{E}) + \frac{\mathbf{Y}_1^{(j)}}{\mu^{(j)}}\right\|_F^2 \\ =& \frac{\mu^{(j)}}{\zeta_1}\left(\mathbf{X}_0^{(j)}\tilde{\mathbf{W}}^{(j)} - \frac{\mathbf{Y}_1^{(j)}}{\mu^{(j)}}\right)\tilde{\mathbf{W}}^{(j)T},\end{aligned} \tag{5.19}$$

where $\tilde{\mathbf{W}} = \mathbf{I} - \mathbf{W}$, $\mathbf{I}$ being the identity matrix.

3. Updating $\mathbf{J}^{(j+1)}$:

$$\begin{aligned}\mathbf{J}^{(j+1)} = \arg\min_{\mathbf{J}} \zeta_2\|\mathbf{J}\|_{1,2} \\ + \frac{\mu^{(j)}}{2}\left\|\mathbf{J} - \mathbf{W}^{(j)}\mathbf{R} + \frac{\mathbf{Y}_2^{(j)}}{\mu^{(j)}}\right\|_F^2.\end{aligned} \tag{5.20}$$

Let $\mathbf{U}^{(j)} = \mathbf{W}^{(j)}\mathbf{R} - \frac{\mathbf{Y}_2^{(j)}}{\mu^{(j)}}$, then the closed-form solution for $\mathbf{J}^{(j+1)}$ is given by

$$\mathbf{J}_l^{(j+1)} = \begin{cases}\frac{\|\mathbf{U}_l^{(j)}\| - \frac{\zeta_2}{\mu^{(j)}}}{\|\mathbf{U}_l^{(j)}\|} \cdot \mathbf{U}_l^{(j)}, & \text{if } \|\mathbf{U}_l^{(j)}\| > \frac{\zeta_2}{\mu^{(j)}} \\ 0, & \text{otherwise.}\end{cases} \tag{5.21}$$

4. The Lagrange multipliers $\mathbf{Y}_1^{(j+1)}, \mathbf{Y}_2^{(j+1)}$ are updated as

$$\begin{aligned}\mathbf{Y}_1^{(j+1)} &= \mathbf{Y}_1^{(j)} + \mu^{(j+1)}(\mathbf{X}_0^{(j+1)}\mathbf{W}^{(j+1)} - \mathbf{X}_0^{(j+1)}),\\ \mathbf{Y}_2^{(j+1)} &= \mathbf{Y}_2^{(j)} + \mu^{(j+1)}(\mathbf{J}^{(j+1)} - \mathbf{W}^{(j+1)}\mathbf{R}),\end{aligned} \tag{5.22}$$

   where $\mathbf{X}_0^{(j+1)} = \mathbf{X} - \mathbf{E}^{(j+1)}$.

5. The penalty parameter $\mu^{(j+1)}$ is updated according to

$$\mu^{(j+1)} = \rho\mu^{(j)}, \tag{5.23}$$

   $\rho > \|\mathbf{X}_0\|^2 + \|\mathbf{R}\|^2$ being a constant that ensures a steady increase of $\mu^{(j)}$.

Note that the variable updates are performed in parallel and always depend on their former values from the last iteration, allowing for efficient parallel computing.

Convergence of the algorithm is achieved when the following two conditions are met:

$$\|\mathbf{X}_0\mathbf{W}^{(j+1)} - \mathbf{X}_0\|_F^2 < \epsilon_1; \quad \|\mathbf{J}^{(j+1)} - \mathbf{W}^{(j+1)}\mathbf{R}\|_F^2 < \epsilon_1 \tag{5.24}$$

$$\mu^{(j)} \cdot \max\{\|\mathbf{W}^{(j+1)} - \mathbf{W}^{(j)}\|_F^2, \|\mathbf{E}^{(j+1)} - \mathbf{E}^{(j)}\|_F^2, \|\mathbf{J}^{(j+1)} - \mathbf{J}^{(j)}\|_F^2\} < \epsilon_2. \tag{5.25}$$

Although, the global convergence guarantee of the algorithm is given in [LLL13], it is presented here tailored to this specific problem.

Theorem 2. *If the sequence* $\{\mu^{(j)}\}$ *is non-decreasing and* $\sum_{j=1}^{+\infty}(\mu^{(j)})^{-1} = +\infty$, $\rho > \|\mathbf{X}_0\|^2 + \|\mathbf{R}\|^2$, $\partial\|\mathbf{W}\|_{1,2}, \partial\|\mathbf{E}\|_F^2, \partial\|\mathbf{J}\|_{1,2}$ *are bounded, then the sequence* $\{(\mathbf{W}^{(j)}, \mathbf{E}^{(j)}, \mathbf{J}^{(j)})\}$ *generated by LADMPSAP converges to an optimal solution of Eq. (5.12).*

A more detailed treatment of the convergence of this algorithm can be found in [LLL13].

5.3.3 The Ordered Temporal Sparse Subspace Clustering (OTSSC): Few data vectors are contaminated

In the previous section a data model is assumed, in which all data vectors may be corrupted by noise. Here an optimisation problem is derived that relaxes this assumption, by assuming that only few data vectors will be corrupted. This results in the following data model

$$\mathbf{X} = \mathbf{X}\mathbf{W} + \mathbf{E}. \tag{5.26}$$

This approach is termed Ordered Temporal Sparse Subspace Clustering (OTSSC) and is given by the following optimisation problem:

$$\begin{aligned} \underset{\mathbf{W}}{\text{minimize}} \quad & \frac{1}{2}\|\mathbf{E}\|_F^2 + \zeta_1\|\mathbf{W}\|_{1,2} + \zeta_2\|\mathbf{J}\|_{1,2} \\ & \text{subject to } \mathbf{X} = \mathbf{X}\mathbf{W} + \mathbf{E}. \end{aligned} \tag{5.27}$$

As in Eq. (5.12), the group sparse norm on $\mathbf{W}$ enforces the choice of only few representatives, which is reasonable for segmentation. The term $\|\mathbf{W}\mathbf{R}\|_{1,2}$ ensures a smooth segmentation of the image pixels, as previously clarified.

A three-variable problem is attained by substituting $\mathbf{J} = \mathbf{W}\mathbf{R}$, resulting in the augmented Lagrange function as

$$\begin{aligned} \mathcal{L}(\mathbf{E}, \mathbf{W}, \mathbf{J}, \mathbf{Y}_1, \mathbf{Y}_2, \mu) = & \frac{1}{2}\|\mathbf{E}\|_F^2 + \zeta_1\|\mathbf{W}\|_{1,2} + \zeta_2\|\mathbf{J}\|_{1,2} \\ & + \langle \mathbf{Y}_1, \mathbf{X}\mathbf{W} - \mathbf{X} + \mathbf{E}\rangle + \frac{\mu}{2}\|\mathbf{X}\mathbf{W} - \mathbf{X} + \mathbf{E}\|_F^2 \\ & + \langle \mathbf{Y}_2, \mathbf{J} - \mathbf{W}\mathbf{R}\rangle + \frac{\mu}{2}\|\mathbf{J} - \mathbf{W}\mathbf{R}\|_F^2. \end{aligned} \tag{5.28}$$

LADMPSAP is used to solve the multi-block separable convex problem in Eq. (5.28) by alternating between $\mathbf{E}, \mathbf{W}$ and $\mathbf{J}$, similarly as in Sec. 5.3.2.

1. Updating $\mathbf{W}^{(j+1)}$:

$$\mathbf{W}^{(j+1)} = \arg\min_{\mathbf{W}} \zeta_1\|\mathbf{W}\|_{1,2} + F(\mathbf{W}) \tag{5.29}$$

and $F(\mathbf{W})$ is given by

$$\begin{aligned} F(\mathbf{W}) = & \frac{\mu^{(j)}}{2}\left\|\mathbf{X}\mathbf{W} - \mathbf{X} + \mathbf{E}^{(j)} + \frac{\mathbf{Y}_1^{(j)}}{\mu^{(j)}}\right\|_F^2 \\ & + \frac{\mu^{(j)}}{2}\left\|\mathbf{J}^{(j)} - \mathbf{W}\mathbf{R} + \frac{\mathbf{Y}_2^{(j)}}{\mu^{(j)}}\right\|_F^2, \end{aligned} \tag{5.30}$$

where

$$\nabla F(\mathbf{W}^{(j)}) = \mu^{(j)}\mathbf{X}^T(\mathbf{X}\mathbf{W}^{(j)} - \mathbf{X} + \mathbf{E}^{(j)} + \frac{\mathbf{Y}_1^{(j)}}{\mu^{(j)}}) - \mu^{(j)}(\mathbf{J}^{(j)} - \mathbf{W}^{(j)}\mathbf{R} + \frac{\mathbf{Y}_2^{(j)}}{\mu^{(j)}})\mathbf{R}^T. \tag{5.31}$$

By linearising $F(\mathbf{W})$, Eq. (5.29) can be solved as

$$\begin{aligned} \mathbf{W}^{(j+1)} \approx & \arg\min_{\mathbf{W}} \zeta_1\|\mathbf{W}\|_{1,2} \\ & + \frac{\mu^{(j)}}{2}\|\mathbf{W} - \left(\mathbf{W}^{(j)} - \frac{1}{\mu^{(j)}}\nabla F(\mathbf{W}^{(j)})\right)\|_F^2. \end{aligned} \tag{5.32}$$

If $\mathbf{V}^{(j)} = \left(\mathbf{W}^{(j)} - \frac{1}{\mu^{(j)}}\nabla F(\mathbf{W}^{(j)})\right)$, then according to [LLY10]

$$\mathbf{W}_l^{(j+1)} = \begin{cases} \frac{\|\mathbf{V}_l^{(j)}\| - \frac{\zeta_1}{\mu^{(j)}}}{\|\mathbf{V}_l^{(j)}\|} \cdot \mathbf{V}_l^{(j)}, & \text{if } \|\mathbf{V}_l^{(j)}\| > \frac{\zeta_1}{\mu^{(j)}} \\ 0, & \text{otherwise.} \end{cases} \tag{5.33}$$

2. Updating $\mathbf{E}^{(j+1)}$:

$$\mathbf{E}^{(j+1)} = \arg\min_{\mathbf{E}} \frac{1}{2}\|\mathbf{E}\|_F^2 + \frac{\mu}{2}\left\|\mathbf{X}\mathbf{W} - \mathbf{X} + \mathbf{E} + \frac{\mathbf{Y}_1}{\mu}\right\|_F^2. \tag{5.34}$$

This has a very straightforward closed-form solution

$$\mathbf{E}^{(j+1)} = \frac{\mathbf{X} - \mathbf{X}\mathbf{W}^{(j)} - \frac{1}{\mu^{(j)}}\mathbf{Y}_1^{(j)}}{1 + \frac{\zeta_1}{\mu^{(j)}}}. \tag{5.35}$$

3. Updating $\mathbf{J}^{(j+1)}$:

$$\mathbf{J}^{(j+1)} = \arg\min_{\mathbf{J}} \zeta_2\|\mathbf{J}\|_{1,2} + \frac{\mu}{2}\left\|\mathbf{J} - \mathbf{W}\mathbf{R} + \frac{\mathbf{Y}_2}{\mu}\right\|_F^2. \tag{5.36}$$

For $\mathbf{U}^{(j)} = \mathbf{W}^{(j)}\mathbf{R} - \frac{\mathbf{Y}_2^{(j)}}{\mu^{(j)}}$, the solution for $\mathbf{J}^{(j+1)}$ is obtained by

$$\mathbf{J}_l^{(j+1)} = \begin{cases} \frac{\|\mathbf{U}_l^{(j)}\| - \frac{\zeta_2}{\mu^{(j)}}}{\|\mathbf{U}_l^{(j)}\|} \cdot \mathbf{U}_l^{(j)}, & \text{if } \|\mathbf{U}_l^{(j)}\| > \frac{\zeta_2}{\mu^{(j)}} \\ 0, & \text{otherwise.} \end{cases} \tag{5.37}$$

4. The Lagrange multipliers $\mathbf{Y}_1, \mathbf{Y}_2$ are updated as

$$\begin{aligned} \mathbf{Y}_1^{(j+1)} &= \mathbf{Y}_1^{(j)} + \mu^{(j)} \cdot (\mathbf{X}\mathbf{W}^{(j+1)} - \mathbf{X} + \mathbf{E}^{(j+1)}) \\ \mathbf{Y}_2^{(j+1)} &= \mathbf{Y}_2^{(j)} + \mu^{(j)} \cdot (\mathbf{J}^{(j+1)} - \mathbf{W}^{(j+1)}\mathbf{R}). \end{aligned} \tag{5.38}$$

5. The penalty parameter μ is updated as

$$\mu^{(j+1)} = \rho\mu^{(j)}, \tag{5.39}$$

$\rho > \|\mathbf{X}\|^2 + \|\mathbf{R}\|^2$ being a constant that ensures a steady increase of $\mu^{(j)}$.

Convergence of the algorithm is achieved when Eq. (5.24) and Eq. (5.25) are fulfilled.

The global convergence guarantee is given by following theorem:

Theorem 3. *If the sequence* $\{\mu^{(j)}\}$ *is non-decreasing and* $\sum_{j=1}^{+\infty}(\mu^{(j)})^{-1} = +\infty$, $\rho > \|\mathbf{X}\|^2 + \|\mathbf{R}\|^2$, $\partial\|\mathbf{W}\|_{1,2}, \partial\|\mathbf{E}\|_F^2, \partial\|\mathbf{J}\|_{1,2}$ *are bounded, then the sequence* $\{(\mathbf{W}^{(j)}, \mathbf{E}^{(j)}, \mathbf{J}^{(j)})\}$ *generated by LADMPSAP converges to an optimal solution of Eq. (5.27).*

Note that this is very similar to Theorem 2, however, $\mathbf{X}_0$ has been substituted by $\mathbf{X}$.

5.3.4 Spectral Clustering

Once the coefficient matrix $\mathbf{W}$ has been determined, it is used to perform spectral clustering [VL07, NJW+02] and segment the data. For this purpose an affinity matrix is formed as

$$\mathbf{C} = |\mathbf{W}| + |\mathbf{W}|^T, \tag{5.40}$$

representing the data as an undirected graph. The element c_{ij} indicates the connection between elements i and j of the data matrix $\mathbf{X}$. The normalised Laplacian of the affinity matrix is computed as

$$\mathbf{L_C} = \mathbf{D} - \mathbf{C}, \tag{5.41}$$

with $\mathbf{D}$ being the diagonal degree matrix $\mathbf{D} = \text{diag}\left(\sum_{j=1}^{L} c_{1j}, \ldots, \sum_{j=1}^{L} c_{Lj}\right)$, which contains the summation over all rows of the adjacency matrix $\mathbf{C}$.

The first k eigenvectors of $\mathbf{L_C}$ are calculated as $\mathbf{u}_1, \ldots, \mathbf{u}_k$ and form the matrix $\mathbf{U} \in \mathbb{R}^{L \times k}$. The k-means algorithm [DHS12] is, then, applied to the matrix comprised of the normalised rows of $\mathbf{U}$. The k clusters obtained using the k-means algorithm correspond to the segmented regions. This technique, termed Normalised Cuts method (NCuts) [SM00], has proven to be very robust for image segmentation applications using spectral clustering and remains the current state-of-the-art approach. [EV13, LLY+13].

NCuts requires the number of subspaces k to be either given or estimated from the data. In the glucose application no knowledge of k is available, as presented in more detail in Chapter 4. Hence, the following criterion is proposed to estimate it. The sorted eigenvalues of the Laplacian $\mathbf{L_C}$ are denoted by $\lambda_l, l = 1, ..., L$ and the eigenvalue gap $\Delta_{\lambda_l} = \lambda_l - \lambda_{l+1}$ is computed. The number of subspaces can be estimated as

$$\hat{k} = \sum_{l=1}^{L-1} \mathbf{1}\{\Delta_{\lambda_l} > \delta_{\text{eigen}}\}, \tag{5.42}$$

with $\mathbf{1}$ being the indicator function and δ_{eigen} a predefined threshold. Hence, the estimated number of subspaces $\hat{k}$ is calculated by counting the number of eigenvalues that have a large eigenvalue gap. This is a slight modification of the criterion in [VL07].

The whole procedure for ORoSuRe is summarised in the sequel. By replacing equations Eq. (5.18) to Eq. (5.23) by Eq. (5.33) to Eq. (5.39) the procedure for OTSSC is obtained. The inputs to the algorithm are the data matrix $\mathbf{X} \in \mathbb{R}^{F \times L}$, the regularisation parameters ζ_1, ζ_2 as well as $\rho, \epsilon_1, \epsilon_2$ and the threshold for determining the underlying number of subspaces δ_{eigen}.

- **Step 1.** Initialise $\mathbf{W}^{(0)}, \mathbf{E}^{(0)}, \mathbf{J}^{(0)}, \mathbf{Y}_1^{(0)}, \mathbf{Y}_2^{(0)}, \mu^{(0)}$.
- **Step 2.** While not converged, i.e., Eq. (5.24) and Eq. (5.25) do not hold
  - $j := j + 1$.
  - Find $\mathbf{W}^{(j+1)}$ using Eq. (5.18).
  - Find $\mathbf{E}^{(j+1)}$ using Eq. (5.19).
  - Find $\mathbf{J}^{(j+1)}$ using Eq. (5.21).
  - Update $\mathbf{Y}_1^{(j+1)}$ and $\mathbf{Y}_2^{(j+1)}$ using Eq. (5.22).
  - Update $\mu^{(j+1)}$ using Eq. (5.23).
- **Step 3.** Build adjacency matrix $\mathbf{C} = |\mathbf{W}| + |\mathbf{W}|^T$.
- **Step 4.** Calculate the normalised Laplacian $\mathbf{L}_\mathbf{C}$.
- **Step 5.** Compute the first $\hat{k}$ eigenvectors $\mathbf{u}_1, \ldots, \mathbf{u}_k$ of $\mathbf{L}_\mathbf{C}$ and let them form the columns of $\mathbf{U}$. The number of clusters $\hat{k}$ is estimated by Eq. (5.42).
- **Step 6.** Normalise the matrix $\mathbf{U}$ w.r.t. to its rows.
- **Step 7.** Cluster the rows of the normalised matrix $\mathbf{U}$ using k-means algorithm into $\hat{k}$ clusters [DHS12].

The initialisation of the algorithm can be performed with a warm start [BPC+11] if knowledge of the matrices is given, otherwise they are initialised with zero matrices. The output is given by $\mathbf{E}, \mathbf{X}_0$ and the segmented video.

Finally, the relative remission estimate $\hat{r}$ is obtained from the segmentation results as follows: For each segmented region, the intensity value of the underlying region is estimated as the median of all data vectors corresponding to this region. For the ROI this becomes

$$\hat{r} = \text{median}\{\mathbf{x}_i^{\text{ROI}} | i \in \text{ ROI segment}\}, \tag{5.43}$$

where $\mathbf{x}_i^{\text{ROI}}$ are the data points associated with the ROI.

5.4 Sequential Subspace Clustering

The UoS methods presented in the last sections all process the whole data vector in a batch. In the glucose application, the frames arrive sequentially and it is beneficial

to process them as they come in, in order to stop the processing when convergence is reached, hereby saving computation and delivering the result to the user as early as possible. In this section, sequential processing of UoS methods is discussed.

One advantage of UoS methods is that they are able to separate the data from underlying corruptions. This section proposes a method to sequentially perform LRR, subtracting the error matrix calculated in the previous step from the current data matrix. As presented in Sec. 5.2, the underlying error matrix $\mathbf{E}$ is a result of the LRR and SSC methods, hence they can be adopted in their standard forms. The proposed procedure can be executed frame-wise, i.e., by sequentially adding a new frame to the data matrix, or block-wise by sequentially processing blocks of frames. In the sequel, the proposed frame-wise approach is presented.

Recalling the LRR formulation [LLY10, LLY+13] and assuming Gaussian noise, Eq. (5.7) becomes

$$\begin{aligned} &\underset{\mathbf{W},\mathbf{E}}{\text{minimize}} \ \|\mathbf{W}\|_* + \zeta\|\mathbf{E}\|_F, \\ &\text{subject to } \mathbf{X} = \mathbf{X}\mathbf{W} + \mathbf{E}. \end{aligned} \tag{5.44}$$

It is typically solved by Augmented Lagrange Multiplier (ALM) methods [LCM10, Ber14]. Basically, ALM methods formulate the augmented Lagrangian. Hereby, they replace constrained optimisation problems by unconstrained problems and add a penalty term to the objective function. The LADMPSAP method, introduced previously, is considered a special case of ALM methods. Evidently, the LRR-problem results in a low-rank representation matrix $\mathbf{W}$ as well as an estimated error matrix $\mathbf{E}$.

The proposed sequential approach solves LRR at each time step $n = 1, \ldots, N$, calculating the error matrix $\mathbf{E}^{(n)}$ at each time step and updating the data matrix by subtracting the error from the old data matrix and appending the new incoming frame. For each incoming glucose frame $\mathbf{I}^{(n)}$, the data matrix is updated as

$$\mathbf{X}^{(n)} = [\hat{\mathbf{X}}^{(n-1)}; \mathrm{vec}(\mathbf{I}^{(n)})^T], \tag{5.45}$$

where $\hat{\mathbf{X}}^{(n-1)}$ is the error-free matrix from the previous step calculated as

$$\hat{\mathbf{X}}^{(n)} = \mathbf{X}^{(n)} - \mathbf{E}^{(n)}. \tag{5.46}$$

This process proceeds sequentially for every new incoming frame, until the final frame N has been reached. The resulting low-rank representation matrix $\mathbf{W}^{(N)}$ is used to form an affinity matrix $\mathbf{C}$ as in Eq. (5.40). Spectral clustering, as presented in

Sec. 5.3.4, using the resulting affinity matrix, is performed to segment the original data matrix.

The sequential LRR algorithm is presented in the sequel. It takes as its input the glucose frames $[\mathbf{I}^{(1)}, \mathbf{I}^{(2)}, \ldots, \mathbf{I}^{(N)}]$, as well as the threshold δ_{eigen} for spectral clustering. The first error-free data matrix is initialised as $\hat{\mathbf{X}}^{(0)} = [\text{vec}(\mathbf{I}^{(1)}), ..., \text{vec}(\mathbf{I}^{(n_D)})]^T$, i.e.,it contains the first glucose frames before the drop is detected (recall the kinetic behaviour of the glucose reaction in Sec. 2.4.1). For each subsequent incoming frame the following is performed:

- **Step 1.** For each incoming frame $n = \{1, \ldots, N\}$
  - $n = n + 1$.
  - Update the data vector to include the incoming frame as in Eq. (5.45).
  - Perform LRR using $\mathbf{X}^{(n)}$, obtaining $\mathbf{W}^{(n)}$ and $\mathbf{E}^{(n)}$.
  - Remove the calculated error matrix from the current data vector as in Eq. (5.46).
  - Update the data vector for the Next, time step as $\mathbf{X}^{(n+1)} = [\hat{\mathbf{X}}^{(n)}; \text{vec}(\mathbf{I}^{(n+1)})^T]$,
- **Step 2.** Perform LRR using $\mathbf{X}^{(N)}$, obtaining $\mathbf{W}^{(N)}$ and $\mathbf{E}^{(N)}$.
- **Step 3.** Construct an affinity matrix using $\mathbf{W}^{(N)}$.
- **Step 4.** Perform spectral clustering as in Sec. 5.3.4.

This procedure can be performed equivalently for SSC, resulting in a sequential SSC formulation.

5.5 Experimental Results and Discussion

This section evaluates the methods proposed in this chapter using Set C from Table 2.1, which contains $N_M = 78$ glucose measurements of $N_g = 16$ different glucose concentrations. Binning with $B = 5$, as presented in Sec. 3.2.1, is performed resulting in a frame size of $M \times N = 55 \times 22$. First, the ordered temporal subspace clustering methods developed in this thesis are evaluated and compared. Then, the results of sequential subspace clustering are presented

5.5.1 Sparse Subspace Clustering for Temporal Video Segmentation

Table 5.1: Parameters used for the experiments. The empty fields indicate that this value is not required for the corresponding method.

Method	ζ_1	ζ_2	μ	ϵ_1	ϵ_2	ρ	δ_{eigen}
SSC	0.1		0.01	$2 \cdot 10^{-4}$	$2 \cdot 10^{-4}$		0.01
RoSuRe	0.1		0.01	10^{-2}	10^{-2}	1.5	0.01
OSC	0.1	1	1	10^{-2}	10^{-4}	$\|\mathbf{X}\| \cdot 1.1$	0.01
ORoSuRe	0.6	1	0.01	10^{-2}	10^{-4}	$\|\mathbf{X}\| \cdot 1$	0.01
OTSSC	0.6	1	0.01	10^{-2}	10^{-4}	$\|\mathbf{X}\| \cdot 1.1$	0.01

The parameters used for the different methods are presented in Table 5.1. First, the segmentation results, presented in Fig. 5.2 and Fig. 5.3, are analysed. As described previously, at least three areas are expected to be visible in the image after it has converged: the ROI, the background area and a transition area between both. If artefacts are present, they represent a further segment in the image. Low glucose measurements usually just show two areas, as the change in colour is not strong. Figure 5.2 (a) is an artefact-free image with an underlying glucose concentration of 350 mg/dl. It is very clear to see in Fig. 5.2 (d) that ORoSuRe segments the three image regions correctly, identifying even the bridge between both parts of the ROI. The transition region between the ROI and the background is also segmented very clearly in this case. The SSC result in Fig. 5.2 (b) shows an over-segmentation behaviour and does not result in a smooth segmentation. The result is improved slightly when using RoSuRe in Fig. 5.2 (c), however the segmentation is patchy. The same behaviour is seen for the segmentation of Fig. 5.2 (e), which contains two air bubbles in the ROI and has an underlying concentration of 450 mg/dl. All methods manage to segment the artefacts. Again the proposed approach outperforms the others in clearly identifying the different areas in the image. Finally, Fig. 5.2 (i) shows a glucose image of low glucose concentration 90 mg/dl containing an artefact.

Generally the lower the glucose concentration, the more difficult the problem becomes as the contrast in the images decreases. Clearly, all methods are able to segment the artefact, and the differences in segmentation are not very significant. It is, however, noteworthy that SSC manages to detect the transition region between ROI and background, while all other methods interpret it as part of the background. More importantly, ORoSuRe outperforms the other methods in detecting the bridge between both segments of the ROI.

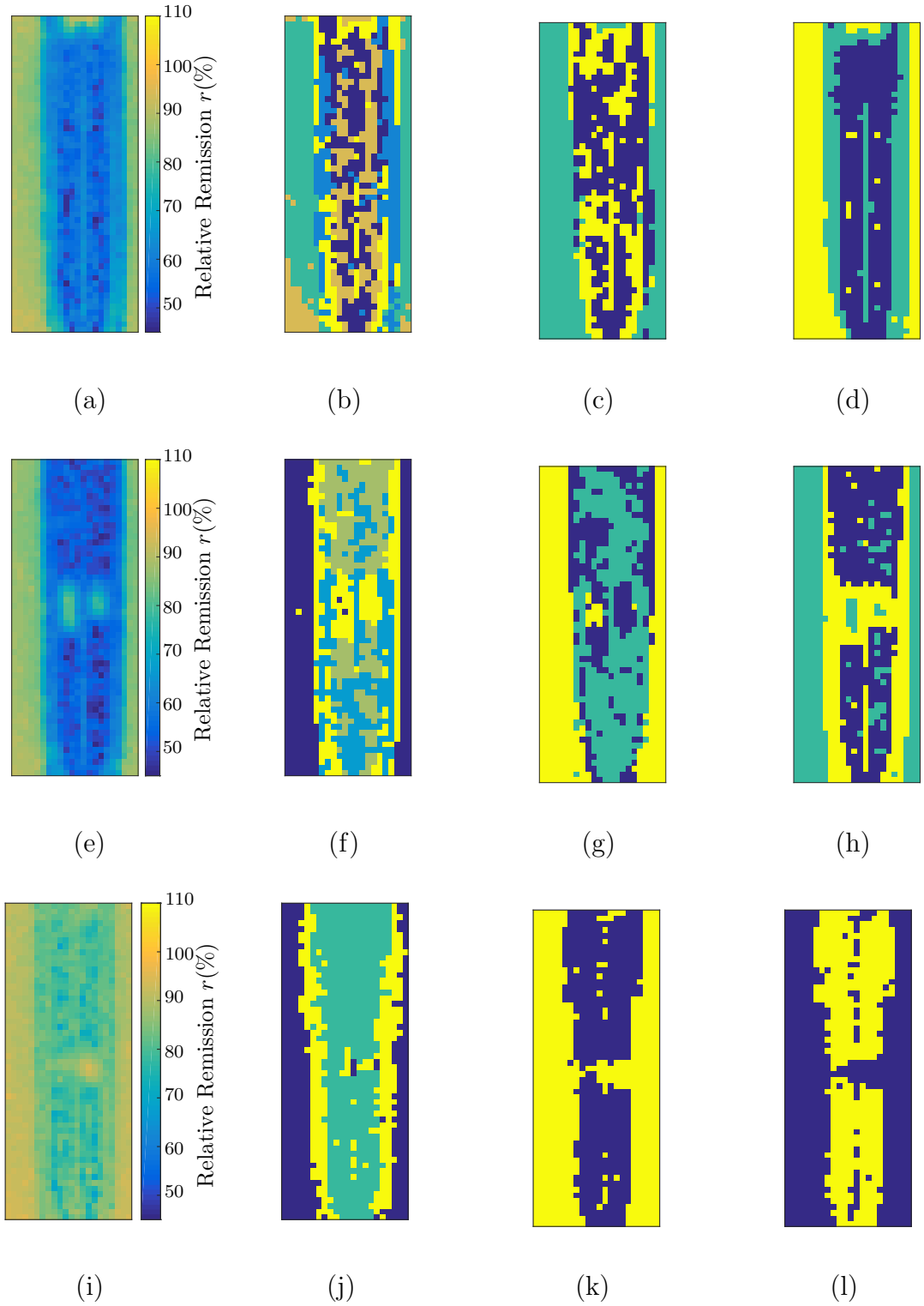


Figure 5.2: Examples of the original glucose images for an underlying glucose concentration of (a) 350 mg/dl, (e) 450 mg/dl and (i) 90 mg/dl. The corresponding segmentations using: (b), (f) and (j) SSC, (c), (g) and (k) RoSuRe, and Finally, (d), (h) and (l) ORoSuRe.

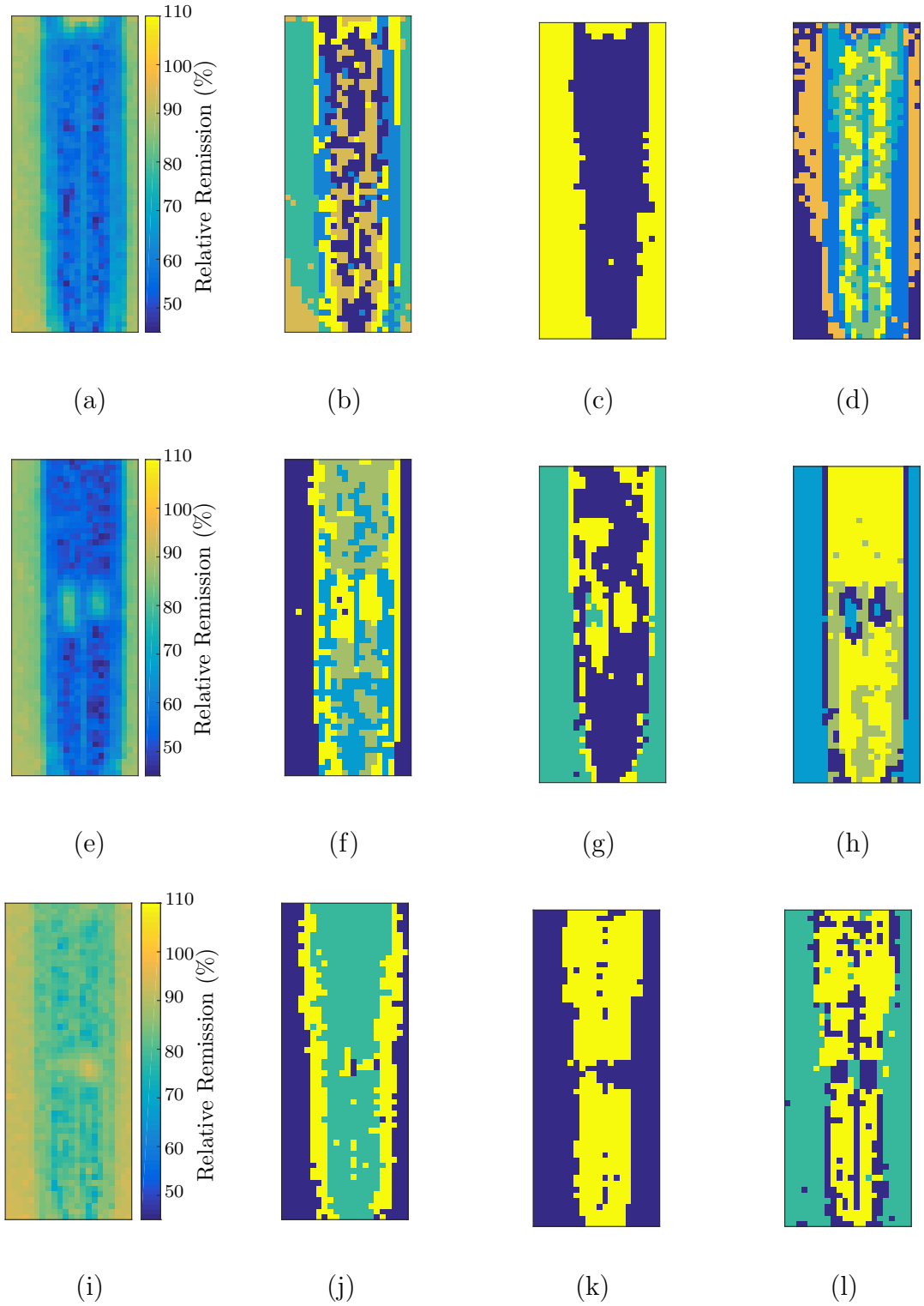


Figure 5.3: Examples of the original glucose images for an underlying glucose concentration of (a) 350 mg/dl, (e) 450 mg/dl and (i) 90 mg/dl. The corresponding segmentations using: (b), (f) and (j) SSC, (c), (g) and (k) OSC, and Finally, (d), (h) and (l) OTSSC.

Table 5.2: $CV_{\hat{r}}$ and gMAD comparison for all methods. Low and high indicate low glucose concentration values $\leq 75\,\mathrm{mg/dl}$ and high glucose concentration values $> 75\,\mathrm{mg/dl}$.

Method	SSC	OSC	RoSuRe	ORoSuRe	OTSSC
$CV_{\hat{r}}$	1.37	0.93	0.96	**0.83**	0.88
gMAD	12.09	7.62	7.74	**6.93**	7.17
gMAD - low	11.80	6.21	6.34	**5.01**	6.24
gMAD - high	12.70	8.03	8.16	**7.59**	7.47

Coming to the segmentation results of OTSSC in Fig. 5.3, we can see that OTSSC over-segments the first image, resulting in a segmentation that is close to SSC. In this case, OSC in Fig. 5.3 (c) achieves a very smooth segmentation, but fails at detecting the transition region and the bridge. For the second image in Fig. 5.3 (e) OTSSC seems to perform slightly better than the other methods. The best performance of OTSSC is identifiable for the low glucose image in Fig. 5.3 (i). Here, OTSSC seems to outperform even ORoSuRe by identifying both the bridge and the transition region, as well as the ROI and the background.

These results underline the importance of the assumption made for ORoSuRe in most cases, i.e., that most of the data samples may be corrupted by noise. However, for low glucose cases OTSSC presents a more suitable assumption. The images in this case are more homogenous in colour and the different regions are not as clearly separated as in the high glucose case. Assuming that all data samples are corrupted, may lead to actual data samples be mistaken for noise in the ORoSuRe case. This is the reason why OTSSC obtains better segmentation results for low glucose cases.

Next, the quantitative analysis, presented in Table 5.2, is discussed. For the relative remission variation coefficient $CV_{\hat{r}}$, the proposed ORoSuRe approach clearly outperforms all others. OTSSC comes in second place, which proves the necessity of the noise model in Eq. (5.9). SSC shows particularly poor results, which is due to the over-segmentation behaviour. The gMAD results show the same trend with ORoSuRe clearly outperforming the other methods, and OTSSC coming in second place.

Subsequently the CEG analysis, presented in Fig. 5.4, is evaluated. The overall results for OTSSC, RoSoRe and ORoSuRe are very good, as all points lie in the A-region, signifying clinically accurate results. SSC shows a larger spread of the points lying in the mid-glucose-level region. For the high glucose concentrations ORoSuRe outperforms OTSSC and RoSuRe.

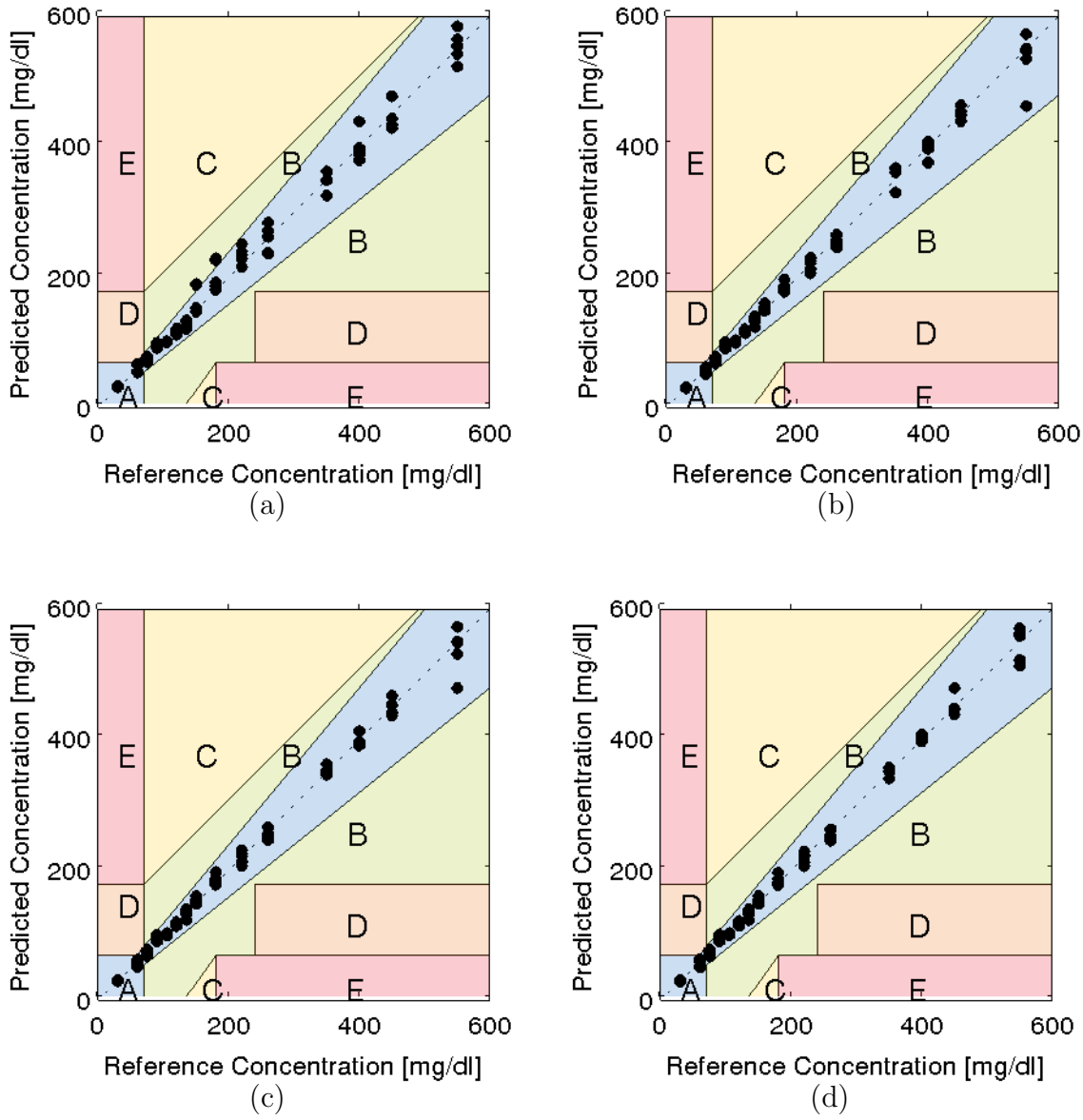


Figure 5.4: Clarke's Error Grid Analysis for (a) SSC, (b) RoSuRe, (c) OTSSC, and (d) ORoSuRe.

Finally, the effect of the regularisation parameter ζ_1 on the performance of ORoSuRe is analysed. Figure 5.5 shows the performance of ORoSuRe, given by the gMAD criterion, for different values of ζ_1. For $\zeta_1 < 0.7$ the performance is stable and does not vary significantly. This analysis is omitted for OTSSC, as this method does not show a dependancy on ζ_1.

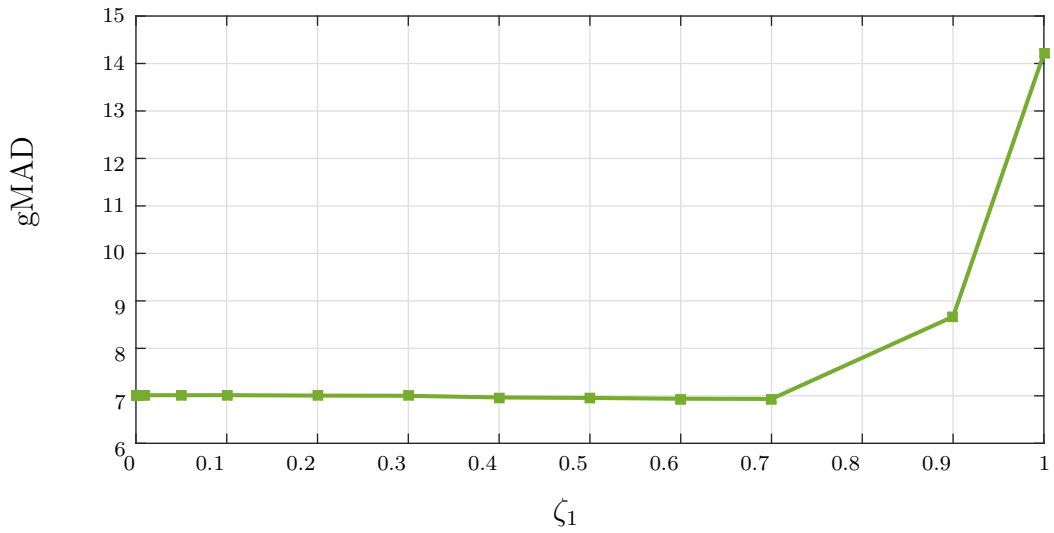


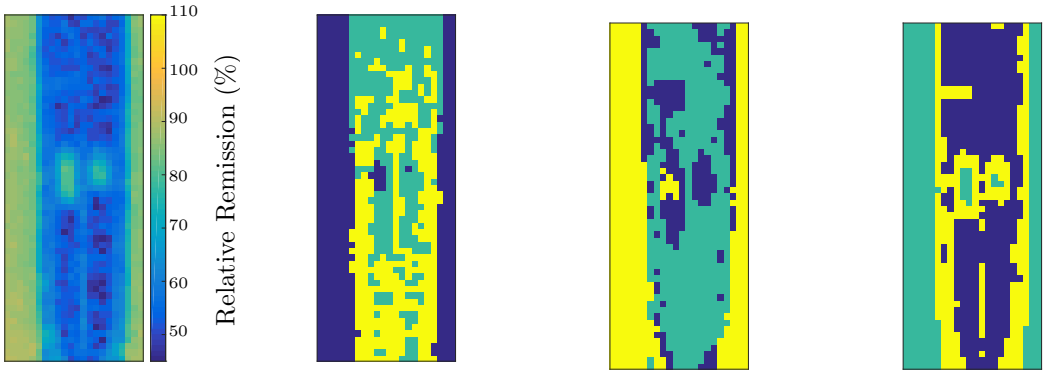
Figure 5.5: Analysis of the effect of the parameter ζ_1 on the performance of ORoSuRe.

5.5.2 Sequential Subspace Clustering

The sequential LRR algorithm corrects the data matrix by the previously calculated error as presented in Sec. 5.4. The aim behind processing the frames as they arrive sequentially is that a correct segmentation can be achieved using a much lower number of frames, due to the error correction performed for each frame. The algorithm is set to consider one frame every $t_{\text{skip}} = 1/10\,\text{s}$, which is equivalent to skipping three captured frames. Sequential LRR is compared to the LRR algorithm from Eq. (5.7), which operates in batch mode, as well as to the spatial LRR algorithm from Eq. (5.8).

First, the segmentation results are evaluated in Fig. 5.6. Sequential LRR considers only the first 200 frames, while LRR and spatial LRR process the whole 582 frames as a batch. Figure 5.6 (a) shows a glucose image from a measurement containing 350 mg/dl, contaminated by an air bubble. The segmentation obtained by LRR, shown in Fig. 5.6 (b), does not clearly identify the different regions of the image. Spatial LRR improves the segmentation slightly, by producing more connected regions. Sequential LRR, presented in Fig. 5.6 (d), performs best as it is able to segment the image correctly into its different regions at a much earlier stage of the chemical reaction.

In Table 5.3 the $\text{CV}_{\hat{r}}$ and gMAD values are presented for LRR, spatial LRR and their sequential versions. Sequential spatial LRR, clearly, achieves better results in terms of remission variation coefficient for most cases. This can be attributed to



(a) (b) (c) (d)

Figure 5.6: (a) Original image with a glucose concentration of 450 mg/dl, (b) segmentation using standard LRR at frame $N = 582$, (c) segmentation using spatial LRR at frame $N = 582$, (d) segmentation using sequential LRR at frame $N = 200$.

Table 5.3: $CV_{\hat{r}}$ and gMAD comparison for LRR and spatial LRR, comparing both the batch and sequential versions. Low and high indicate low glucose concentration values ≤ 75 mg/dl and high glucose concentration values > 75 mg/dl.

Method	LRR	SpatLRR	Seq. LRR	Seq. SpatLRR
$CV_{\hat{r}}$	1.16	1.01	**0.91**	0.99
gMAD	10.20	7.96	7.48	**5.33**
gMAD - low	6.38	6.95	6.22	**5.81**
gMAD - high	11.20	8.29	7.91	**4.46**

the removal of the error in each step from the data. Both methods show clinically accurate results, however spatial LRR seems to outperform regular LRR w.r.t. the variance of same-glucose measurements. Table 5.3 underpins this finding, as both the $CV_{\hat{r}}$ and the gMAD results of spatial LRR are better, specifically for high glucose concentrations.

Next, Clarke's Error Grid Analysis is performed to evaluate the accuracy of glucose estimates obtained using LRR, spatial LRR and sequential LRR in Figure 5.7. All three methods produce results that fall in zones A and B, which indicates clinically accurate results. The sequential LRR produces estimates that show a lower variance for the same underlying glucose concentration.

Figure 5.8 portrays the computation times for segmenting one image using the different UoS methods presented in this chapter. As mentioned previously, these

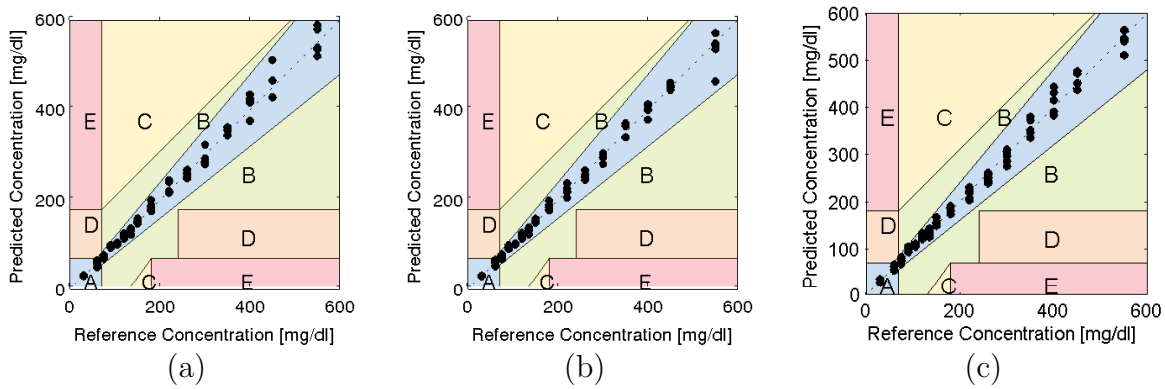


Figure 5.7: Clarke's Error Grid Analysis for (a) LRR, (b) spatial LRR and (c) sequential LRR.

values have been obtained using the MATLAB simulation runtimes and should not be taken in their absolute sense. Rather, they merely serve as a comparative indicator of computational complexity between the different methods. Evidently, SSC and RoSuRe are the most efficient among all methods. The reason for this, is that in both cases a two-block optimisation problem is solved. All other cases solve three-block optimisation problems. However, much research is currently being performed in deriving algorithms to solve such problems efficiently [TGG15]. Comparing them to the mean-shift methods of Chap. 4, it can be concluded that UoS methods outperform all mean-shift methods in efficiency. The only exception is given by sparse-mean shift which compares to SSC and RoSuRe in efficiency.

5.6 Summary and Conclusions

In this chapter the Union of Subspaces (UoS) model has been introduced to the blood glucose estimation problem. Using methods of sparse subspace clustering and low-rank representation, the glucose videos are segmented not only based on the intensity value of the ROI but also on the temporal evolution of the intensity value of the ROI compared to the other surrounding regions.

First, two novel sparse subspace clustering formulations, namely Ordered Robust Subspace Recovery (ORoSuRe) and Ordered Temporal Sparse Subspace Clustering (OTSSC), have been derived that include a term to enforce neighbouring pixels to be similar, hereby producing smoother segmentations, as well as including a term that favours the choice of only few representatives to describe the remaining data vectors.

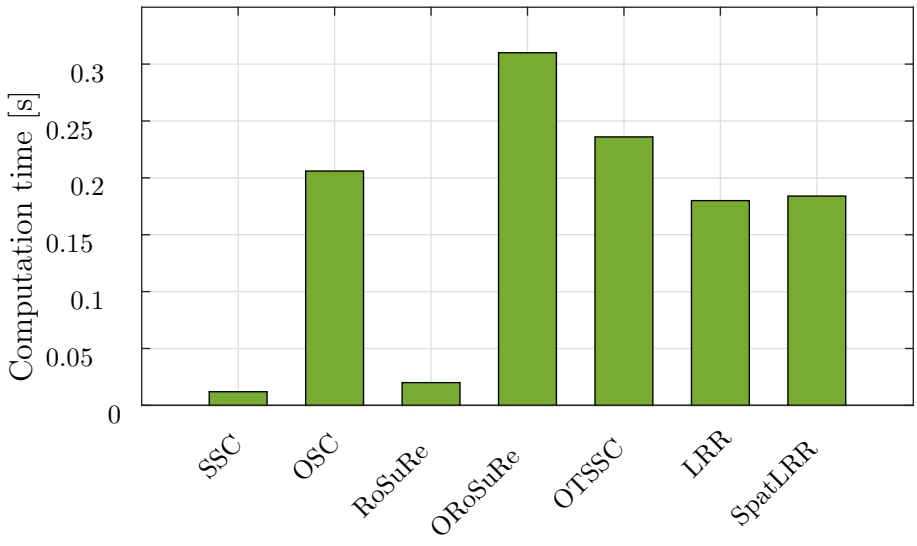


Figure 5.8: Computation time for the segmentation of one image, for the different UoS algorithms.

The first proposed formulation assumes that all data vectors may be corrupted by noise, while the second formulation relaxes this assumption by assuming that only a few data vectors will be corrupted. Both optimisation problems have been solved by the recently proposed LADMPSAP.

Experiments using real glucose data sets have been performed, elucidating that both methods produce much smoother and more meaningful segmentations than state-of-the-art methods. The stricter noise assumption shows better results in most cases. A quantitative analysis has been performed for all methods using the previously presented validation methods. ORoSuRe has outperformed all other methods both w.r.t. accuracy and precision. OTSSC has maintained the second place while still outperforming the state of the art.

Furthermore, a sequential version of subspace clustering has been presented that processes the frames sequentially. It calculates an error matrix for each time instant and uses it to correct the Next, frame. The sequential approach has been applied to LRR and has been compared to standard and spatial LRR. The sequential approach has not only been able to produce better segmentation results but has done so at earlier stages of the chemical reaction, due to the error correction step.

Chapter 6
Conclusions and Future Work

In this thesis, glucose concentration estimation from an image-based photometric measurement principle, intended for use in hand-held glucometers, has been addressed. This thesis develops signal and image processing methods to accurately estimate the glucose concentration from the photometric images and reduce measurement time. Saving in computational complexity is given importance, as the glucose meters need to be small in size and cheap in price. However, accuracy remains the most important requirement. A comparison of the achieved accuracy using the developed techniques to the accuracy of state-of-the-art blood glucose meters has not been possible, as this information is typically not disclosed by the respective companies. Therefore, throughout the thesis, the results have been benchmarked w.r.t. the most recent ISO standards for glucose meters [Int13].

It is important to note that the framework presented in this thesis, as well as the developed methods are not confined to the glucose measurement problem. Recent years have shown much research in diagnosis methods that use pinprick blood samples and are based on the photometric principle. The methods developed in this thesis can easily be adapted to such problems.

Section 6.1 provides a summary and the main conclusions of the work performed in this thesis. Section 6.2 highlights recommendations for future work.

6.1 Conclusions

6.1.1 A Framework for Image-Based Photometry in Glucose Measurements

As much of the work in diabetes care is proprietary, one aim of this work has been to develop and present a complete processing chain to estimate the glucose concentration from the raw photometric images. This framework has been presented in Chap. 3.

First, a pre-processing scheme has been developed that constitutes cropping the raw images. This has been followed by normalising the cropped images with the help

of calibration frames recorded prior to the chemical reaction. Subsequent binning of the images has led to the reduction of noise, as well as the reduction of the image size. A suitable binning size has been identified that offers a trade-off between noise reduction and the conservation of the image structure. This has been followed by the derivation of a hypothesis test to detect the onset of the chemical reaction. The purpose of this is to avoid unnecessary computation prior to the begin of the chemical reaction.

A tool has been developed to detect faulty glucose measurements at an early stage. To this end, texture analysis based on Haralick's grey-level co-occurrence matrices has been employed. The feature maximum probability has proved to be the most suitable in detecting faulty measurements.

The framework continues by segmenting the glucose images and estimating the relative remission value of each image. This results in an estimated kinetic curve, that describes the chemical reaction. As the segmentation step constitutes a major part of this thesis, the conclusions relating to segmentation are presented separately in Sec. 6.1.2.

Given the estimated kinetic curve, the framework proceeds with convergence detection of the chemical reaction. The converged remission value is, then, mapped to an estimate of the glucose concentration. The conclusions of the work done in convergence detection are summarised in Sec. 6.1.3.

Extensive experiments have been performed applying the developed framework to six different real data sets. Of these six data sets four were obtained using blood samples volumes in the range 10 nl to 100 nl, one set was obtained using blood sample volumes of 1 nl and one set was obtained using blood sample volumes in the µl-range. This study has shown that using blood sample volumes of only 10 nl to 100 nl high measurement accuracy can be achieved, which is comparable to the accuracy of state-of-the-art blood sample volumes in the µl range. Furthermore, a lower limit of 1 nl in blood sample volume has been established to be at the edge of what is acceptable in terms of clinical accuracy.

6.1.2 Segmentation of Glucose Images

The segmentation of the glucose images can be a tricky problem. The reaction region, termed region of interest (ROI) which contains the sought-after colour change,

depends on the blood flow of the blood sample over the test strip and on the chemical reagent. This can be inhomogeneous over the test strip. This results in a ROI that is not necessarily spatially connected. In this thesis, segmentation has been interpreted as a clustering problem, where the different regions are represented by cluster centres indicating the intensity value of the respective region. The number of clusters in this problem is unknown. Two different approaches have been developed in this thesis to perform the segmentation: the mean-shift (MS) approach presented in Chap. 4, and the union of subspaces (UoS) approach presented in Chap. 5. The conclusions of each approach are summarised in the sequel.

The problem of kernel and bandwidth selection for the mean-shift approach has been examined. It has been established that using a spatial kernel can degrade the performance of mean-shift as the ROI is often disconnected. Furthermore, the choice of a fixed bandwidth parameter has been compared to an adaptive bandwidth. The adaptive bandwidth has led to a slight improvement of high-glucose measurements. The reason for this is that high glucose measurements are characterised by a multi-modal pdf. No effect could be established for the low and mid-level glucose measurements. Interpreting the kernel density estimator in terms of the reproducing kernel Hilbert space has allowed for the derivation of a robust mean-shift variation and a sparse mean-shift variation. The robust mean-shift has been proved to improve accuracy in heavy-tailed noise scenarios. The sparse mean-shift has enabled the reduction of computation. For this purpose, a subspace of the original data vector is selected on the basis of an incoherence criterion. It has been shown that when the subset size is large enough a gain in estimation accuracy can be achieved, as a noise reduction effect sets in. An approach has been developed to choose a suitable subspace cardinality. The medoid-shift approach has also been considered in this thesis and has outperformed the mean-shift approach. A robust and a sparse variation of the medoid-shift have been derived. It has been established that the combination of the robust and sparse approaches clearly improves the performance of the sparse methods. Finally, the convergence proofs of all derived algorithms have been provided.

While the mean-shift approach segments each image separately, the UoS model uses the intensity information in each image, as well as the temporal evolution of the intensity over the frames to perform clustering. Two novel formulations of sparse subspace clustering have been derived. They have shown to be better suited than the state of the art for segmenting the glucose video data. They incorporate a term enforcing neighbouring pixels in the vertical direction to be similar, hereby resulting in smoother segmentations. Furthermore, a term is included that favours the choice of only few representatives making the methods better suited for segmentation. Ordered

Robust Subspace Recovery (ORoSuRe) has assumed that all data vectors may be corrupted by noise. It has been established that this noise model is in most cases better suited for the glucose segmentation problem. Ordered Temporal Subspace Clustering (OTSSC) has assumed that only few data vectors may be corrupted. This method has shown superiority over ORoSuRe in low glucose cases only. In these cases the strong noise assumption can lead to falsely clustering data samples as noise samples. Both methods have outperformed the state of the art in all experiments.

Finally, a sequential subspace clustering approach has been applied to LRR. It has been shown that estimating the error matrix from previous frames and using it to correct subsequent clustering leads to better and earlier segmentations, producing more meaningful segmentation results and strongly improving accuracy of the estimates.

A comparison of the mean-shift approach to the UoS approach for segmentation of the glucose images shows similar results in accuracy. The mean-shift approaches are advantageous in that they do not require knowledge on the number of clusters, but provide this as an output as long as the bandwidth parameter is chosen well. Furthermore, the mean-shift approaches combine segmentation of the images with the estimation of the relative remission value. UoS approaches on the other hand require either knowledge on the number of subspaces or a method to estimate this number as shown in Chap. 5. However, the advantages of UoS approaches is that they are able to process several frames simultaneously, hereby incorporating more information from the data as well as requiring less computation. Furthermore, the segmentation results obtained using the developed ORoSuRe and OTSSC show better delineation of the different image regions. Finally, UoS methods are more efficient.

6.1.3 Kinetic Filtering and Convergence Detection

There is a limited number of available studies regarding the issue of convergence detection for glucose kinetics. To the best of the author's knowledge convergence in state-of-the-art devices is detected by analysing the slope of the kinetic curve. This thesis has shown that this approach can easily lead to erroneous glucose concentration estimates. An existing non-linear model of the chemical kinetics has been extended in this work. A linear relation between the reaction rate and the convergence value of the chemical reaction has been established. The derivation of a model of the chemical kinetics that is only dependent on the convergence value has enabled using state estimation approaches to estimate the convergence value with higher accuracy and ahead of time. For this purpose, an extended Kalman filter (EKF) approach

and a region-based particle filter (RBPF) approach have been employed. Extensive results have shown that the EKF outperforms the state of the art both in terms of estimation accuracy as well as measurement time. A gain in accuracy of up to approximately 2% in relative remission has been achieved and a gain in measurement time of around $30 - 40\%$. The RBPF, on the other hand, has been able to drastically reduce measurement time by more than 50%. However, the accuracy of estimates did not improve w.r.t. to the state of the art. In some cases the RBPF even showed a degradation in accuracy.

These results highlight that using state-estimation techniques combined with a derived model of the chemical kinetics can drastically improve convergence detection both w.r.t. to accuracy and measurement time. The EKF is considered more suitable for this application, as accuracy constitutes the most important objective in glucose measurement.

6.2 Future Work

6.2.1 Segmentation of Glucose Images

The segmentation using the mean-shift algorithm requires 0.4 s to 1.6 s per image, depending on the used variation. This is hardly acceptable in a real-time scenario. To speed up the process, incremental versions of mean-shift and medoid-shift should be derived. For this purpose, only an update of previous segmentation results and mode location estimates with the help of the incoming frame should be performed. As the glucose images change gradually over time, the difference between consecutive frames is not large and an incremental approach would greatly reduce computation.

An incremental approach would also be beneficial for subspace clustering methods. Here the segmentation of each image requires 0.02 s to 0.2 s. In this thesis subspace clustering has been performed in batch for the whole video as well as sequentially, recalculating the subspace clustering problem for each incoming frame. An incremental approach would process the data sequentially, only updating the coefficient matrix for each incoming frame instead of re-calculating it.

Furthermore, the presented sequential approach requires calculation of subspace clustering for each incoming frame. In a real-time scenario a block-sequential approach would be more suitable. This has, in fact, been implemented for a block size of

$1\,\mathrm{s} \approx 30$ frames and has proved to outperform the sequential approach. The effect of the block size should be thoroughly analysed in future work.

6.2.2 Kinetic Filtering and Convergence Detection

In the real data sets used throughout this thesis the only variable has been the glucose concentration. Other affecting parameters such as surrounding temperature, humidity or haematocrit values in the blood sample have been kept constant. These parameters have been shown to greatly affect the kinetics of the chemical reaction [SNL13]. The derived kinetic model in Chap. 3 would greatly benefit from considering these affecting parameters. To this end, a measurement series in which these parameters are varied should be performed to better understand and analyse the effect of these parameters.

6.2.3 Clinical Validation

The developed concepts for image segmentation and convergence detection perform well in the experiments performed with real data. The results lie in the range of permissible accuracy according to the most recent ISO standards. The next logical step is to perform clinical tests to validate the developed techniques for implementation in real glucose meters. The derived model of the chemical kinetics in Chap. 3 should be adapted accordingly.

List of Acronyms

ADM Alternating Direction Method

ADMM Alternating Direction Method of Multipliers

AKDE Adaptive Kernel Density Estimator

ALM Augmented Lagrange Multiplier

AMISE Asymptotic Mean Integrated Squared Error

BIC Bayesian Information Criterion

CCD Charge-Coupled Device

CEG Clarke Error Grid

CV Coefficient of Variation

EKF Extended Kalman Filter

EM Expectation Maximization

GDM Gestational Diabetes Mellitus

GLCM grey-level co-occurrence matrix

gMAD Glucose-Specific Mean Absolute Deviation

IDF International Diabetes Foundation

i.i.d. independently identically distributed

ISO International Standards Organisation

IRWLS Iteratively Re-Weighted Least-Squares

KDE Kernel Density Estimator

KIRWLS Kernalised Iteratively ReWeighted Least-Squares

LADM Linearized Alternating Direction Method

LADMPSAP Linearized Alternating Direction Method with Parallel Splitting and Adaptive Penalty

LRR Low Rank Representation

LS Least-Squares

MAD Mean Absolute Deviation

MISE Mean Integrated Squared Error

MMSE Minimum Mean Squared Error

MS Mean-Shift

MedS Medoid-Shift

MSE Mean Squared Error

ORoSuRe Ordered Robust Subspace Recovery

OSC Ordered Subspace Clustering

pdf probability density function

PF Particle Filter

POC Point-of-Care

RBPF Region-based Particle Filter

RKHS Reproducing Kernel Hilbert Space

R-MS Robust Mean-Shift

R-MedS Robust Median-Shift

ROI Region Of Interest

RoSuRe Robust Subspace Recovery

RS-MS Robust Sparse Mean-Shift

RS-MedS Robust Sparse Medoid-Shift

SE Structuring Element

SpatLRR Spatial Low Rank Representation

SpatSC Spatial Subspace Clustering

SSC Sparse Subspace Clustering

S-MS Sparse Mean-Shift

S-MedS Sparse Medoid-Shift

List of Symbols

The following list contains symbols in alphabetic order, which occur more than once in this thesis. The remaining symbols are directly introduced where they are used.

$\mathbf{1}$	Vector of ones
$\underset{x}{\operatorname{argmax}}\, f(x)$	Returns the value of x that maximizes $f(x)$
$\underset{x}{\operatorname{argmin}}\, f(x)$	Returns the value of x that maximizes $f(x)$
$\operatorname{diag}(\cdot)$	Returns a diagonal matrix when the argument is a vector, or returns a vector containing the elements of the main diagonal when the argument is a matrix
B	Binning value
$\mathbf{C}$	Affinity matrix
$\det(\cdot)$	Returns the determinant of a square matrix
$\in$	Element of
$\mathrm{E}\{\cdot\}$	Expectation operator
h	Bandwidth parameter
$\mathcal{H}$	Hilbert space
$\mathcal{H}_0$	Null hypothesis
$\mathcal{H}_1$	Alternative hypothesis
n	Discrete time value
n_C	Convergence time of the chemical reaction
n_D	Drop time of the chemical reaction
f_s	Sampling frequency
$f_{\mathcal{K}}$	Kernel density estimator with kernel function $\mathcal{K}$
$\mathbf{f_x}$	probability density function of $\mathbf{x}$
g	Glucose concentration value
$\hat{g}$	Glucose concentration estimate
$\mathbf{I}$	Matrix containing a glucose image
$\mathbf{I}^{(n)}$	Matrix containing a glucose image obtained at time n
$\mathcal{I}$	Set of indices
$\mathcal{I}^*$	Optimal set of indices
j	Iteration step
J	Number of iterations
$K(\cdot)$	Profile of a kernel
$\mathcal{K}(\mathbf{x})$	Kernel density estimator of $\mathbf{x}$

L	Total number of pixels in an image
L_{red}	Reduced number of pixels in an image using the sparse approach
$\mathbf{m}_{\mathbf{H},\mathcal{K}}^{(\text{MS})}(\mathbf{x})$	Mean-shift vector of $\mathbf{x}$ obtained with bandwidth h and kernel function $\mathcal{K}(\cdot)$
M_x	Number of rows in a matrix
M_y	Number of columns in a matrix
N	Number of frames in a measurement
N_g	Number of glucose concentrations in a set
N_M	Number of measurements in a set
N_γ	Number of grey-levels in an image
P_{FA}	Probability of false alarm
$Q(\cdot)$	Complementary cumulative distribution function
r	Relative remission value
$\hat{r}$	Estimate of the remission value
r_C	Convergence relative remission value
r_D	Relative remission value at the drop time n_D
$\hat{\mathcal{R}}_g$	A set of remission estimates pertaining to a glucose concentration g
$\mathbb{R}$	Set of real numbers
t	Continuous time value
$\text{vec}(\cdot)$	Vectorisation operator
$\mathbf{w}$	Vector of weights
w_l	l-th weight
$\mathbf{x}$	Vectorised image
x_l	l-th element of $\mathbf{x}$
$\tilde{\mathbf{x}}$	Mean-adjusted version of $\mathbf{x}$
$\mathbb{Z}$	Set of integer numbers
$(\cdot)^{\text{T}}$	Transpose of a vector or matrix
$(\cdot)^{-1}$	Inverse of a square matrix
$\lvert a \rvert$	Absolute value of a scalar a
$\lvert \mathcal{A} \rvert$	Cardinality of a set $\mathcal{A}$
$\lVert \cdot \rVert$	Euclidean norm or 2-norm of a vector
$\lVert \cdot \rVert_1$	1-norm of a vector
$\lVert \cdot \rVert_{1,2}$	$l_{1,2}$-norm of a vector
$\lVert \cdot \rVert_F$	Frobenius of a vector
$\lVert \cdot \rVert_*$	Nuclear norm of a vector
$\langle \cdot \rangle$	Inner product

$\boldsymbol{\alpha}$	Vector of sparse weights
α_i	i-th sparse weight
γ	Grey-level values
λ	Eigenvalue
$\Phi(\cdot)$	Mapping function to the reproducing kernel Hilbert space
$\rho(\cdot)$	Robust loss function
$\Psi(\cdot)$	Robust score function
ν	Incoherence function
$\tilde{\nu}$	Normalised incoherence function
σ^2	Noise variance
τ	Decay rate of the chemical reaction
Υ	Resolution of an image
ζ	Regularisation parameter

Bibliography

[Abr82] I. S. Abramson, "On bandwidth variation in kernel estimates-a square root law," *The Annals of Statistics*, vol. 10, no. 4, pp. 1217–1223, 1982.

[AF05] R. M. Anderson and M. M. Funnell, "Patient empowerment: reflections on the challenge of fostering the adoption of a new paradigm," *Patient education and counseling*, vol. 57, no. 2, pp. 153–157, 2005.

[AF10] ——, "Patient empowerment: myths and misconceptions," *Patient education and counseling*, vol. 79, no. 3, pp. 277–282, 2010.

[AH01] S. Aksoy and R. Haralick, "Feature normalization and likelihood-based similarity measures for image retrieval," *Pattern Recognition Letters*, vol. 22, no. 5, pp. 563–582, April 2001.

[AHKP11] J. Asfour, H. Haar, G. Kocherscheidt, and W. Petrich, "Analyse Optischer Daten Mit Hilfe Von Histogrammen." December 2011, European Patent Application, EP1843148A1.

[Alv08] S. Alva, "FreeStyle Litea blood glucose meter that requires no coding," *Journal of diabetes science and technology*, vol. 2, no. 4, pp. 546–551, 2008.

[Ame10] *Diagnosis and classification of diabetes mellitus*, American Diabetes Association Std. Supplement 1, 2010.

[AMGC02] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking," *IEEE Transactions on Signal Processing*, vol. 50, no. 2, pp. 174–188, 2002.

[AMM09] U. Ali, M. B. Malik, and K. Munawar, "FPGA/soft-processor based real-time object tracking system," in *5th Southern Conference on Programmable Logic. SPL.*, April 2009, pp. 33–37.

[BBW89] P. J. Besl, J. B. Birch, and L. T. Watson, "Robust window operators," *Machine Vision and Applications*, vol. 2, no. 4, pp. 179–191, 1989.

[BDE+09] M. Bass, C. DeCusatis, J. Enoch, V. Lakshminarayanan, G. Li, C. Macdonald, V. Mahajan, and E. Van Stryland, *Handbook of optics, Volume II: Design, fabrication and testing, sources and detectors, radiometry and photometry.* McGraw-Hill, Inc., 2009.

[Bee52] A. Beer, "Bestimmung der Absorption des rothen Lichts in farbigen Fluessigkeiten," *Annalen der Physik und Chemie*, vol. 86, p. 7888, 1852.

[Ber14] D. P. Bertsekas, *Constrained optimization and Lagrange multiplier methods.* Academic press, 2014.

[Bis08] H. Bisswanger, *Enzyme kinetics: principles and methods.* John Wiley & Sons, 2008.

[BK14] X. Bian and H. Krim, "Robust Subspace Recovery via Bi-Sparsity Pursuit," *arXiv preprint arXiv:1403.8067*, 2014.

[BK15] ——, "Bi-Sparsity pursuit for robust subspace recovery," in *IEEE International Conference on Image Processing (ICIP)*, 2015, pp. 3535–3539.

[BM03] G. R. D. B. Brent Modzelewski, Steven Gilmour, "Method for determining concentration of an analyte in a test strip," April 2003, US Patent 6,541,266.

[BOW95] E. Baumann, W. Obermeier, and K. Werner, "Light source pulsed with irregular pulse sequence in analog photometric signal evaluation for a test carrier analysis system," October 1995, US Patent 5,463,467.

[BPC+11] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends® in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.

[CC97] B. Cameron and G. Cote, "Noninvasive glucose sensing utilizing a digital closed-loop polarimetric approach," *IEEE Transactions on Biomedical Engineering*, vol. 44, no. 12, pp. 1221–1227, Dec 1997.

[CCAG09] C. Chevrefils, F. Cheriet, C. Aubin, and G. Grimard, "Texture Analysis for Automatic Segmentation of Intervertebral Disks of Scoliotic Spines From MR Images," *IEEE Transactions on Information Technology in Biomedicine*, vol. 13, no. 4, pp. 608–620, April 2009.

[CD05] D. A. Clausi and H. Deng, "Design-based texture feature fusion using Gabor filters and co-occurrence probabilities," *IEEE Transactions on Image Processing*, vol. 14, no. 7, pp. 925–936, 2005.

[CF12] S. Clarke and J. R. Foster, "A history of blood glucose meters and their role in self-monitoring of diabetes mellitus," *British Journal of Biomedical Science*, vol. 69 (2), pp. 83–93, March 2012.

[Che95] Y. Cheng, "Mean shift, mode seeking, and clustering," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 17, no. 8, pp. 790–799, 1995.

[Chu97] F. R. Chung, *Spectral graph theory*. American Mathematical Soc., 1997, vol. 92.

[CLMW11] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *Journal of the ACM (JACM)*, vol. 58, no. 3, p. 11, 2011.

[CM02] D. Comaniciu and P. Meer, "Mean shift: a robust approach toward feature space analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 5, pp. 603–619, 2002.

[Com00] D. Comaniciu, "Nonparametric robust methods for computer vision," Ph.D. dissertation, Rutgers, The State University of New Jersey, 2000.

[Com03] ——, "An algorithm for data-driven bandwidth selection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 2, pp. 281–288, 2003.

[CP00] M. A. Carreira-Perpinán, "Mode-finding for mixtures of Gaussian distributions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 11, pp. 1318–1323, 2000.

[CP06] ——, "Acceleration Strategies for Gaussian Mean-Shift Image Segmentation," vol. 1, pp. 1160–1167, 2006. [Online]. Available: http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=1640881

[CP15] ——, "A review of mean-shift algorithms for clustering," *arXiv preprint arXiv:1503.00687*, 2015.

[CR00] D. Comaniciu and V. Ramesh, "Mean shift and optimal prediction for efficient object tracking," in *Proceedings International Conference on Image Processing (ICIP)*, vol. 3, 2000, pp. 70–73.

[CRM00] D. Comaniciu, V. Ramesh, and P. Meer, "Real-time tracking of non-rigid objects using mean shift," in *Proceedings IEEE Conference on Computer Vision and Pattern Recognition*, vol. 2, 2000, pp. 142–149.

[CRM01] ——, "The variable bandwidth mean shift and data-driven scale selection," in *Proceedings Eighth IEEE Int. Conf. Computer Vision (ICCV)*, vol. 1, 2001, pp. 438–445.

[CS14] E. C. Cortés and C. Scott, "Scalable sparse approximation of a sample mean," in *IEEE 39th International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2014.

[CS15] ——, "Sparse Approximation of a Kernel Mean," *ArXiv e-prints*, Mar. 2015.

[CWG+13] S. Craciun, G. Wang, A. D. George, H. Lam, and J. C. Principe, "A scalable RC architecture for mean-shift clustering," in *IEEE 24th International Conference on Application-Specific Systems, Architectures and Processors*, June 2013, pp. 370–374.

[DFFC12] S. Del Favero, A. Facchinetti, and C. Cobelli, "A Glucose-Specific Metric to Assess Predictors and Identify Models," *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 5, pp. 1281–1290, May 2012.

[DHS12] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern classification.* John Wiley & Sons, 2012.

[Dia93] *The effect of intensive treatment of diabetes on the development and progression of long-term complications in insulin-dependent diabetes mellitus*, The Diabetes Control and Complications Trial Research Group Std. 14, 1993.

[DKZ16] N. Demitri, H. Krim, and A. M. Zoubir, "An Ordered Subspace Clustering Approach for Temporal Segmentation of Blood Glucose Measurements. under review," *IEEE Transactions on Biomedical Engineering*, 2016.

[DZ13] N. Demitri and A. Zoubir, "Mean-shift Based Algorithm for the Measurement of Blood Glucose in Hand-Held Devices," in *Proceedings of the 21st European Signal Processing Conference (EUSIPCO)*, September 2013.

[DZ14a] ——, "Detection of faulty glucose measurements using texture analysis," in *Signal Processing Conference (EUSIPCO), 2014 Proceedings of the 22nd European*, 2014, pp. 2480–2484.

[DZ14b] ——, "A Robust Kernel Density Estimator Based Mean-Shift Algorithm," in *IEEE 39th International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, May 2014.

[DZ15] ——, "Estimating Glucose Concentration Using A Sparse Scalable Mean-Shift Algorithm." in *IEEE 11th International Symposium on Biomedical Imaging (ISBI)*, April 2015.

[DZ16] ——, "Measuring Blood Glucose Concentrations in Photometric Glucometers Requiring Very Small Sample Volumes," *IEEE Transactions on Biomedical Engineering*, vol. PP, no. 99, pp. 1–1, 2016.

[EK88] L. Edelstein-Keshet, *Mathematical models in biology*. Siam, 1988, vol. 46.

[ESV12] E. Elhamifar, G. Sapiro, and R. Vidal, "See all by looking at a few: Sparse modeling for finding representative objects," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012, pp. 1600–1607.

[EV09] E. Elhamifar and R. Vidal, "Sparse subspace clustering," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2009, pp. 2790–2797.

[EV10] ——, "Clustering disjoint subspaces via sparse representation," in *Proceedings IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP)*, 2010, pp. 1926–1929.

[EV13] ——, "Sparse subspace clustering: Algorithm, theory, and applications," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 11, pp. 2765–2781, 2013.

[FH75] K. Fukunaga and L. Hostetler, "The estimation of the gradient of a density function, with applications in pattern recognition," *IEEE Transactions on Information Theory*, vol. 21, no. 1, pp. 32–40, 1975.

[FWO+11] O. Fuerst, D. Weintraub, D. Oppenheim, M. Plevinski, and E. Cohen, "Fluids testing apparatus and methods of use," May 8 2011, US Patent App. 13/697,022.

[GDG+03] B. Guerci, P. Drouin, V. Grang, P. Bougnres, P. Fontaine, V. Kerlan, P. Passa, C. Thivolet, B. Vialettes, and B. Charbonnel, "Self-monitoring of blood glucose significantly improves metabolic control in patients with type 2 diabetes mellitus: the Auto-Surveillance Intervention Active (ASIA) study," *Diabetes & Metabolism*, vol. 29, no. 6, pp. 587 – 594, 2003.

[GDM] "http://www.idf.org/gestational-diabetes, retrieved March 2016."

[GGL13] Y. Guo, J. Gao, and F. Li, "Spatial subspace clustering for hyperspectral data segmentation," in *The Third International Conference on Digital Information Processing and Communications (ICDIPC)*. The Society of Digital Information and Wireless Communication, 2013, pp. 180–190.

[GGL+15] Y. Guo, J. Gao, F. Li, S. Tierney, and M. Yin, "Low rank sequential subspace clustering," in *International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2015, pp. 1–8.

[GM13] V. A. Gault and N. H. McClenaghan, *Understanding bioanalytical chemistry: principles and applications*. John Wiley & Sons, 2013.

[GR11] S. Z. A. Gilani and N. I. Rao, "Data driven bandwidth for medoid shift algorithm," in *Computational Science and Its Applications-ICCSA 2011*. Springer, 2011, pp. 534–546.

[GSS93] N. Gordon, D. Salmond, and A. Smith, "Novel approach to nonlinear/non-Gaussian Bayesian State Estimation," *IEEE Proceedings For Radar and Signal Processing*, vol. 140, no. 2, pp. 107–113, April 1993.

[GW02] R. C. Gonzalez and R. E. Woods, "Digital image processing," 2002.

[Ham74] F. R. Hampel, "The influence curve and its role in robust estimation," *Journal of the American Statistical Association*, vol. 69, no. 346, pp. 383–393, 1974.

[Har79] R. M. Haralick, "Statistical and structural approaches to texture," *Proceedings of the IEEE*, vol. 67, no. 5, pp. 786–804, 1979.

[HRHW15] E. A. Holmes, S. Roy, J. Howard, and C. Wang, "Medical device for analyte monitoring and drug delivery," Aug. 3 2015, US Patent App. 14/816,426.

[HSD73] R. Haralick, K. Shanmugam, and I. Dinstein, "Textural Features for Image Classification," *IEEE Transactons on Systems, Man, And Cybernetics*, vol. SMC-3, no. 6, pp. 610–621, November 1973.

[Hub64] P. J. Huber, "Robust estimation of a location parameter," *The Annals of Mathematical Statistics*, vol. 35, no. 1, pp. 73–101, 1964.

[Hub11] ——, *Robust statistics*. Springer, 2011.

[HWMS04] W. Härdle, A. Werwatz, M. Müller, and S. Sperlich, *Introduction,Nonparametric and semiparametric models.* Springer, 2004.

[IDF14] "Annual Report," International Diabetes Foundation, Tech. Rep., 2014.

[Int13] *In vitro diagnostic test system - requirements for blood-glucose monitoring ssystem for self-testing in managing diabetes mellitus.*, International Organization for Standardization: ISO 15197 Std., 2013.

[JA13] J. Jordan and E. Angelopoulou, "Mean-Shift Clustering For Interactive Multispectral Image Analysis," *20th IEEE International Conference on Image Processing (ICIP)*, 2013.

[JAT02] M. J. T. J. A. Tamada, M. Lesho, "Keeping watch on glucose," *IEEE Spectrum*, vol. 39, no. 4, pp. 52–57, 2002.

[JCS+12] A. Janowczyk, S. Chandran, R. Singh, D. Sasaroli, G. Coukos, M. Feldman, and A. Madabhushi, "High-Throughput Biomarker Segmentation on Ovarian Cancer Tissue Microarrays via Hierarchical Normalized Cuts," *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 5, pp. 1240–1252, May 2012.

[JD00] S. A. Julious and C. A. M. Debarnot, "Why are Pharmacokinetic Data Summarized by Arithmetic Means ?" *Journal of Biopharmaceutical Statistics*, vol. 10:1, p. 55:72, 2000.

[JPD07] S. R. Joshi, R. M. Parikh, and A. Das, "Insulin-history, biochemistry, physiology and pharmacology," *Journal-Association of Physicians of India*, vol. 55, no. L, p. 19, 2007.

[JYS13] K.-J. L. J.-Y. Shin, H.-H. Nam, "Estimation of glucose concentration using adaptive calibration curve in different hematocrit levels," *Electronics Letters*, vol. 49, no. 9, pp. 584–585, 2013. [Online]. Available: http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=6528106

[KAK+91] S. Kayashima, T. Arai, M. Kikuchi, N. Sato, N. Nagata, O. Takatani, N. Ito, J. Kimura, and T. Kuriyama, "New noninvasive transcutaneous approach to blood glucose monitoring: successful glucose monitoring on human 75 g OGTT with novel sampling chamber," *IEEE Transactions on Biomedical Engineering*, vol. 38, no. 8, pp. 752–757, Aug 1991.

[Kay93] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory.* Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1993.

[KC10] J. S. Krouwer and G. S. Cembrowski, "A Review of Standards and Statistics Used to Describe Blood Glucose Monitor Performance," *Journal of Diabetes Science and Technology*, vol. 4, no. 1, pp. 75–83, January 2010.

[KS08] J. S. Kim and C. Scott, "Robust kernel density estimation," in *IEEE International Conference on Acoustics, Speech and Signal Processing, 2008.*, 2008, pp. 3381–3384.

[KS11] ——, "On the Robustness of Kernel Density M-Estimators," in *Proceedings of the 28th International Conference on Machine Learning (ICML)*, 2011, pp. 697–704.

[KS12] ——, "Robust kernel density estimation," *The Journal of Machine Learning Research*, vol. 13, no. 1, pp. 2529–2565, 2012.

[Lau15] D. C. Lau, "Diabetes Technology and Devices Transform the Lives of People with Diabetes," *Canadian journal of diabetes*, vol. 39, no. 3, pp. 174–175, 2015.

[LCM10] Z. Lin, M. Chen, and Y. Ma, "The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices," *arXiv preprint arXiv:1009.5055*, 2010.

[Lei09] S. Leier, "Methods of Signal Detection and Processing in Medical Engineering," Master's thesis, TU Darmstadt, 2009.

[LLL13] Z. Lin, R. Liu, and H. Li, "Linearized alternating direction method with parallel splitting and adaptive penalty for separable convex programs in machine learning," *Machine Learning*, vol. 99, no. 2, pp. 287–325, 2013.

[LLS11] Z. Lin, R. Liu, and Z. Su, "Linearized alternating direction method with adaptive penalty for low-rank representation," *Advances in Neural Information Processing Systems*, pp. 612–620, 2011.

[LLY10] G. Liu, Z. Lin, and Y. Yu, "Robust subspace segmentation by low-rank representation," in *Proceedings of the 27th International Conference on Machine Learning (ICML)*, 2010, pp. 663–670.

[LLY+13] G. Liu, Z. Lin, S. Yan, J. Sun, Y. Yu, and Y. Ma, "Robust recovery of subspace structures by low-rank representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 1, pp. 171–184, 2013.

[LMJM13] C. Lu, M. Mahmood, N. Jha, and M. Mandal, "Automated Segmentation of the Melanocytes in Skin Histopathological Images," *IEEE Journal of Biomedical and Health Informatics*, vol. 17, no. 2, pp. 284–296, March 2013.

[Lue97] D. G. Luenberger, *Optimization by vector space methods*. John Wiley & Sons, 1997.

[LW01] J. Liu and M. West, "Combined parameter and state estimation in simulation-based filtering," in *Sequential Monte Carlo methods in practice*. Springer, 2001, pp. 197–223.

[MG09] A. Mayer and H. Greenspan, "An Adaptive Mean-Shift Framework for MRI Brain Segmentation," *IEEE Transactions on Medical Imaging*, vol. 28, no. 8, pp. 1238–1250, 2009.

[MMY06] R. A. Maronna, R. D. Martin, and V. J. Yohai, *Robust statistics*. J. Wiley, 2006.

[Moh97] B. Mohar, *Some applications of Laplace eigenvalues of graphs*. Springer, 1997.

[Moo96] T. K. Moon, "The expectation-maximization algorithm," *IEEE Signal Processing Magazine*, vol. 13, no. 6, pp. 47–60, 1996.

[NHR12] Z. Noumir, P. Honeine, and C. Richard, "One-class machines based on the coherence criterion," in *Statistical Signal Processing Workshop (SSP), 2012 IEEE*. IEEE, 2012, pp. 600–603.

[NJW+02] A. Y. Ng, M. I. Jordan, Y. Weiss *et al.*, "On spectral clustering: Analysis and an algorithm," *Advances in neural information processing systems*, vol. 2, pp. 849–856, 2002.

[Ots75] N. Otsu, "A threshold selection method from gray-level histograms," *Automatica*, vol. 11, no. 285-296, pp. 23–27, 1975.

[P+10] L. Poretsky *et al.*, *Principles of diabetes mellitus*. Springer, 2010.

[Par62] E. Parzen, "On Estimation of a Probability Density Function and Mode," *The Annals of Mathematical Statistics*, vol. 33, no. 3, pp. 1065–1076, 1962.

[PM90] B. U. Park and J. S. Marron, "Comparison of data-driven bandwidth selectors," *Journal of the American Statistical Association*, vol. 85, no. 409, pp. 66–72, 1990.

[PW13] W. Petrich and F. Weidner, "Device for detecting an analyte in a bodily fluid," Jan. 18 2013, US Patent App. 13/745,371.

[RRF12] A. Rebel, M. A. Rice, and B. G. Fahy, "The accuracy of point-of-care glucose measurements," *Journal of diabetes science and technology*, vol. 6, no. 2, pp. 396–411, 2012.

[SAS09] L. Shapira, S. Avidan, and A. Shamir, "Mode-detection via median-shift," in *IEEE 12th International Conference on Computer Vision, 2009*. IEEE, 2009, pp. 1909–1916.

[SC08] I. Steinwart and A. Christmann, *Support vector machines*. Springer Science & Business Media, 2008.

[SC+12] M. Soltanolkotabi, E. J. Candes *et al.*, "A geometric analysis of subspace clustering with outliers," *The Annals of Statistics*, vol. 40, no. 4, pp. 2195–2238, 2012.

[SDZ16] A.-K. Seifert, N. Demitri, and A. Zoubir, "Decreasing the Measurement Time of Blood Sugar Tests using Particle Filtering." *Proceedings of 41st IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, March 2016.

[Sei15] A.-K. Seifert, "Particle Filter for Image-based Photometric Glucose Measurement in Hand-held Devices," Master's thesis, Technische Universität Darmstadt, May 2015.

[Sil86] B. Silverman, *Density Estimation For Statistics And Data Analysis*. Chapman and Hall, London, 1986.

[Sim02] H. Simon, *Adaptive filter theory*, 2002, vol. 2.

[SKK07] Y. A. Sheikh, E. A. Khan, and T. Kanade, "Mode-seeking by medoidshifts," in *IEEE 11th International Conference on Computer Vision (ICCV)*, 2007, pp. 1–8.

[SM00] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 8, pp. 888–905, 2000.

[SNL13] J.-Y. Shin, H.-H. Nam, and K.-J. Lee, "Estimation of glucose concentration using adaptive calibration curve in different hematocrit levels," *Electronics Letters*, vol. 49, no. 9, pp. 584–585, 2013.

[SS00] P. Sonksen and J. Sonksen, "Insulin: understanding its action in health and disease," *British journal of anaesthesia*, vol. 85, no. 1, pp. 69–79, 2000.

[SS02] B. Schölkopf and A. J. Smola, *Learning with kernels*. The MIT Press, 2002.

[SS04] P. Stoica and Y. Selen, "Model-order selection: a review of information criterion rules," *Signal Processing Magazine, IEEE*, vol. 21, no. 4, pp. 36–47, 2004.

[ST99] L.-K. Soh and C. Tsatsoulis, "Texture Analysis of SAR Sea Ice Imagery Using Gray Level Co-Occurrence Matrices," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 37, no. 2, pp. 780–795, 1999.

[SV10] F. Sahba and A. Venetsanopoulos, "Mean shift based algorithm for mammographic breast mass detection," in *Proceedings 17th IEEE International Image Processing Conference (ICIP)*, 2010, pp. 3629–3632.

[Tew15] M. Tewari, "A functional extracellular transcriptome in animals? implications for biology, disease and medicine," *Genome biology*, vol. 16, no. 1, p. 47, 2015.

[TGG14] S. Tierney, J. Gao, and Y. Guo, "Subspace clustering for sequential data," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014, pp. 1019–1026.

[TGG15] S. Tierney, Y. Guo, and J. Gao, "Segmentation of Subspaces in Sequential Data," *arXiv preprint arXiv:1504.04090*, 2015.

[TJ93] M. Tuceryan and A. Jain, *Handbook of Pattern Recognition and Computer Vision*. Singapore: World Scientific, 1993.

[TN09] K. Tonyushkina and J. H. Nichols, "Glucose meters: a review of technical challenges to obtaining accurate results," *Journal of Diabetes Science and Technology*, vol. 3, no. 4, pp. 971–980, 2009.

[Tro04] J. A. Tropp, "Greed is good: Algorithmic results for sparse approximation," *IEEE Transactions on Information Theory*, vol. 50, no. 10, pp. 2231–2242, 2004.

[TVS+15] A. Triantafyllidis, C. Velardo, D. Salvi, S. A. Shah, V. Koutkias, and L. Tarassenko, "A Survey of Mobile Phone Sensing, Self-reporting and Social Sharing for Pervasive Healthcare," *IEEE Journal of Biomedical and Health Informatics*, vol. PP, no. 99, pp. 1–1, 2015.

[Vid10] R. Vidal, "A tutorial on subspace clustering," *IEEE Signal Processing Magazine*, vol. 28, no. 2, pp. 52–68, 2010.

[VL07] U. Von Luxburg, "A tutorial on spectral clustering," *Statistics and computing*, vol. 17, no. 4, pp. 395–416, 2007.

[VMS05] R. Vidal, Y. Ma, and S. Sastry, "Generalized principal component analysis (GPCA)," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 12, pp. 1945–1959, 2005.

[VS08] A. Vedaldi and S. Soatto, "Quick shift and kernel methods for mode seeking," in *Computer vision–ECCV 2008*. Springer, 2008, pp. 705–718.

[VS13] R. Vandermeulen and C. Scott, "Consistency of robust kernel density estimators," in *Conference on Learning Theory*, 2013, pp. 568–591.

[V.V01] V.V.Vazirani, *Approximation algorithms*. Springer, 2001.

[WB95] G. Welch and G. Bishop, "An Introduction to the Kalman Filter," Chapel Hill, NC, USA, Tech. Rep., 1995.

[WDZ16] K. Wahby, N. Demitri, and A. M. Zoubir, "Sequential Low Rank Representation For Blood Glucose Measurements (under review). Budapest, Hungary. August 2016." in *In Proceedings 24th European Signal Processing Conference(EUSIPCO).*, August 2016.

[WHG+15] F. Wu, Y. Hu, J. Gao, Y. Sun, and B. Yin, "Ordered Subspace Clustering With Block-Diagonal Priors," *IEEE Transactions on Cybernetics*, vol. PP, no. 99, pp. 1–11, 2015.

[WHO] "http://www.who.int/diabetes/en/, retrieved October 2015."

[WHO16] "Global report on diabetes." World Health Organization 2016, Tech. Rep., 2016.

[WJ94] M. P. Wand and M. C. Jones, *Kernel smoothing*. CRC Press, 1994.

[WPS+15] L. Wang, P. C. Pedersen, D. M. Strong, B. Tulu, E. Agu, and R. Ignotz, "Smartphone-Based Wound Assessment System for Patients With Diabetes," *IEEE Transactions on Biomedical Engineering*, vol. 62, no. 2, pp. 477–488, Feb 2015.

[YMK98] M. Yamaguchi, M. Mitsumori, and Y. Kano, "Noninvasively measuring blood glucose using saliva," *IEEE Engineering in Medicine and Biology Magazine*, vol. 17, no. 3, pp. 59–63, May 1998.

[ZKCM12] A. M. Zoubir, V. Koivunen, Y. Chakhchoukh, and M. Muma, "Robust Estimation in Signal Processing: A Tutorial-Style Treatment of Fundamental Concepts," *IEEE Signal Processing Magazine*, vol. 29, no. 4, pp. 61–80, 2012.

Curriculum Vitae

Name:	Nevine Demitri
Date of birth:	06.08.1987
Place of Birth:	Cairo, Egypt
Family status:	single

Education

10/2009-04/2011	Technische Universität Darmstadt, Germany Electrical Engineering and Information Technology Master of Science
10/2005-09/2008	Technische Universität Darmstadt, Germany Information and Communication Technology Bachelor of Science
09/1997-05/2005	Deutsche Schule der Borromäerinnen, Cairo, Egypt High School Degree (Abitur)

Work Experience

since 05/2011	Research Associate at Signal Processing Group Technische Universität Darmstadt, Germany
04/2015-06/2015	Visiting Research Fellow North Carolina State University, Raleigh, USA
08/2010-04/2011	Master thesis Siemens Audiologische Technik, Erlangen, Germany
04/2010-08/2010	Internship Siemens Audiologische Technik, Erlangen, Germany
06/2008-07/2008	Internship Signal Processing Group Technische Universität Darmstadt, Germany
04/2007	Mentor Graphics Corporation, Cairo, Egypt

Erklärung laut §9 der Promotionsordnung

Ich versichere hiermit, dass ich die vorliegende Dissertation allein und nur unter Verwendung der angegebenen Literatur verfasst habe. Die Arbeit hat bisher noch nicht zu Prüfungszwecken gedient.

Darmstadt, 26. April 2016,