Advances in the theory and applications of (dis)similarity functions in pattern recognition

PhD thesis

György Kovács

Supervisor: Dr. Attila Fazekas

University of Debrecen
Doctoral School of Informatics
Debrecen, 2016.

Advances in the theory and applications of (dis)similarity functions in pattern recognition

PhD thesis

György Kovács

Supervisor: Dr. Attila Fazekas

University of Debrecen
Doctoral School of Informatics
Debrecen, 2016.

Ezen értekezést a Debreceni Egyetem Természettudományi Doktori Tanács Informatikai Tudományok Doktori Iskola *Diszkrét matematika, képfeldolgozás és komputergeometria* programja keretében készítettem a Debreceni Egyetem természettudományi doktori (PhD) fokozatának elnyerése céljából.

Debrecen, 2016, április 3.

.....
Kovács György

Tanusítom, hogy Kovács György doktorjelölt 2007–2010 között a fent megnevezett Doktori Iskola *Diszkrét matematika, képfeldolgozás és komputergeometria* programjának keretében irányításommal végezte munkáját. Az értekezésben foglalt eredményekhez a jelölt önálló alkotó tevékenységével meghatározóan hozzájárult. Az értekezés elfogadását javasolom.

Debrecen, 2016, április 3.

.....
Dr. Fazekas Attila

Advances in the theory and applications of (dis)similarity functions in pattern recognition

Értekezés a doktori (Ph.D.) fokozat megszerzése érdekében az informatika tudományágban

Írta: Kovács György okleveles programtervező matematikus, okleveles alkalmazott matematikus, okleveles fizikus

Készült a Debreceni Egyetem Informatikai Tudományok doktori iskolája (Diszkrét matematika, képfeldolgozás és komputergeometria programja) keretében

Témavezető: Dr. Fazekas Attila

A doktori szigorlati bizottság:

elnök:	Dr. Pethő Attila	.....
tagok:	Dr. Hajdu András	.....
	Dr. Palágyi Kálmán	.....

A doktori szigorlat időpontja: 2010. december 20.

Az értekezés bírálói:

Dr. Zichar Marianna	.....
Dr. Nyúl G. László	.....

A bírálóbizottság:

elnök:	Dr. Kruppa András	.....
tagok:	Dr. Vaszil György	.....
	Dr. Halász Gábor	.....
	Dr. Balázs Péter	.....
	Dr. Barkóczi Péter	.....

Az értekezés védésének időpontja: 2016. .....

Acknowledgements

I would like to express my endless gratitude to my mom, Éva Kovácsné Angalét, my dad, György Kovács, my brother Gabi, and all the other members of my family who always supported my studies; to my friends who accepted the objection many times: "I can't go, because I'm working on a paper"; to my colleagues, teachers, supervisors and managers, primarily to Szilvia Szeghalmy: we went through the same long road and it was great to have a company; to Dr. István Juhász, who encouraged me to spend time on science; to Dr. Katalin Nyakóné Juhász, who helped me to my first job; to Dr. Miklós Emri, who thought me programming in C++; to Dr. András Hajdu, who thought me how to write papers; to Dr. István Fazekas, who showed me how beautiful statistics is; to Dr. Zsolt Gulácsi, from whom I learned that even the most complicated formula can be beaten with a pen and a piece of paper; and finally to Dr. Attila Fazekas, my PhD supervisor, who thought me how to plan, accomplish and disseminate scientific work independently. I would like to express my thanks to all my co-authors, especially to Dr. Tamás Donkó, Dr. Sándor Attila Kis and Dr. Gábor Opposits, it was great to work with you!

Köszönetnyilvánítás

Ezúton szeretném kifejezni végtelen hálámat anyukámnak, Kovácsné Angalét Évának, apukámnak, Kovács Györgynek, kisöcsémnek Gabinak, valamint szűkebb és bővebb családom minden tagjának, akik mindig támogattak tanulmányaimban; barátaimnak, akik számtalanszor fogadták el kifogásként, hogy „nem tudok menni, épp egy cikken dolgozom”; korábbi kollégáimnak, tanáraimnak, témavezetőimnek és főnökeimnek, elsősorban Szeghalmy Szilviának, akivel hasonlóan hosszú utat jártunk végig: nagy segítség volt, hogy nem vagyok egyedül; Dr. Juhász Istvánnak, aki bíztatott, hogy a tanulmányaimon túl is foglalkozzak tudománnyal; Nyakóné Dr. Juhász Katalinnak, aki első igazi munkámhoz segített, Dr. Emri Miklósnak, akitől megtanultam C++-ban programozni; Dr. Hajdu Andrásnak, akitől megtanultam cikket írni; Dr. Fazekas Istvánnak, aki felhívta a figyelmemet a statisztika szépségére; Dr. Gulácsi Zsoltnak, akitől megtanultam, hogy a legbonyolultabb formuláktól sem kell megijedni; és természetesen Dr. Fazekas Attilának, PhD témavezetőmnek, aki mellett megtanultam, hogyan kell önállóan megtervezeni, kivitelezni és publikálni egy tudományos munkát. Köszönettel tartozom továbbá azoknak a szerzőtársaknak, akikkel valamilyen formában az elmult években együtt dolgoztam és publikáltam, közülük is szeretném kiemelni Dr. Donkó Tamást, Dr. Kis Sándor Attilát és Dr. Opposits Gábort, élmény volt veletek dolgozni!

Table of Contents

Chapter 1

Introduction

> "Don't judge each day by the harvest you reap, but by the seeds that you plant."
>
> Robert Louis Stevenson

In the recent years pattern recognition and its applications definitely entered into everyday life. People have smart devices recognizing spoken commands, applications are able to recognize friends on photos and even digital cameras can recognize smiles in real time to take the best pictures. Industry puts more and more efforts and resources to analyse unimaginable amounts of data searching for patterns and causalities to predict the future, like the behaviour of customers or the development of economic processes. One famous and revolutionary example is the Netflix Prize[1] from 2009, which was awarded to the team providing the best solution to predict user ratings for movies. Since then, many start-up companies have been founded with the flag of pattern recognition and machine intelligence, graduation programs have been started, online communities and platforms appeared[2] to share and solve problems similar to the Netflix Prize. Many fields of natural sciences became dependent on pattern recognition since the increasing amount of more and more complex and diverse data acquired by digital sensors can hardly be filtered, processed and evaluated without

[1] www.netflixprize.com

[2] `www.kaggle.com`

intelligent software solutions. A good example is the *Higgs Boson Machine Learning Challenge*. The famous discovery of the boson meant the beginning of a long quest to measure its characteristics as accurately as possible: a sufficiently large amount of Higgs related signals have to be extracted from the unbelievable amounts of data (1 petabyte per seconds) acquired by the ATLAS detector, and this is the point where pattern recognition was applied in particle physics: a competition[3] was started on Kaggle to find the best software solution which is able to classify numeric vectors to the classes *signal* and *background* accurately. As stated on the webpage of the competition: "*No knowledge of particle physics is required*". These simple examples illustrate the various areas where pattern recognition is applied successfully. But this is only the present. We can expect that the upcoming results, solutions and applications will change life, industry and science in a fundamental way.

Deep under the hood, most of the pattern recognition driven applications are built on the concept of measuring the *similarity* or the inversely proportional *dissimilarity* of real world objects described by numerical vectors. Correspondingly, similarity and dissimilarity are usually formulated in terms of mathematical functions operating on vectors. Obviously, what similar and dissimilar mean, depends on the problem and field of application. As a common property, (dis)similarity functions are usually desired to be invariant to some certain classes of transformations. For example, one can see the well-known image of *Lena* and its distorted variants in Figure 1.1: despite the various distortions like under- and over-exposure, noise, geometric and color space transformations, the human mind can recognize that the images contain the same face in the same scene. Thus, in face recognition applications – having similar performance as human mind – the (dis)similarity measure is desired to be invariant to all of these distortions, since they definitely conserve the piece of information important from the application's point of view: the face. In pattern recognition the requirements of the application define which transformations a (dis)similarity measure has to be invariant to, and the invariance implies that the information (pattern) is conserved under this class of transformations[4]. If the

[3] `https://www.kaggle.com/c/higgs-boson`

[4] As a slightly philosophical remark, one can recognize an analogy with one of the basic principles of theoretical physics. The systems described by physics may have so

coordinates of the vectors can be interpreted similarly (e.g. all pixels of a grayscale represent an intensity level; all elements of stock price time series are prices), the needs of most applications can be satisfied by (dis)similarity measures invariant to some general classes of transformations, like linear, non-linear, monotonic or non-monotonic ones. Consequently, the research of (dis)similarity measures invariant to some of these classes is of high value in the various fields and applications of pattern recognition.

1.1 Outline

In this work, original results of my recent research activity are presented. The results are organized into three main chapters, each of them closely related to (dis)similarity measures and their applications. Although (dis)similarity measures are the common divisor, the problems described and solved have different background. For the ease of discussion, the motivation and overview of related work are arranged into the first sections of the three main chapters. The main part of the dissertation is organized as follows:

1. In Chapter 2 two novel dissimilarity measures are derived, both of them are approximately invariant to even non-linear monotonically increasing transformations and can be computed in linear time. The proposed measures are tested and evaluated in various scenarios of template matching and compared to other techniques designed to be invariant to monotonic transformations. The results show that the measures are highly competitive, in certain cases giving remarkably better performance than conventional techniques. The measures are independent from the spatial dimensionality of the objects being

called *symmetries*. In this context a symmetry means that some transformation of the system does not change its description and behaviour. For example, if a system can be described by the very same laws and formulation in times t_0 and t_1, it is said to have time symmetry. Another systems may have translation symmetry or rotation symmetry, if the formulation of its description does not change by the spatial translation or rotation of the system. All of these symmetries induce conservation laws. Time symmetry implies that the energy of the system is conserved; the translation and rotational symmetry imply that the momentum and angular momentum of the system is preserved, respectively. Thus, invariances to some transformations imply that some quantitative properties of the system are conserved.

Figure 1.1: The *Lena* image (a) and some distorted variants: grayscale (b), under-exposed (c), over-exposed (d), Gaussian noise (e), salt-and-pepper noise (f), color transformation (g), mirroring (h)

compared, thus, besides template matching in images (which we use for demonstration and evaluation) both of the proposed measures can be used to compare and measure the dissimilarity of 1D time series or even 3D images.

2. In Chapter 3 a problem related to general pattern recognition techniques is discussed. The so called *kernel trick* (replacing the inner product to a bivariate function which acts like an inner product in a high dimensional space) is a well-known technique to adapt the general formulation of machine learning techniques to a specific problem or improve their properties in some sense, by keeping the computational demands similar to the inner product based variants. For example, Support Vector Machines (SVM) are usually used with non-linear kernels for classification – although they are formulated with the conventional inner product – and the kernel PCA is also a

well-known and theoretically established technique to extract information from noisy, distorted vectors. Similarly, the kernel trick can be applied to (dis)similarity measures if they can be formulated as the linear combinations of inner products (like the Euclidean distance and the correlation coefficient) to improve their discrimination power. We have examined the kernelization of the *Pearson correlation coefficient* (PCC) and *centered Euclidean distance* (cEUC) measures with the commonly used polynomial kernel and found that there are infinitely many different ways to compute them in the polynomial kernel space, and out of them at least one can be calculated in terms of kernel evaluations. The discrimination power of the kernelized PCC and cEUC measures is evaluated in classification problems using the k-nearest neighbour classifier. The results show that the performance of the kernelized measures is highly competitive, and in certain problems better than that of the non-kernelized variants.

3. In contrast with the mainly theoretical problems and results of the previous chapters, in Chapter 4 we go to the field of applications and propose a novel technique for the segmentation of the vasculature in retinal images. The method is based on template matching operators, and correspondingly, on the use of (dis)similarity measures. The template matching framework we developed for the accurate representation of the vasculature can be generalized to many other problems of pattern recognition. The proposed method was tested on the images of two publicly available datasets and compared to many previously published techniques in terms of the conventional measures of binary classification. We found that the proposed method outperforms the previous techniques on the *de facto* standard image databases, showing that the template matching based representation of the vasculature is a reasonable technique and the features extracted by the method have high discriminative power.

The discussion of the new results is followed by a summary in Chapter 5. Although some of the discussed problems are closely related to the field of computer vision, we emphasize that most of the ideas, methods and approaches presented in the three main chapters can be used in a wide variety of applications in pattern recognition and data science, none of

them is restricted to images.

1.2 Notations

Throughout the dissertation we use the following notations. Lowercase, boldface and uppercase letters are used to denote scalars, vectors and matrices, respectively (e.g. $x \in \mathbb{R}$, $d, b \in \mathbb{N}$, $\mathbf{t} \in \mathbb{R}^d$, $\boldsymbol{\alpha} \in \mathbb{R}^b$, $S \in \mathbb{R}^{d \times b}$). The coordinates of vectors and elements of matrices are referred by subscripts (e.g. $\mathbf{t}_i$ denoting the ith coordinate of vector $\mathbf{t}$ and S_{ij} denoting the element in the ith row and jth column of matrix S). Sets are denoted by calligraphic letters (like $\mathcal{I} \subset \{0, 1, \ldots, d\}$). We use the notation $\mathcal{U}([a, b])$ for the uniform probability distribution on the interval $[a, b]$, and $\mathcal{G}(m, \sigma)$ stands for the Gaussian distribution with mean m and standard deviation σ. Further notations used only in some specific parts of the document are introduced in the body of the text.

Chapter 2

Novel dissimilarity measures invariant to monotonic tone mappings

2.1 Motivation and Introduction

Many (dis)similarity functions have been proposed in the literature of pattern recognition, some of them being developed to satisfy the needs of particular applications, others being general enough to be applied in various problems efficiently. The general ones are usually invariant to certain

- geometric transformations – enabling the recognition of similar patterns when geometric distortions like rotation and scaling are present;
- photometric transformations – enabling the recognition of similar patterns in images taken under various lighting circumstances and by various acquisition techniques (like patterns in corresponding CT and PET images).

Throughout this chapter we focus on photometric invariance. Even if images of the same scene are taken, the changes in the lighting conditions and the usually differing optical transfer functions of various imaging devices can introduce non-linear deviations in the intensity distributions of the images. Thus, (dis)similarity functions invariant to certain classes

of non-linear intensity transformations are especially useful in computer vision applications.

As a novel contribution to the field, a dissimilarity function called Matching By Tone Mapping (MTM) was proposed recently [1]. The authors show that the MTM measures are invariant to non-linear intensity transformations; has extremely low computational demands (can be computed in terms of some convolution operations); can be considered as an extension of PCC for non-linear transformations; and outperform many previous measures in certain template matching and registration scenarios. Interestingly, the applications of MTM are somewhat limited by its most advantageous property: its invariance to non-linear intensity transformations. Particularly, the MTM dissimilarity of a vector and its additive inverse is zero, although, there are many problems where the distinction of such patterns is necessary. The issue is illustrated in Figure 2.1 by an example from the widely researched field of retinal image analysis. Figure 2.1(a) shows a fragment of a retinal image containing two vessels running parallel. There are many techniques proposed for the segmentation of the vasculature by template matching [2, 3, 4], using the correlation coefficient of the image content and an ideal vessel template as the primary feature of segmentation. Clearly, the deviation of real vessels from ideal vessel templates can rarely be described by linear transformations, thus, the performance of the segmentation techniques could be improved by replacing the correlation coefficient with the MTM measure. Accordingly, we have selected a real vessel segment template (white frame in Figure 2.1(a)), and applied template matching using MTM. The outcome is expected to have low scores at the center lines of vessel segments being parallel with the one in the template, since MTM is invariant to the possible non-linear deviations in the intensity profiles of vessels. However, as one can see in the outcome (Figure 2.1(b)), low scores appear between the vessels, as well, since the inverse of the template is similar to the thin stripe of background between the parallel vessels. In order to enable the use of MTM in problems where invariance to only monotonic intensity transformations is desired, we formulate the measure *Matching by Monotonic Tone Mapping* (MMTM) by keeping the outstanding properties and discriminative power of MTM, but restricting it to be invariant to monotonic intensity transformations only. For illustration, we anticipate the outcome of template matching

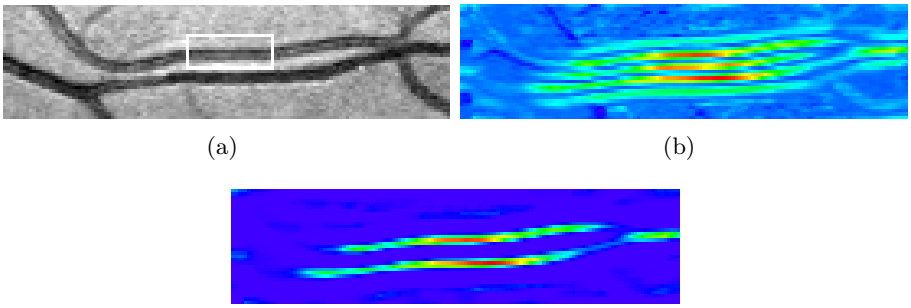


Figure 2.1: Illustration of a problem where MTM can not be used: a segment of a retinal fundus image and a template marked by white rectangle (a); the result of template matching by MTM (b); the result of template matching by the proposed MMTM measure (c)

using the proposed MMTM measure in Figure 2.1(c): one can qualitatively assess that the proposed technique gives the expected response.

The chapter is organized as follows. We give a brief overview of the most widely used techniques and measures in Section 2.2, in order to enable the qualitative comparison and positioning of the proposed methods in the world of (dis)similarity functions. Then, we give a short introduction to the Matching by Tone Mapping measure in Section 2.3 and derive the Matching by Monotonic Tone Mapping measure as a novel contribution to the field in Section 2.4. MMTM is evaluated and compared to other (dis)similarity measures in Section 2.5, some conclusions are drawn in Section 2.6 and a brief outlook is given in Section 2.7. The results of this chapter can be found in the paper [5].

2.2 Brief overview of similarity functions

Although the terms *similarity* and *dissimilarity* are widely used in pattern recognition and related fields, no single definition can be given: on the one hand the terms are highly intuitive, on the other hand, what similarity and dissimilarity mean highly depends on the problem. Given a set of vectors $\mathcal{X} \subset \mathbb{R}^d$, basically any bivariate function of the form $F : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ can

be considered as a similarity or dissimilarity function, as well. Particularly, one can find a problem and construct a representation $\mathcal{X}$, where F can be used as a similarity measure or a dissimilarity measure. For the sake of completeness, we give two general definitions based on the most common properties and use the terms in this sense in the rest of the dissertation.

Definition 2.2.1. *(Similarity measure) Let $\mathcal{X} \subset \mathbb{R}^d$, then, we call the function $S : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ a similarity measure, if for any $\mathbf{x} \in \mathcal{X}$,*

$$S(\mathbf{x}, \mathbf{x}) = \max_{\mathbf{y} \in \mathcal{X}} S(\mathbf{x}, \mathbf{y}). \tag{2.1}$$

Definition 2.2.2. *(Dissimilarity measure) Let $\mathcal{X} \subset \mathbb{R}^d$, then, we call the function $D : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ a disimilarity measure, if for any $\mathbf{x} \in \mathcal{X}$,*

$$D(\mathbf{x}, \mathbf{x}) = \min_{\mathbf{y} \in \mathcal{X}} S(\mathbf{x}, \mathbf{y}). \tag{2.2}$$

In words, what we expect from a similarity measure is that no vector can be more similar to $\mathbf{x}$ than $\mathbf{x}$. Analogously from a dissimilarity measure we expect that no vector can be less dissimilar from $\mathbf{x}$ than $\mathbf{x}$.

One can readily see that similarity and dissimilarity measures can be transformed into each other in many ways. For example, for any similarity measure S, a dissimilarity measure can be constructed in the form $D(\mathbf{x}, \mathbf{y}) = -S(\mathbf{x}, \mathbf{y})$, and vice versa. Due to this duality of similarity and dissimilarity functions, in the rest of the dissertation we neglect the composite term *(dis)similarity*. Depending on the topic we use either *similarity* or *dissimilarity*. Nevertheless, whenever we talk about the *higher* or *highest* score of a *similarity* function, the statements are equally true for *dissimilarity* functions by changing the adverbs to *lower* or *lowest*, respectively.

For the ease of discussion, we use the terminology of template matching in this chapter: let $\mathbf{t} \in \mathbb{R}^d$ denote a template compared to each window $\mathbf{w} \in \mathbb{R}^d$ of an image, seeking for the window of highest similarity score. By the term *monotonic* we mean *monotonically increasing* functions throughout the chapter.

2.2.1 The simplest methods

The simplest general purpose similarity functions are based on the inner product of the vectors. We can mention the inner product itself, but it is

rarely used in practice due to its unbounded nature:

$$S_{IP}(\mathbf{t}, \mathbf{w}) = \langle \mathbf{t}, \mathbf{w} \rangle. \tag{2.3}$$

Invariance to translations can be added to S_{IP} by subtracting the mean of the elements from the vectors. This measure is called *covariance*:

$$S_{COV}(\mathbf{t}, \mathbf{w}) = \langle \mathbf{t} - \bar{\mathbf{t}}\mathbf{1}, \mathbf{w} - \overline{\mathbf{w}}\mathbf{1} \rangle, \tag{2.4}$$

where $\bar{\mathbf{t}}$ denotes the mean of the elements in $\mathbf{t}$ and $\mathbf{1}$ denotes the vector consisting of all ones. Although S_{COV} is invariant to translations, that is,

$$S_{COV}(\mathbf{t}, \mathbf{t} + a\mathbf{1}) = Var(\mathbf{t}), \tag{2.5}$$

for any $a \in \mathbb{R}$, again, it is rarely used in practice due to its unboundedness. The correlation coefficient (CC)

$$S_{CC}(\mathbf{t}, \mathbf{w}) = \frac{\langle \mathbf{t}, \mathbf{w} \rangle}{\|\mathbf{t}\| \|\mathbf{w}\|}, \tag{2.6}$$

can be considered as another extension of the inner product measure. It can be readily seen that S_{CC} is bounded and invariant to scaling:

$$S_{CC}(\mathbf{t}, a\mathbf{t}) = 1, \tag{2.7}$$

for any $a \in \mathbb{R}$, $a \neq 0$. The last variant of IP based measures is the so called *Pearson correlation coefficient* (PCC) (also known as centered or mean-shifted correlation coefficient):

$$S_{PCC}(\mathbf{t}, \mathbf{w}) = \frac{\langle \mathbf{t} - \overline{\mathbf{t}\mathbf{1}}, \mathbf{w} - \overline{\mathbf{w}\mathbf{1}} \rangle}{\sqrt{\langle \mathbf{t} - \overline{\mathbf{t}\mathbf{1}}, \mathbf{t} - \overline{\mathbf{t}\mathbf{1}} \rangle \langle \mathbf{w} - \overline{\mathbf{w}\mathbf{1}}, \mathbf{w} - \overline{\mathbf{w}\mathbf{1}} \rangle}}, \tag{2.8}$$

which is well-known to be invariant to linear transformations (both scaling and translation). Particularly,

$$S_{PCC}(\mathbf{t}, a\mathbf{t} + b\mathbf{1}) = 1, \tag{2.9}$$

for any $a, b \in \mathbb{R}$, $b \neq 0$.

Probably the simplest dissimilarity functions are the L_p distances (L_1 and L_2 also known as *Manhattan* and *Euclidean distance (EUC)*, respectively):

$$D_{L_p}(\mathbf{t}, \mathbf{w}) = \left(\sum_{i=1}^{d} |\mathbf{t}_i - \mathbf{w}_i|^p \right)^{\frac{1}{p}}, \tag{2.10}$$

where $p \in \mathbb{R}$, $p \geq 1$. These measures reflect the natural concept of dissimilarity, without invariance to any transformations: it can be readily seen that there are no vectors $\mathbf{w} \neq \mathbf{t}$ providing $D_{L_p}(\mathbf{t}, \mathbf{w}) = 0$. One can observe that starting from the inner product measure the widely used PCC can be constructed in two steps: subtracting the mean (centering) to ensure invariance to translations and dividing by the norms (normalization) of the vectors to add invariance to scaling. The question arises: can we do the same steps and extend the L_p distances to be invariant to linear transformations? The answer is positive, and for the Euclidean distance some analogies can be observed with the inner product based similarity measures. We can add invariance to translation, scaling or both translation and scaling to the Euclidean distance and introduce the centered (cEUC), normalized (nEUC) and both centered and normalized (cnEUC) Euclidean distances:

$$D_{cEUC}(\mathbf{t}, \mathbf{w}) = \left(\sum_{i=1}^{d} \left(\mathbf{t}_i - \overline{\mathbf{t}} - \mathbf{w}_i + \overline{\mathbf{w}} \right)^2 \right)^{\frac{1}{2}}, \tag{2.11}$$

$$D_{nEUC}(\mathbf{t}, \mathbf{w}) = \left(\sum_{i=1}^{d} \left(\frac{\mathbf{t}_i}{\sigma_{\mathbf{t}}} - \frac{\mathbf{w}_i}{\sigma_{\mathbf{w}}} \right)^2 \right)^{\frac{1}{2}}, \tag{2.12}$$

$$D_{cnEUC}(\mathbf{t}, \mathbf{w}) = \left(\sum_{i=1}^{d} \left(\frac{\mathbf{t}_i - \overline{\mathbf{t}}}{\sigma_{\mathbf{t}}} - \frac{\mathbf{w}_i - \overline{\mathbf{w}}}{\sigma_{\mathbf{w}}} \right)^2 \right)^{\frac{1}{2}}, \tag{2.13}$$

where $\sigma_{\mathbf{t}}$ and $\sigma_{\mathbf{w}}$ denote the sample standard deviation of $\mathbf{t}$ and $\mathbf{w}$, that is,

$$\sigma_{\mathbf{t}} = \left(\frac{1}{d} \sum_{i=1}^{d} (\mathbf{t}_i - \overline{\mathbf{t}})^2 \right), \quad \sigma_{\mathbf{w}} = \left(\frac{1}{d} \sum_{i=1}^{d} (\mathbf{w}_i - \overline{\mathbf{w}})^2 \right). \tag{2.14}$$

The analogy between the IP based similarity measures and the variants of EUC comes from the fact that EUC can be rewritten in terms of inner

products:

$$D_{EUC}(\mathbf{t},\mathbf{w}) = \langle \mathbf{t},\mathbf{t}\rangle + \langle \mathbf{w},\mathbf{w}\rangle - 2\langle \mathbf{t},\mathbf{w}\rangle. \tag{2.15}$$

Exploiting the bilinearity of the inner product, one easily derives

$$D_{cEUC}(\mathbf{t},\mathbf{w}) = \left(\sigma_{\mathbf{t}}^2 + \sigma_{\mathbf{w}}^2 - 2S_{COV}(\mathbf{t},\mathbf{w})\right)^{\frac{1}{2}}, \tag{2.16}$$

$$D_{nEUC}(\mathbf{t},\mathbf{w}) = \sqrt{2}\left(1 - S_{CC}(\mathbf{t},\mathbf{w})\right)^{\frac{1}{2}}, \tag{2.17}$$

$$D_{cnEUC}(\mathbf{t},\mathbf{w}) = \sqrt{2}\left(1 - S_{PCC}(\mathbf{t},\mathbf{w})\right)^{\frac{1}{2}}. \tag{2.18}$$

It can be readily seen, that D_{nEUC} and D_{cnEUC} are shifted and scaled variants of S_{CC} and S_{PCC}, thus, we do not deal with them in the rest of the dissertation. On the other hand, although S_{COV} is part of D_{cEUC}, there is an important difference: D_{cEUC} has a lower bound at zero. Particularly, for $\mathbf{w} = \mathbf{t} + a\mathbf{1}$, $D_{cEUC}(\mathbf{t},\mathbf{w}) = 0$. Due its boundedness, D_{cEUC} can be used to infer on the dissimilarity: if D_{cEUC} is relatively small, the vectors are similar. Although these simple measures and their variants (like the *Tanimoto measure* [6]) are invariant to linear photometric transformations at most, they usually serve as building blocks of the more advanced techniques.

2.2.2 Advanced techniques

A large family of similarity measures (like *Spearman's Rho coefficient* [7], *Kendall's tau coefficient* [8, 9], the *greatest deviation* [10], and the *ordinal measure* [11]) is based on the rank transformation $R : \mathbb{R}^d \to \mathbb{Z}^d$ assigning the rank of $\mathbf{t}_i$ to $R(\mathbf{t})_i$. Most of these methods use some of the already mentioned simple techniques to measure the similarity of $R(\mathbf{t})$ and $R(\mathbf{w})$ instead of $\mathbf{t}$ and $\mathbf{w}$. Although these techniques are invariant to monotonic intensity transformation (since monotonic transformations do not change the rankings), a common drawback is that the outcome of R is highly sensitive to noise and ties among the elements of $\mathbf{t}$ and $\mathbf{w}$. Another class of commonly used measures is based on the *number of sign changes* [12, 13, 14] in the vector $(\mathbf{t} - \mathbf{w})$. These techniques provide the best performance if the intensities of $\mathbf{t}$ and $\mathbf{w}$ vary slowly and only small amount of zero mean noise is present. Several widely used similarity functions [15, 16, 17] are based on information theoretical approaches by measuring the *mutual information*

(MI) content in the intensity distribution of $\mathbf{t}$ and $\mathbf{w}$. Some variants of MI are based on the joint distribution of derived local quantities, like phase [18] or gradient orientation [19]. The MI based measures are intended to be invariant to even non-linear intensity transformations, although, the proper estimation of joint densities can be challenging, especially in the case of small templates. Following a different approach, *correlation ratio* and its variants [20, 21] measure the degree at which $\mathbf{w}$ is a single-valued function of $\mathbf{t}$. In registration problems these measures are shown to give comparable results as the widely used MI based techniques [22]. Invariance to certain photometric and geometric transformations can be achieved by extracting invariant features from $\mathbf{t}$ and $\mathbf{w}$ and measuring the similarity of the feature vectors. *Hu's descriptors* [23, 24] (combinations of statistical moments) and *local binary patterns* (LBP) [25, 26] (based on the intensity differences of a pixel and its neighbors) are two representative examples of these approaches. Both of these measures are invariant to geometrical rotations and LBP is also invariant to monotonic intensity transformations. The drawback of LBP is its high sensitivity to additive noise. There are many techniques developed to be invariant to even affine or projective geometrical transformations: instead of characterizing the entire template by a feature vector, these methods usually identify keypoints, describe their local environment and try to find keypoints with similar descriptors in the target image. Examples for these descriptors are the *histogram of gradients* (HOG), *scale-invariant feature transform* (SIFT) [27], *speeded up robust features* (SURF) [28], *gradient location and orientation histograms* (GLOH) [29], *local energy based shape histogram* (LESH) [30]. Obviously, these techniques cannot be used efficiently if the templates are small or do not contain characteristic keypoints.

Recent advances in the field include methods to reduce the computational demands of various similarity functions [31, 32] and the development of measures being invariant to specific geometric transformations [33, 34], noise [35] and non-linear intensity transformations [36]. Novel approaches based on higher order statistics [37] and evolutionary computations [38] are also proposed. Some excellent overviews of similarity functions can be found in the books [39], [40], [41], [42].

Our contribution to the field is the development of novel dissimilarity functions being approximately invariant to monotonic non-linear intensity

transformations. Before we give a detailed description of the proposed measures, we give a short introduction into the Matching by Tone Mapping technique, that inspired our work.

2.3 Brief Introduction to Matching by Tone Mapping

Following the notations of [1], a special class of functions called intensity transforms (or *tone mappings* in the terminology of [1]) are denoted by calligraphic letters, like $\mathcal{M} : \mathbb{R} \to \mathbb{R}$. With this formalism $\mathcal{M}(\mathbf{t})$ denotes the application of the function $\mathcal{M}$ to each coordinate of $\mathbf{t}$, independently. The MTM dissimilarity [1] of $\mathbf{t}$ and $\mathbf{w}$ is defined as

$$D(\mathbf{t}, \mathbf{w}) = \min_{\mathcal{M}} \left\{ \frac{\|\mathcal{M}(\mathbf{t}) - \mathbf{w}\|^2}{d\mathrm{var}(\mathbf{w})} \right\}. \tag{2.19}$$

In words, the numerator measures how close $\mathbf{t}$ can be transformed to $\mathbf{w}$ by applying some tone mapping $\mathcal{M}$ coordinate-wise and the denominator ensures the invariance to intensity scaling. It is worth noting that (2.19) is not symmetric: another measure can be defined by interchanging $\mathbf{t}$ and $\mathbf{w}$ in the minimization problem. In the rest of the chapter we focus on the form (2.19) referred as the *Pattern-to-Window* (PtW) case in [1], but emphasize that all results can be derived analogously for the *Window-to-Pattern* (WtP) case.

2.3.1 Piecewise constant approximation

Obviously, the minimization problem (2.19) cannot be solved explicitly, however, approximate solutions can be obtained by the linearization of the optimization problem, namely, replacing the term $\mathcal{M}(\mathbf{t})$ with a linear approximation. In order to introduce the linearization, we suppose that tone mappings are piecewise constant (PWC) functions and $\mathbf{t}$ is also replaced with a PWC approximation. Let the coordinates of $\mathbf{t}$ be quantized into $b \in \mathbb{N}$ bins and let the boundaries of the bins arranged into the vector $\mathbf{q} \in \mathbb{R}^{b+1}$, supposing that $\mathbf{q}_1 \leq \min_i \mathbf{t}_i$, $\max_i \mathbf{t}_i < \mathbf{q}_{b+1}$ and each bin $[\mathbf{q}_i, \mathbf{q}_{i+1}[$, $i = 1, \ldots, b$ contains at least one element. One can form the

PWC slice transform matrix $S(\mathbf{t}) \in \{0,1\}^{d\times b}$ of $\mathbf{t}$ as

$$S(\mathbf{t})_{ij} = \begin{cases} 1, & \text{if } \mathbf{q}_j \le \mathbf{t}_i < \mathbf{q}_{j+1}, \\ 0, & \text{otherwise.} \end{cases} \tag{2.20}$$

It can be readily seen that the matrix $S(\mathbf{t})$ contains structural information about $\mathbf{t}$, each column is related to a bin, and the ith element of column j is set to 1 only if $\mathbf{t}_i$ falls in the bin $[\mathbf{q}_j, \mathbf{q}_{j+1}[$. The columns of the matrix $S(\mathbf{t})$ are also referred as *slices*. Given $S(\mathbf{t})$, one can approximate $\mathbf{t}$ in the form $\mathbf{t} \simeq S(\mathbf{t})\boldsymbol{\beta}$, using $\boldsymbol{\beta} \in \mathbb{R}^b$,

$$\boldsymbol{\beta}_j = \frac{\mathbf{q}_j + \mathbf{q}_{j+1}}{2} \tag{2.21}$$

or $\boldsymbol{\beta}_j = \mathbf{q}_j$.

Similarly to the approximation of $\mathbf{t}$, the matrix $S(\mathbf{t})$ can be used to approximate various coordinate-wise transformations of $\mathbf{t}$, for example, the vector $\mathbf{u} = S(\mathbf{t})\boldsymbol{\beta}$, $\boldsymbol{\beta}_j = \mathbf{q}_j^2$ can be considered as an approximation of the vector $\mathcal{M}(\mathbf{t}) \in \mathbb{R}^d$ derived from $\mathbf{t}$ by applying the tone mapping $\mathcal{M}(x) = x^2$ coordinate-wise. Analogously, for any $\boldsymbol{\beta} \in \mathbb{R}^b$, the expression $S(\mathbf{t})\boldsymbol{\beta}$ can be considered as the PWC approximation of some possibly non-linear coordinate-wise transformation of $\mathbf{t}$. Obviously, the quality of approximation highly depends on the intensity distribution of $\mathbf{t}$, the number of bins, and the smoothness of $\mathcal{M}$. Nevertheless, the linearization of the minimization problem (2.19) by $\mathcal{M}(\mathbf{t}) \simeq S(\mathbf{t})\boldsymbol{\beta}$ is reasonable, and the PWC MTM becomes

$$D_{MTM}^{PWC}(\mathbf{t}, \mathbf{w}) = \min_{\boldsymbol{\beta}\in\mathbb{R}^b} \left\{ \frac{\|S(\mathbf{t})\boldsymbol{\beta} - \mathbf{w}\|^2}{d\text{var}(\mathbf{w})} \right\} = \frac{\|S(\mathbf{t})\hat{\boldsymbol{\beta}} - \mathbf{w}\|^2}{d\text{var}(\mathbf{w})}, \tag{2.22}$$

where $\hat{\boldsymbol{\beta}}$ denotes the exact optimal solution of the linear minimization problem in the numerator:

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}\in\mathbb{R}^b} \|S(\mathbf{t})\boldsymbol{\beta} - \mathbf{w}\|^2 = (S^T(\mathbf{t})S(\mathbf{t}))^{-1}S^T(\mathbf{t})\mathbf{w}, \tag{2.23}$$

Clearly, the PWC approximation ignores the fine details within the slices, since each element within a slice is approximated by the same scalar. If the fine details within the slices matter, better results can be expected if the PWC approximation is replaced by piecewise linear (PWL) approximation.

2.3.2 Piecewise linear approximation

In the PWL approximation the PWC slice transform matrix $S(\mathbf{t})$ of $\mathbf{t}$ is replaced by the *PWL slice transform* matrix $Q(\mathbf{t}) \in [0,1]^{d\times(b+1)}$,

$$Q(\mathbf{t})_{ij} = \begin{cases} 1 - r(\mathbf{t}_i, \mathbf{q}_j, \mathbf{q}_{j+1}), & \text{if } \mathbf{t}_i \in [\mathbf{q}_j, \mathbf{q}_{j+1}[, \\ r(\mathbf{t}_i, \mathbf{q}_j, \mathbf{q}_{j+1}), & \text{if } \mathbf{t}_i \in [\mathbf{q}_{j-1}, \mathbf{q}_j[, \\ 0, & \text{otherwise,} \end{cases} \tag{2.24}$$

where

$$r(x, a_1, a_2) = \frac{x - a_1}{a_2 - a_1}. \tag{2.25}$$

The elements of $Q(\mathbf{t})$ can be interpreted in the following way: row i contains the coefficients of the linear interpolation scheme reconstructing $\mathbf{t}_i$ from the bin boundaries $\mathbf{q}$. Particularly, $\mathbf{t}_i = \sum_{j=1}^{b+1} Q_{ij}(\mathbf{t})\mathbf{q}_j$, and the vector $\mathbf{t}$ can be perfectly reconstructed by the product $\mathbf{t} = Q(\mathbf{t})\mathbf{q}$. Similarly to the case of the PWC approximation, replacing $\mathbf{q}$ with an arbitrary vector $\boldsymbol{\zeta} \in \mathbb{R}^{b+1}$, the product $Q(\mathbf{t})\boldsymbol{\zeta}$ can be considered as the approximation of a vector $\mathcal{M}(\mathbf{t})$, where $\mathcal{M}$ denotes an unknown tone mapping. Thus, one can replace the term $\mathcal{M}(\mathbf{t})$ by $Q(\mathbf{t})\boldsymbol{\zeta}$ in (2.19), and the PWL MTM becomes

$$D_{MTM}^{PWL}(\mathbf{t}, \mathbf{w}) = \min_{\boldsymbol{\zeta}\in\mathbb{R}^{b+1}} \left\{ \frac{\|Q(\mathbf{t})\boldsymbol{\zeta} - \mathbf{w}\|^2}{d\mathrm{var}(\mathbf{w})} \right\} = \frac{\|Q(\mathbf{t})\hat{\boldsymbol{\zeta}} - \mathbf{w}\|^2}{d\mathrm{var}(\mathbf{w})}, \tag{2.26}$$

where $\hat{\boldsymbol{\zeta}} = (Q^T(\mathbf{t})Q(\mathbf{t}))^{-1}Q^T(\mathbf{t})\mathbf{w}$ denotes the exact optimal solution of the linear minimization problem in (2.26).

The approximate invariance of the MTM measures to non-linear tone mappings is a consequence of their design: if $\mathcal{M}$ is relatively smooth, the intensities falling in the same bin are transformed similarly, and the structure of $\mathcal{M}(\mathbf{t})$ remains similar to that of $\mathbf{t}$. Accordingly, both $D_{MTM}^{PWC}(\mathbf{t}, \mathcal{M}(\mathbf{t}))$ and $D_{MTM}^{PWL}(\mathbf{t}, \mathcal{M}(\mathbf{t}))$ become small and indicate high similarity.

2.4 Matching by Monotonic Tone Mapping

One can readily see that $D_{MTM}^{PWC}(\mathbf{t}, -\mathbf{t})$ is close to zero (proportional to the error of the PWC approximation), and $D_{MTM}^{PWL}(\mathbf{t}, -\mathbf{t}) = 0$. As we have

demonstrated in the introduction, there are problems where invariance to non-linear transformations is too permissive, and invariance to only monotonic non-linear intensity transformations would be desired. The main goal of this work is to derive dissimilarity measures giving the same results as $D^{PWC}_{MTM}(\mathbf{t}, \mathcal{M}(\mathbf{t}))$ and $D^{PWL}_{MTM}(\mathbf{t}, \mathcal{M}(\mathbf{t}))$ if $\mathcal{M}$ is a monotonic tone mapping, and smoothly deviate from the results of the MTM measures as $\mathcal{M}$ deviates from being monotonic.

2.4.1 Piecewise constant approximation

In this subsection we introduce the proposed Matching by Monotonic Tone Mapping (MMTM) measure with piecewise constant approximation. For the ease of notations we leave the argument of the PWC slice transform matrices $S(\mathbf{t})$, however, we emphasize that S is always a function of the template $\mathbf{t}$ and the vector of bin boundaries $\mathbf{q}$. Based on the PWC approximation of MTM (2.22), the desired dissimilarity function can be formulated easily:

$$\Delta^{PWC}_{MMTM}(\mathbf{t}, \mathbf{w}) = \min_{\boldsymbol{\gamma} \in \mathbb{R}^b} \left\{ \frac{\|S\boldsymbol{\gamma} - \mathbf{w}\|^2}{d\mathrm{var}(\mathbf{w})} \right\}, \\ \text{subject to } \boldsymbol{\gamma}_1 \leq \boldsymbol{\gamma}_2 \leq \cdots \leq \boldsymbol{\gamma}_b. \tag{2.27}$$

Two questions arise: can we solve this constrained optimization problem exactly? If so, can we solve it efficiently? From the general point of view, this problem is a quadratic optimization problem with linear constraints (QPLC) [43]. If the the quadratic form appearing in the target function is positive definite (i.e. the optimization problem is convex), exact solutions can be found by standard techniques (like the ellipsoid method) in polynomial time [44]. The numerator of the target function can be expanded by utilizing the bilinearity of the inner product: $\|S\boldsymbol{\gamma} - \mathbf{w}\|^2 = \langle S\boldsymbol{\gamma}, S\boldsymbol{\gamma}\rangle + 2\langle S\boldsymbol{\gamma}, \mathbf{w}\rangle + \langle \mathbf{w}, \mathbf{w}\rangle$, and the quadratic term can be rewritten as $\langle S\boldsymbol{\gamma}, S\boldsymbol{\gamma}\rangle = \boldsymbol{\gamma}^T A \boldsymbol{\gamma}$, where $A = S^T S$. Since the columns of S are orthogonal and $S \in \{0,1\}^{d \times b}$, A is diagonal and the element A_{ii} equals the cardinality of slice i. Thus, A has only positive eigenvalues, that is, A is positive definite, and polynomial time techniques can find the exact solution of (2.27). However, the execution of a polynomial time algorithm for each pixel of an image would make MMTM inefficient in practice. In the next subsection we show how the optimization problem (2.27) can be solved efficiently.

2.4.2 Transformation of PWC MMTM to an isotonic regression problem

The main result of this subsection shows that (2.27) can be transformed into a weighted isotonic regression problem and solved in linear time.

Theorem 2.4.1. *For any* $\mathbf{t}, \mathbf{w} \in \mathbb{R}^d$, $\Delta^{PWC}_{MMTM}(\mathbf{t}, \mathbf{w})$ *is equivalent to*

$$D^{PWC}_{MMTM}(\mathbf{t}, \mathbf{w}) = \frac{\|S\hat{\boldsymbol{\gamma}} - \mathbf{w}\|^2}{dvar(\mathbf{w})},$$

$$\hat{\boldsymbol{\gamma}} = \arg\min_{\boldsymbol{\gamma} \in \mathbb{R}^b} \sum_{i=1}^{b} |\mathcal{I}_i| (\boldsymbol{\gamma}_i - \hat{\boldsymbol{\beta}}_i)^2,$$

$$\text{subject to } \boldsymbol{\gamma}_1 \leq \boldsymbol{\gamma}_2 \leq \cdots \leq \boldsymbol{\gamma}_b, \tag{2.28}$$

where $\hat{\boldsymbol{\beta}}$ *is defined by (2.23) and* $\mathcal{I}_i \subseteq \{1, \ldots, d\}$ *denotes the set of indices corresponding to slice* i*, that is,* $j \in \mathcal{I}_i$ *if* $\mathbf{t}_j \in [\mathbf{q}_i, \mathbf{q}_{i+1}[$.

Proof. First we note that due to the special structure of S, the elements of $\hat{\boldsymbol{\beta}} = (S^T S)^{-1} S^T \mathbf{w}$ can be written in the form

$$\hat{\boldsymbol{\beta}}_i = \frac{1}{|\mathcal{I}_i|} \sum_{j \in \mathcal{I}_i} \mathbf{w}_j. \tag{2.29}$$

We have to show that the optimal solution vector $\hat{\boldsymbol{\gamma}}$ of (2.28) also minimizes the optimization problem in (2.27) and vice versa. We use indirect reasoning. Suppose, $\hat{\boldsymbol{\gamma}}$ minimizes the optimization problem in (2.28), but $\hat{\boldsymbol{\gamma}}$ does not minimize the numerator of (2.27). Then, there must be a vector $\boldsymbol{\delta} \in \mathbb{R}^b$, satisfying $\boldsymbol{\delta}_1 \leq \boldsymbol{\delta}_2 \leq ... \leq \boldsymbol{\delta}_b$, and

$$\|S(\mathbf{t})\boldsymbol{\delta} - \mathbf{w}\|^2 < \|S(\mathbf{t})\hat{\boldsymbol{\gamma}} - \mathbf{w}\|^2. \tag{2.30}$$

Expanding both sides of the inequality, one gets

$$\sum_{i=1}^{b} \sum_{j \in \mathcal{I}_i} \left(\boldsymbol{\delta}_i^2 - 2\boldsymbol{\delta}_i \mathbf{w}_j + \mathbf{w}_j^2 \right) < \sum_{i=1}^{b} \sum_{j \in \mathcal{I}_i} \left(\hat{\boldsymbol{\gamma}}_i^2 - 2\hat{\boldsymbol{\gamma}}_i \mathbf{w}_j + \mathbf{w}_j^2 \right). \tag{2.31}$$

Subtracting $\sum_{i=1}^{b} \sum_{j \in \mathcal{I}_i} \mathbf{w}_j^2$ from both sides and utilizing (2.29),

$$\sum_{i=1}^{b} |\mathcal{I}_i| \left(\boldsymbol{\delta}_i^2 - 2\boldsymbol{\delta}_i \hat{\boldsymbol{\beta}}_i \right) < \sum_{i=1}^{b} |\mathcal{I}_i| \left(\hat{\boldsymbol{\gamma}}_i^2 - 2\hat{\boldsymbol{\gamma}}_i \hat{\boldsymbol{\beta}}_i \right). \tag{2.32}$$

Adding $\sum_{i=1}^{b} |\mathcal{I}_i| \hat{\boldsymbol{\beta}}_i^2$ to both sides,

$$\sum_{i=1}^{b} |\mathcal{I}_i| \left(\boldsymbol{\delta}_i^2 - 2\boldsymbol{\delta}_i \hat{\boldsymbol{\beta}}_i + \hat{\boldsymbol{\beta}}_i^2 \right) < \sum_{i=1}^{b} |\mathcal{I}_i| \left(\hat{\boldsymbol{\gamma}}_i^2 - 2\hat{\boldsymbol{\gamma}}_i \hat{\boldsymbol{\beta}}_i + \hat{\boldsymbol{\beta}}_i^2 \right), \tag{2.33}$$

implying

$$\sum_{i=1}^{b} |\mathcal{I}_i| (\boldsymbol{\delta}_i - \hat{\boldsymbol{\beta}}_i)^2 < \sum_{i=1}^{b} |\mathcal{I}_i| (\hat{\boldsymbol{\gamma}}_i - \hat{\boldsymbol{\beta}}_i)^2, \tag{2.34}$$

that is, $\hat{\boldsymbol{\gamma}}_i$ is not the optimal solution of (2.28): contradiction. The other direction of the statement can be proved by applying the same steps of derivation backwards. □

One can observe, that the optimization problem in (2.28) is a *weighted isotonic regression problem in L_2 norm*, thus, exact solution can be found in $O(b)$ linear time by the well-known and many times rediscovered *Pool Adjacent Violators Algorithm* (PAVA) [45, 46], given in Algorithm 1. The input of PAVA is the vector $\hat{\boldsymbol{\beta}} \in \mathbb{R}^b$, and the cardinalities $|\mathcal{I}_i|$, $i = 1, \ldots, b$ playing the role of the weights in the weighted isotonic regression problem. The output of PAVA is proved [45, 46] to be the optimal solution $\hat{\boldsymbol{\gamma}}$ of the minimization problem in (2.28).

As a summary, we define the proposed PWC MMTM dissimilarity as formulated in (2.28), where $\hat{\boldsymbol{\beta}}$ is defined by (2.23), and $\hat{\boldsymbol{\gamma}}$ is determined by PAVA as the exact solution of the minimization problem in (2.28).

According to [1] the time complexity of computing PWC MTM for all windows of an image containing p pixels is $O(dp+bp)$, where d and b denote the size of the template and the number of bins used to construct the slice transform matrix, respectively. From (2.22) and (2.28) one can readily see that the only additional operation required to turn PWC MTM into PWC MMTM is PAVA. Since PAVA runs in $O(b)$ time, the time complexity of PWC MMTM remains linear $O(dp + 2bp)$.

2.4.3 Properties of the PWC MMTM

In this subsection some properties of PWC MMTM are discussed from the theoretical point of view. So far, we have supposed that the vector of bin boundaries $\mathbf{q}$ is fixed for a given template $\mathbf{t}$. In this subsection we

Algorithm 1 Pool Adjacent Violators Algorithm

1. Initialize the partitioning of coordinates $1, \dots, b$ into the sets $\mathcal{P}_i = \{i\}$, $i = 1, \dots, b$.
2. PAVA iteratively seeks for adjacent sets violating the constraints $\mu(\mathcal{P}_i) \leq \mu(\mathcal{P}_{i+1})$, where $\mu(\mathcal{P}_i)$ denotes the weighted mean corresponding to the indices in $\mathcal{P}_i$:
$$\mu(\mathcal{P}_i) = \frac{\sum\limits_{j \in \mathcal{P}_i} |\mathcal{I}_j| \hat{\boldsymbol{\beta}}_j}{\sum\limits_{j \in \mathcal{P}_i} |\mathcal{I}_j|}. \tag{2.35}$$
3. Suppose $\mathcal{P}_i$ and $\mathcal{P}_{i+1}$ violate the constraints, i.e. $\mu(\mathcal{P}_i) > \mu(\mathcal{P}_{i+1})$. The violation is resolved by unifying the sets: updating $\mathcal{P}_i$ to $\mathcal{P}_i \cup \mathcal{P}_{i+1}$, and removing $\mathcal{P}_{i+1}$ from the partitioning.
4. When no more adjacent pairs violating the constraints remain, the partitioning consists of $m \leq b$ sets and $\mu(\mathcal{P}_i) \leq \mu(\mathcal{P}_{i+1})$, $i = 1, \dots, m-1$ holds. The optimal solution vector $\hat{\boldsymbol{\gamma}} \in \mathbb{R}^b$ is constructed by the rule
$$\begin{aligned} \hat{\boldsymbol{\gamma}}_1 = \hat{\boldsymbol{\gamma}}_2 = \cdots = \hat{\boldsymbol{\gamma}}_{|\mathcal{P}_1|} &= \mu(\mathcal{P}_1) \\ \hat{\boldsymbol{\gamma}}_{|\mathcal{P}_1|+1} = \hat{\boldsymbol{\gamma}}_{|\mathcal{P}_1|+2} = \cdots = \hat{\boldsymbol{\gamma}}_{|\mathcal{P}_1|+|\mathcal{P}_2|} &= \mu(\mathcal{P}_2) \\ &\cdots \\ \hat{\boldsymbol{\gamma}}_{b-|\mathcal{P}_m|+1} = \hat{\boldsymbol{\gamma}}_{b-|\mathcal{P}_m|+2} = \cdots = \hat{\boldsymbol{\gamma}}_b &= \mu(\mathcal{P}_m). \end{aligned} \tag{2.36}$$

deal with various bin configurations, therefore, we introduce the notations $D^{PWC}_{MTM}(\mathbf{t}, \mathbf{w}|\mathbf{q})$, $S(\mathbf{q})$ and $\mathcal{I}_i(\mathbf{q})$ to emphasize that these quantities are functions of the bin boundaries. We also introduce $\mathcal{B}(\mathbf{t}, b) \subset \mathbb{R}^{b+1}$ denoting the set of all vectors specifying b bins for $\mathbf{t}$. Particularly, $\mathcal{B}(\mathbf{t}, b) = \{\mathbf{q} \in \mathbb{R}^{b+1} | \mathbf{q}_1 \leq \min_i \mathbf{t}_i, \max_i \mathbf{t}_i < \mathbf{q}_{b+1}$ and each bin $[\mathbf{q}_i, \mathbf{q}_{i+1}[$ contains at least one element of $\mathbf{t}\}$. First, a general property of PWC MTM is described: the finer the quantization is, the smaller the dissimilarity of $\mathbf{t}$ and $\mathbf{w}$ becomes.

Proposition 2.4.1. *Let* $\mathbf{t}, \mathbf{w} \in \mathbb{R}^d$, $\mathbf{q} \in \mathcal{B}(\mathbf{t}, b)$, $\mathbf{p} \in \mathcal{B}(\mathbf{t}, b-1)$, $\mathbf{p} = (\mathbf{q}_1, \dots, \mathbf{q}_{i-1}, \mathbf{q}_{i+1}, \dots, \mathbf{q}_{b+1})$. *Then,*
$$D^{PWC}_{MTM}(\mathbf{t}, \mathbf{w}|\mathbf{q}) < D^{PWC}_{MTM}(\mathbf{t}, \mathbf{w}|\mathbf{p}). \tag{2.37}$$

Proof. Introducing $\mathbf{s}_i(\mathbf{q}) = \sum_{j \in \mathcal{I}_i(\mathbf{q})} \mathbf{w}_j$, PWC MTM becomes

$$D^{PWC}_{MTM}(\mathbf{t}, \mathbf{w}|\mathbf{q}) = \frac{\sum_{i=1}^{b} \sum_{j \in \mathcal{I}_i} \left(\frac{\mathbf{s}_i(\mathbf{q})}{|\mathcal{I}_i(\mathbf{q})|} - \mathbf{w}_j \right)^2}{d\mathrm{var}(\mathbf{w})}. \tag{2.38}$$

Expanding and simplifying both sides of (2.37), one gets

$$\frac{\mathbf{s}_i^2(\mathbf{q})}{|\mathcal{I}_i(\mathbf{q})|} + \frac{\mathbf{s}_{i+1}^2(\mathbf{q})}{|\mathcal{I}_{i+1}(\mathbf{q})|} > \frac{(\mathbf{s}_i(\mathbf{q}) + \mathbf{s}_{i+1}(\mathbf{q}))^2}{|\mathcal{I}_i(\mathbf{q}) \cup \mathcal{I}_{i+1}(\mathbf{q})|} \tag{2.39}$$

This expression has the form

$$\frac{r_1^2}{z_1} + \frac{r_2^2}{z_2} > \frac{(r_1 + r_2)^2}{z_1 + z_2}, \tag{2.40}$$

with $r_1, r_2 \in \mathbb{R}$, $z_1, z_2 \in \mathbb{Z}^+$. Introducing $x = \frac{z_2}{z_1}$ and simplifying the expression, one gets the second order inequality

$$r_1^2 x^2 - 2 r_1 r_2 x + r_2^2 > 0. \tag{2.41}$$

The only solution of the second order equation on the left hand side is $x = \frac{r_2}{r_1}$. Thus, the strict inequality holds for any configuration of r_1, r_2, z_1 and z_2, except when $\frac{z_2}{z_1} = \frac{r_2}{r_1}$. However, the latter case implies $\frac{\mathbf{s}_i(\mathbf{q})}{|\mathcal{I}_i(\mathbf{q})|} = \frac{\mathbf{s}_{i+1}(\mathbf{q})}{|\mathcal{I}_{i+1}(\mathbf{q})|}$, which can never hold, since the mean of elements falling in different bins can not be the same. Consequently, the strict inequality of the statement is valid. □

The following proposition provides an interesting insight into the operation of PWC MMTM: using appropriately chosen bin boundaries, PWC MTM can reproduce the outcome of PWC MMTM. In other words, the effect of the monotonicity constraints can be considered as a rule for the automated adjustment of bin boundaries in PWC MTM.

Proposition 2.4.2. *Let* $\mathbf{t}, \mathbf{w} \in \mathbb{R}^d$, $\mathbf{q} \in \mathcal{B}(\mathbf{t}, b)$ *and suppose PAVA resolves* v *violations in* $D^{PWC}_{MMTM}(\mathbf{t}, \mathbf{w}|\mathbf{q})$. *There exists* $\mathbf{p} \in \mathcal{B}(\mathbf{t}, b - v)$, $\{\mathbf{p}_i\} \subseteq \{\mathbf{q}_i\}$ *such, that*

$$D^{PWC}_{MTM}(\mathbf{t}, \mathbf{w}|\mathbf{p}) = D^{PWC}_{MMTM}(\mathbf{t}, \mathbf{w}|\mathbf{q}). \tag{2.42}$$

Proof. Suppose $D^{PWC}_{MMTM}(\mathbf{t}, \mathbf{w}|\mathbf{q})$ is computed. Supposing that PAVA resolved v violations, a partitioning of indices $\{1, \dots, b\}$ into the sets $\mathcal{P}_1, \mathcal{P}_2, \dots, \mathcal{P}_m$ is determined, where $m = b - v$. Since the sets $\mathcal{P}_j \subseteq \{1, \dots, b\}$, $j = 1, \dots, m$ contain consecutive indices, and $i \in \{1, \dots, b\}$ corresponds to the bin $[\mathbf{q}_i, \mathbf{q}_{i+1}[$, the partitioning implies a rebinning. Let $\mathbf{p} \in \mathbb{R}^{m+1}$, $\mathbf{p}_i = \mathbf{q}_{z(i)}$, where $z(i) = 1 + \sum_{j=1}^{i-1} |\mathcal{P}_j|$. Then, $\mathbf{p} \in \mathcal{B}(\mathbf{t}, b - v)$ and $\{\mathbf{p}_i\} \subseteq \{\mathbf{q}_i\}$ holds. In order to prove that $\mathbf{p}$ satisfies $D^{PWC}_{MMTM}(\mathbf{t}, \mathbf{w}|\mathbf{q}) = D^{PWC}_{MTM}(\mathbf{t}, \mathbf{w}|\mathbf{p})$, it is sufficient to show that the vectors $S(\mathbf{q})\hat{\boldsymbol{\gamma}}(\mathbf{q})$ and $S(\mathbf{p})\hat{\boldsymbol{\beta}}(\mathbf{p})$ appearing in the numerators of the measures are the same. Suppose $j \in \mathcal{P}_k$ and $t_i \in [\mathbf{q}_j, \mathbf{q}_{j+1}[\subseteq [\mathbf{p}_k, \mathbf{p}_{k+1}[$. Then,

$$(S(\mathbf{p})\hat{\boldsymbol{\beta}}(\mathbf{p}))_i = \hat{\boldsymbol{\beta}}_k(\mathbf{p}) = \frac{1}{|\mathcal{I}_k(\mathbf{p})|} \sum_{l \in \mathcal{I}_k(\mathbf{p})} \mathbf{w}_l, \tag{2.43}$$

$$(S(\mathbf{q})\hat{\boldsymbol{\gamma}}(\mathbf{q}))_i = \hat{\boldsymbol{\gamma}}(\mathbf{q})_j = \mu(\mathcal{P}_k) = \frac{\sum_{l \in \mathcal{P}_k} |\mathcal{I}_l(\mathbf{q})| \hat{\boldsymbol{\beta}}_l(\mathbf{q})}{\sum_{l \in \mathcal{P}_k} |\mathcal{I}_l(\mathbf{q})|}. \tag{2.44}$$

Since $\mathcal{I}_k(\mathbf{p}) = \cup_{l \in \mathcal{P}_k} \mathcal{I}_l(\mathbf{q})$, (2.43) and (2.44) are equal. □

The most important properties of PWC MMTM are summarized in the next theorem as the consequences of the previous propositions.

Theorem 2.4.2. *Let* $\mathbf{t}, \mathbf{w} \in \mathbb{R}^d$, $\mathbf{q} \in \mathcal{B}(\mathbf{t}, b)$,

1. *if* $\mathcal{M}$ *is a monotonic intensity transformation,* $D^{PWC}_{MMTM}(\mathbf{t}, \mathcal{M}(\mathbf{t})|\mathbf{q}) = D^{PWC}_{MTM}(\mathbf{t}, \mathcal{M}(\mathbf{t})|\mathbf{q})$;
2. $D^{PWC}_{MMTM}(\mathbf{t}, \mathbf{w}|\mathbf{q}) \geq D^{PWC}_{MTM}(\mathbf{t}, \mathbf{w}|\mathbf{q})$;
3. $D^{PWC}_{MMTM}(\mathbf{t}, \mathbf{w}|\mathbf{q}) \in [0, 1]$.

Proof.

1. Since a monotonic transformation $\mathcal{M}$ implies that the elements of $\hat{\boldsymbol{\beta}} \in \mathbb{R}^b$ are also monotonically increasing, PAVA resolves $v = 0$ violations and PWC MMTM turns into PWC MTM.

2. Suppose $\mathbf{t}, \mathbf{w}, \mathbf{q}, \mathbf{p}$ and v satisfy the conditions of Proposition 2.4.2. If $v = 0$, equality holds as a consequence of 1). Let $v \geq 1$. Then, one can construct a sequence of vectors $\mathbf{r}_0 = \mathbf{q}$, $\mathbf{r}_i \in \mathcal{B}(\mathbf{t}, b - i), i =$

$1, \ldots, v-1, \mathbf{r}_v = \mathbf{p}$ such, that the conditions of Proposition 2.4.1 hold for each adjacent pair of vectors. As a consequence of Proposition 2.4.1 and 2.4.2,

$$\begin{aligned} &D^{PWC}_{MTM}(\mathbf{t}, \mathbf{w}|\mathbf{q}) < D^{PWC}_{MTM}(\mathbf{t}, \mathbf{w}|\mathbf{r}_1) < \ldots \\ &< D^{PWC}_{MTM}(\mathbf{t}, \mathbf{w}|\mathbf{r}_{v-1}) < D^{PWC}_{MTM}(\mathbf{t}, \mathbf{w}|\mathbf{p}) = \\ &= D^{PWC}_{MMTM}(\mathbf{t}, \mathbf{w}|\mathbf{q}). \end{aligned} \tag{2.45}$$

3. The third statement is the consequence of Proposition 2.4.2 and the fact that PWC MTM is normalized into $[0, 1]$.

□

Theorem 2.4.2 shows that PWC MMTM meets the requirements we formulated at the beginning of the section. On the one hand, the advantageous properties of PWC MTM (like being an absolute measure and being approximately invariant to monotonic intensity transformations) are inherited by PWC MMTM. On the other hand, since the number of violations resolved by PAVA can be considered as a qualitative measure of how $\mathcal{M}$ deviates from being monotonic, (2.45) implies that the deviation of $D^{PWC}_{MMTM}(\mathbf{t}, \mathcal{M}(\mathbf{t})|\mathbf{q})$ from $D^{PWC}_{MTM}(\mathbf{t}, \mathcal{M}(\mathbf{t})|\mathbf{q})$ increases with the deviation of $\mathcal{M}$ from being a monotonically increasing transformation.

2.4.4 Piecewise linear approximation

Formally, the PWC and PWL MTM measures are highly similar. In both cases, the slice transform matrix S or Q contains structural information about $\mathbf{t}$ and a kind of weight vector $\boldsymbol{\beta}$ or $\boldsymbol{\zeta}$ is determined to transform the structure of $\mathbf{t}$ as close to $\mathbf{w}$ as possible. However, there is a significant difference in the slice transform matrices: S is orthogonal, but Q is not. Changing $\boldsymbol{\beta}_i$ alters only the elements belonging to the ith bin in $S\boldsymbol{\beta}$. On the contrary, changing $\boldsymbol{\zeta}_i$ alters the elements of $Q\boldsymbol{\zeta}$ belonging to bin i and $i+1$, as well. This complicated encoding of structure and intensity transformation gives rise to some strange properties of PWL MTM: the monotonicity of $\mathcal{M}$ does not imply the monotonicity of $\hat{\boldsymbol{\zeta}}$ in $D^{PWL}_{MTM}(\mathbf{t}, \mathcal{M}(\mathbf{t}))$. We use a simple example to demonstrate this behaviour. Let $\mathbf{t} = (2, 3, 8)^T$ be quantized into two bins defined by $\mathbf{q} = (1, 5, 10)^T$, and let the tone mapping be the strictly monotonic sigmoid function

$\mathcal{M}(x) = \dfrac{1}{1+e^{-x}}$. Then, $\mathcal{M}(\mathbf{t}) = (0.88, 0.95, 0.99)^T$, $\hat{\boldsymbol{\zeta}} = (0.8, 1.0, 0.93)^T$ and $Q\hat{\boldsymbol{\zeta}} = (0.88, 0.95, 0.99)^T$. $D^{PWL}_{MTM}(\mathbf{t}, \mathcal{M}(\mathbf{t})) = 0$, since $\mathbf{t}$ is perfectly reconstructed by $Q\hat{\boldsymbol{\zeta}}$, but $\hat{\boldsymbol{\zeta}}$ is not monotonic. Consequently, adding monotonicity constraints to $\boldsymbol{\zeta}$ in the formulation of PWL MTM (2.26) is not the appropriate way to develop the PWL MMTM measure, since the outcome of PWL MTM and PWL MMTM may differ even if $\mathcal{M}$ is monotonic. In the development of the proposed PWL MMTM measure we have taken into consideration the following natural requirements: for a monotonic tone mapping $\mathcal{M}$, $D^{PWL}_{MTM}(\mathbf{t}, \mathcal{M}(\mathbf{t})) = D^{PWL}_{MMTM}(\mathbf{t}, \mathcal{M}(\mathbf{t}))$ is desired and the outcome of PWL MMTM should deviate from PWL MTM as $\mathcal{M}$ deviates from being monotonic; PWL MMTM should be normalized into the range $[0, 1]$ and computed in linear time.

We found that all the requirements can be fulfilled by adding a penalty term to PWL MTM, penalizing the non-monotonic relationship between $\mathbf{t}$ and $\mathbf{w}$:

$$D^{PWL}_{MMTM}(\mathbf{t}, \mathbf{w}) = D^{PWL}_{MTM}(\mathbf{t}, \mathbf{w}) + \left(1 - D^{PWL}_{MTM}(\mathbf{t}, \mathbf{w})\right) P(\mathbf{t}, \mathbf{w}), \tag{2.46}$$

$$P(\mathbf{t}, \mathbf{w}) = D^{PWC}_{MMTM}(\mathbf{t}, S\hat{\boldsymbol{\beta}}). \tag{2.47}$$

The penalty term is based on the quantities computed in PWC MMTM and measures how much the optimal monotonic PWC tone mapping ($\hat{\boldsymbol{\gamma}}$) deviates from the optimal PWC tone mapping ($\hat{\boldsymbol{\beta}}$). The reason why the PWC MMTM distance of $\mathbf{t}$ and $S\hat{\boldsymbol{\beta}}$ is used instead of that of $\mathbf{t}$ and $\mathbf{w}$ is to get rid of the PWC representation error. In this formulation the penalty term becomes zero if $\mathcal{M}$ is monotonic and $D^{PWL}_{MMTM}(\mathbf{t}, \mathcal{M}(\mathbf{t})) = D^{PWL}_{MTM}(\mathbf{t}, \mathcal{M}(\mathbf{t}))$ holds.

Theorem 2.4.3. *Using the notations introduced before,*

1. *if $\mathcal{M}$ is a monotonic intensity transformation, $D^{PWL}_{MMTM}(\mathbf{t}, \mathcal{M}(\mathbf{t})) = D^{PWL}_{MTM}(\mathbf{t}, \mathcal{M}(\mathbf{t}))$;*
2. *$D^{PWL}_{MMTM}(\mathbf{t}, \mathbf{w}) \geq D^{PWL}_{MTM}(\mathbf{t}, \mathbf{w})$;*
3. *$D^{PWL}_{MMTM}(\mathbf{t}, \mathbf{w}) \in [0, 1]$.*

Proof. All statements are simple consequences of the structure of (2.46), the properties of PWC MMTM and the fact that PWL MTM is normalized in the range $[0, 1]$. □

The authors of [1] give an $O(2p(d+b))$ algorithm for the computation of the PWL MTM dissimilarity of a template and all windows of an image containing p pixels. We have previously seen that the time complexity of PWC MMTM is $O(p(d+2b))$, thus, the complexity of the proposed PWL MMTM measure remains linear: $O(p(3d+4b))$.

2.4.5 The Window-to-Pattern cases

The authors of [1] give a detailed description of the Window-to-Pattern (WtP) measures and also give linear time algorithms for their computation. Due to the analogies of the WtP and PtW cases, the monotonicity constraints can be added to the WtP measures in a straightforward way, carrying out the same steps we did in the previous subsections. We do not repeat the derivations here, but highlight that analogous statements can be formulated for the MMTM WtP measures as the ones we made in the PtW case. As expected, the time complexity of the WtP measures remains linear.

2.4.6 Pattern-to-Window or Window-to-Pattern?

As mentioned before, the MTM and MMTM measures are not symmetric: the outcome of the PtW and WtP variants is usually different. One can create symmetric measures based on the MTM or MMTM principles in many ways, e.g. the sum or minimum of the PtW and WtP measures is symmetric. In practice, we found that there is no need to calculate both of the measures. Under the natural assumption that the template being sought for has some characteristic structure, one can expect the WtP measures to have better performance than the PtW ones. We illustrate this strange property with an example. In Figure 2.2 one of our test images can be seen with two windows zoomed-in. Suppose one wants to do template matching with the template in Figure 2.2(b). Either MTM or MMTM PtW measure is used, the window in Figure 2.2(c) gives an almost perfect match, since even a monotonic tone mapping can transform the text on the spine of the book into the corresponding blank region of Figure 2.2(c). However, using the WtP measures, this is not the case. The blank region can not be mapped to the one with structures, so the WtP measures won't indicate a good match. As a summary, supposing that the template

(a) (b) (c)

Figure 2.2: The template in (a) and (b) almost perfectly matches the window (c) when a PtW measure is used

has some characteristic, detailed pattern, the use of the WtP measures is recommended.

2.5 Tests and Results

In the previous section we have introduced the PWC and PWL MMTM measures. In this section we compare the performance of the proposed techniques to that of other measures in various scenarios of template matching. The scenarios are similar to the ones used in [1] and [40] for the evaluation of (dis)similarity measures in template matching.

2.5.1 The measures used for comparison

For comparison we have selected (dis)similarity measures designed to be more-or-less invariant to monotonic intensity transformations: Spearmans's Rho coefficient (SR) [7]; Kendall's Tau coefficient (KT) [8]; the L_1 distance of the rank vectors (RD) [40]; local binary patterns (LBP) [25]; and the Pearson correlation coefficient (PCC). In order to demonstrate the improvements, the test results are also compared to that of the PWC and PWL MTM measures.

Generally, in each of the test scenarios we found that LBP has significantly worse performance than the other measures. This is in accordance with the results reported in [1]. In our interpretation the reason for this is

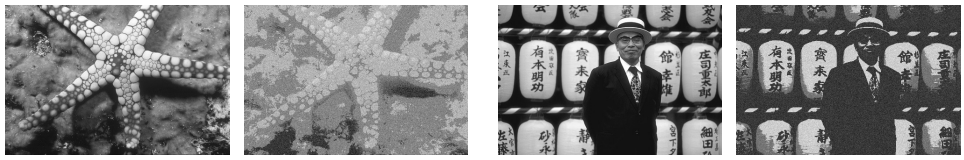


Figure 2.3: Some test images followed by their distorted variants

the high sensitivity of LBP to the distortions used in the tests (monotonic tone mappings, additive noise). In order to improve the readability of the charts and figures, we do not plot and do not discuss the results of LBP in the rest of the chapter.

2.5.2 Simulated images

Similarly to the evaluation in [1], the measures are tested on images distorted by simulated tone mappings and noise. These experiments enable the comparison of performance as a function of the amount of distortion.

2.5.2.1 Test cases and the measure of performance

In the simulations we used the images of the *Berkeley Image Segmentation Dataset* [47]. In every test case we randomly choose an image I from the dataset, and randomly select a window having standard deviation larger than a specific threshold. This window is considered to be the template T. The selection rule ensures that the template has some structure, which is a natural assumption in template matching scenarios. Then, a randomly generated monotonic tone mapping is applied to I, Gaussian white noise is added and template matching is carried out to find T in the distorted image. The position of the highest similarity score is considered to be the detected position of the template. We consider the match correct if the location is inside the window we used as template. The measure we compute to characterize the performance of the various (dis)similarity functions is the *matching rate* defined as

$$\frac{\text{the number of correct matches}}{\text{the total number of tests}}. \tag{2.48}$$

Algorithm 2 Recursive procedure to generate random monotonic tone mappings

```
1: procedure MonMap(b ∈ ℤ, e ∈ ℤ, m ∈ {0,...,255}^256)
2:     c ← ⌊ (b + e)/2 ⌋
3:     if b ≠ c then
4:         m_c ← r ∈ U({m_b, m_b + 1, ..., m_e})
5:         MonMap(b, c, m)
6:         MonMap(c, e, m)
7:     else
8:         return
9:     end if
10: end procedure
```

In each experiment we carry out 1000 tests.

The random monotonic tone mappings are generated in the following way. In the usual 8 bit grayscale representation of images, an arbitrary tone mapping can be represented by a vector $\mathbf{m} \in \{0, \ldots, 255\}^{256}$. The tone mapping is considered to map the intensity $\mathbf{m}_i$ to the gray level i. The first element of the random monotonic mapping is set to zero ($\mathbf{m}_1 = 0$) and the last element to a random integer in the range $[1, 255]$: $\mathbf{m}_{256} = r \in \mathcal{U}(\{1, \ldots, 255\})$. Then, we apply the recursive method of Algorithm 2 to fill the rest of the mapping $\mathbf{m}$ by calling MonMap$(1, 256, \mathbf{m})$.

In order to illustrate the effect of random monotonic tone mappings and additive noise, some images of the Berkeley dataset and their distorted variants are visualized in Figure 2.3.

2.5.2.2 Varying the number of bins

In the first experiment we examined the performance of the proposed MMTM measures as the number of bins used for the slice transform is varied. We used the testing methodology described in the previous subsection. The number of bins was set to $4, 8, \ldots, 24$; the noise level and the linear extent of the templates were selected from the ranges $[1, 25]$ and $[11, 31]$, respectively. The results are plotted and compared to that of the MTM measures in Figure 2.4(a).

Generally, the PWC measures have low performance when the number of bins is low, but the matching rates increase rapidly as the number of bins

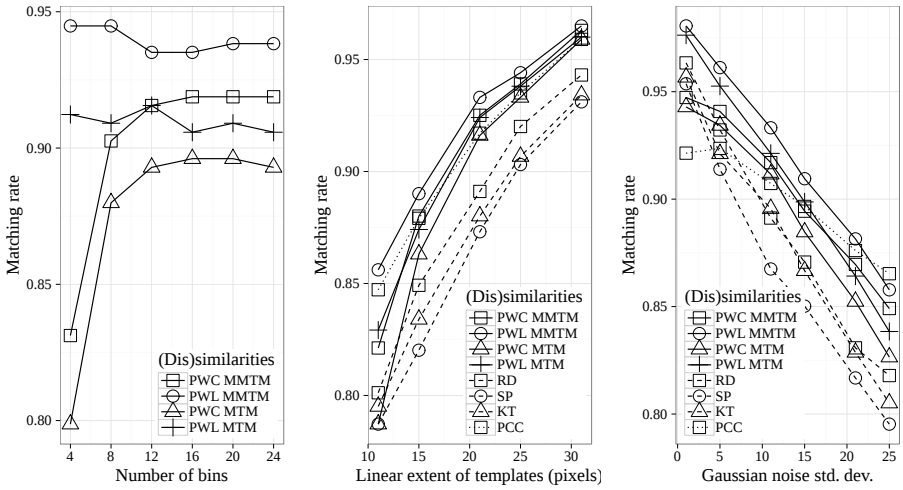
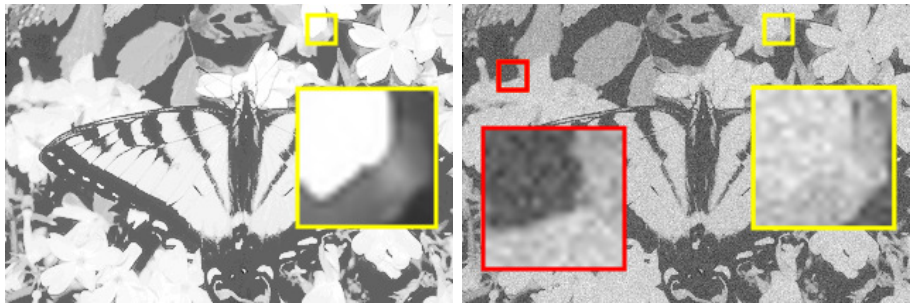


Figure 2.4: The sensitivity of performance to the number of bins (a); template size (b); and level of noise (c)

increases and the performance saturates near 16 bins. The PWL measures seem to be quite insensitive to the number of bins, although a slight drop in performance can be observed when the number of bins exceeds 8. One can observe that there is a remarkable difference in the performance of the MMTM and corresponding MTM measures. Moreover, despite the PWC representation error, even PWC MMTM outperforms PWL MTM when the number of bins is greater than 12. The reason for the relatively low matching rates of MTM is the issue described in the Section 2.1: the MTM measures do not take into consideration the monotonicity properties of the pattern, thus, they may detect the inverse of the template, as well. We illustrate the problem in Figure 2.5: one can see a test image and a randomly selected template (yellow frame, zoomed-in) in Figure 2.5(a); and the distorted image (random monotonic tone mapping and additive noise) with the windows detected by MTM (red frame) and MMTM (yellow frame) in Figure 2.5(b). The MMTM measures find the template correctly, but the MTM measures fail: although the structure of the window in red



(a) (b)

Figure 2.5: The test image and the template in yellow frame (a); the distorted image and the windows identified by MMTM (yellow frame) and MTM (red frame) (b)

frame is similar to the structure of the template, the intensities are inversely related.

As a side-result of the experiment we can select the appropriate number of bins to use for the slice transform: in the rest of the chapter we use the PWC measures with 16 bins and the PWL measures with 8 bins.

2.5.2.3 Varying the size of the template

In the second experiment we examined the sensitivity of performance to the size of the template. In each test case the level of noise was selected from the range $[1, 25]$. The results are plotted in Figure 2.4(b). Although some convergence can be observed as the size of the templates increases, PWL MMTM has the best performance all over the range of template sizes. PWC MMTM has similar performance as PWL MTM, which means that the improvement achieved by taking into consideration the monotonicity properties of the template is comparable to the improvement achieved by using PWL approximation instead of PWC in MTM. The only measure having remarkably better performance than PWC MMTM is PCC at the template size of 11×11 pixels: a template consisting of 121 pixels seems to be too small for an accurate PWC representation using 16 bins. Interestingly, the rank based techniques have remarkably worse performance than PCC: we can conclude that PCC is less sensitive to additive noise and the possible

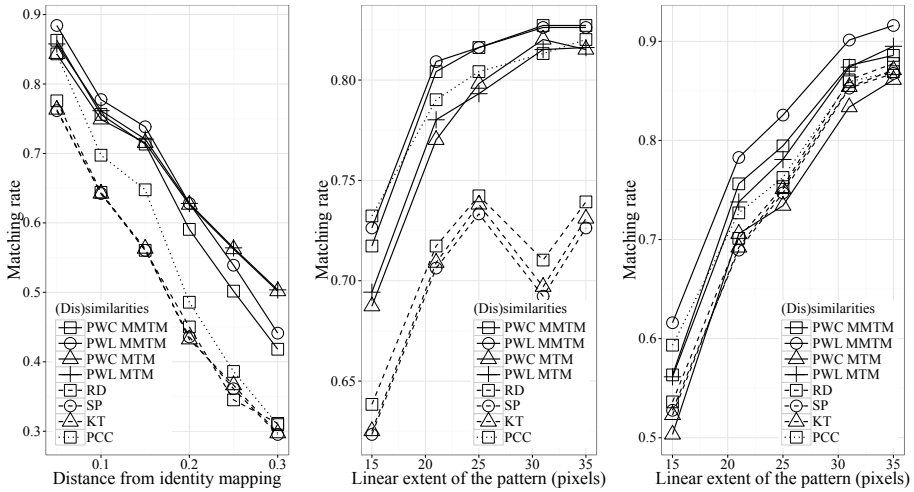


Figure 2.6: Matching rates as the simulated tone mappings deviate from monotonic (a); test results on the *Robust Pattern Matching performance evaluation dataset* (b); tests results on own images (c)

loss of details (due to the monotonic tone mappings) than the rank based techniques.

2.5.2.4 Varying the level of noise

In this experiment the sensitivity of the measures to additive noise is examined. Again, the template sizes were selected from the range $[11, 31]$. As one can observe in the results plotted in Figure 2.4(c), PWL MMTM outperforms the other techniques almost all over the range of noise levels. Interestingly, when the level of noise is close to zero, the rank based techniques have higher matching rates than PCC, PWC MTM and PWC MMTM. This can be explained by recalling that the rank based techniques are exactly invariant to monotonic tone mappings. At low levels of noise their performance is deteriorated only by the loss of details (due to degenerated monotonic tone mappings), while PCC and the PWC measures are also affected by the non-linearity of the tone mapping and the PWC

representation error, respectively. As the level of noise increases, the performance of the rank based methods deteriorates rapidly showing their high sensitivity to additive noise. Although the performance of PCC shows the slowest decrease by the increasing level of noise, the average performance of the MMTM measures – especially that of the PWL MMTM – is highly competitive with that of the other techniques.

2.5.2.5 Sensitivity to the deviation from monotonic tone mappings

In the last experiment of simulations, we examined the performance of the (dis)similarity measures in the presence of slightly non-monotonic tone mappings. The template size and the level of noise were selected randomly from the same ranges as before. The random non-monotonic tone mappings were produced by generating a monotonic tone mapping and randomly interchanging some elements of the mapping vector. The deviation of a non-monotonic tone mapping $\mathbf{m}$ from the identity mapping $\mathbf{i}$ is characterized by computing $D^{PWC}_{MMTM}(\mathbf{i}, \mathbf{m})$ using 256 bins. It is worth noting that the use of PWC MMTM for the quantification of the non-monotonicity of the mapping does not introduce any bias for the proposed methods. As one can see in Figure 2.6(a), the MTM and MMTM techniques highly outperform the rank correlation based methods and PCC as the tone mapping deviates from being monotonic. Interestingly, up to a certain point ($D^{PWC}_{MMTM}(\mathbf{i}, \mathbf{m}) \approx 0.2$), the PWL MMTM method outperforms the MTM measures and even PWC MMTM performs on par with them. We can conclude that the use of monotonicity constraints may improve the performance of the MTM measures even if the tone mappings are not perfectly monotonic.

2.5.3 Real images

In the previous subsection we have shown that the proposed measures perform on par and in certain cases better then previous techniques. However, in real scenarios it is usual that non-monotonic and spatially varying tone mappings appear on the images. In order to demonstrate how the monotonicity constraints improve the performance of the MTM measures in real scenarios, we carried out experiments on real images and summarize

Figure 2.7: Sample images from the Robust Pattern Matching performance evaluation dataset

the results in the rest of this subsection.

2.5.3.1 Robust Pattern Matching performance evaluation dataset

In the literature of image processing we found only a limited number of datasets used for benchmarking (dis)similarity measures. One publicly available database is the *Robust Pattern Matching performance evaluation dataset* [48] containing 10 photos of a structured scene. The images were taken under various lighting conditions and using various imaging devices. The images also contain some geometric distortions: the scene is slightly rearranged and the position of the camera is also varied to some extent. For illustration, some sample images from the dataset are visualized in Figure 2.7.

The testing methodology is similar to the one we used in the previous subsections, except that the simulated distortion of an image is replaced by the random selection of another image from the database. As before, we carried out 1000 tests and plotted the matching rates in Figure 2.6(b) by varying template size. One can observe that MTM, MMTM and PCC highly outperform the rank based measures. The reason for this is that the rank based measures are highly sensitive to geometric distortions: if the template and the target window do not overlap perfectly, the rankings may become far from each other and the matching fails. Regarding the proposed techniques, one can also observe that the MMTM measures give remarkably better results than the MTM techniques and in most of the cases outperform PCC, as well. Two strange things can be observed at the linear size of 31 pixels. On the one hand, the performance of the

Figure 2.8: Some sample images from the database we prepared

MTM, MMTM and PCC measures saturates: the reason for this is that the deteriorating effect of geometric distortions (like slight rotations) increases with the size of the templates. On the other hand, there is a drop in the performance of rank based techniques. We found the reason for this strange behaviour in the presence of highly non-monotonic local distortions (like the spotlight in Figure 2.6(b)). As the size of templates exceeds 25×25 pixels, the probability that a randomly selected structured template is affected by the spatially limited highly non-monotonic distortions increases rapidly and the rank based techniques seem to fail in these situations since the ranking of the template and the window becomes highly different. As the size of templates increases further, the deteriorating effect of spatially limited distortions decreases (as their relative size within the template decreases) and that of geometric distortions dominates: the matching rates increase and just like in the other cases, the performance seems to saturate near the level of 0.74.

2.5.3.2 Own images

Since the publicly available dataset used in the previous experiment is pretty small, we have prepared a larger dataset of test images. The images are arranged into 10 series, each of them containing 20 images of the same scene. The images are taken under various lighting conditions, at various levels of CCD sensitivity and some of them contains out-of-focus

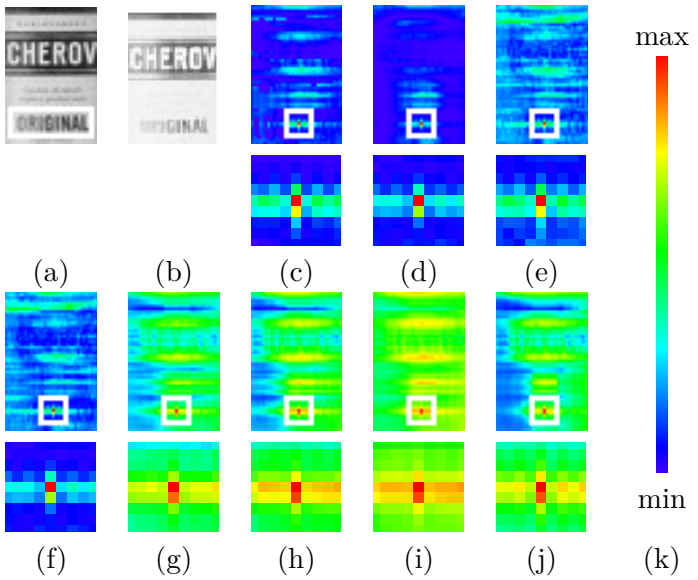


Figure 2.9: A test image and the template in white frame (a); the corresponding region of another test image (b); In each distance map the pixel with the highest score is located in the middle of the small white frame, which is zoomed-in below the distance maps for better visibility: MMTM PWC (c); MMTM PWL (d); MTM PWC (e); MTM PWL (f); RD (g); SR (h); KT (i) and PCC (j). The scale of false coloring: *max* and *min* are related to the best and worst score (k)

blur. For illustration, one can see some sample images in Figure 2.8. We carried out similar experiments as before and plotted the results in Figure 2.6(c). The most conspicuous thing is that the rank based techniques have higher matching rates than in the previous experiment, moreover, they outperform PWC MTM almost all over the range of template sizes. This can be explained by the lack of heavy geometric distortions (like rearrangement) and the limited amount of non-monotonic local intensity distortions. Similarly to the previous experiments, PWL MMTM has the best performance, and in most of the cases even the matching rates of PWC MMTM are higher than that of the other measures used for comparison.

2.5.3.3 Qualitative illustration of the discrimination power

The test results presented so far suggest that – in the presence of monotonic or slightly non-monotonic tone mappings and additive noise – the MMTM measures can distinguish a structured template from the background more efficiently than the methods used for comparison. In Figure 2.9 we demonstrate this characteristics qualitatively. Figure 2.9(a) shows a segment of a test image and a template in white frame. The template was sought in the slightly over-exposed test image of Figure 2.9(b) and the distance maps produced by the various measures are shown in Figures 2.9(c)-2.9(j). For the ease of visualization false coloring is used; the neighborhoods of the best matching pixels (small white frames) are zoomed-in and both the similarity and dissimilarity scores are normalized into the range [0,1], where 0 and 1 denote the worst and best match, respectively. Although each measure finds the template, one can observe that the distance maps of the MMTM measures are less noisy than that of the others, the peaks are sharper and better localized.

2.6 Discussion and Conclusion

In this chapter, two dissimilarity measures referred as PWC and PWL MMTM have been proposed, both of them based on principles of *Matching by Tone Mapping*. The measures are approximately invariant to monotonic tone mappings and can be computed in linear time. The derivation of PWC MMTM leads to a constrained optimization problem, for which we managed to find a linear time solution using the PAVA algorithm. In the PWL case we showed that due to the complex encoding of structure and tone mappings, it is not straightforward to put monotonicity constraints in the formulation of the measure. Instead, the constraints are incorporated into a penalty term to construct PWL MMTM from PWL MTM. We have carried out several experiments on simulated and real images to evaluate the discrimination power of the measures and compare their performance to that of other techniques designed to be invariant to monotonic tone mappings. Regarding the sensitivity to some specific distortions (like additive noise and monotonic tone mappings), we found that the proposed measures are highly competitive with previous techniques: in most of the experiments PWL MMTM produced the highest matching rates and PWC MMTM

had similar results as the best performing method from the competitors. We emphasize that throughout the derivation of the measures we have not utilized the dimensionality of the signal, thus, the proposed methods can be used to measure the dissimilarity of 1D or 3D signals, as well.

Finally, the first thesis of the dissertation is formulated as follows.

Thesis 1.

1. *I have introduced the Matching by Monotonic Tone Mapping (MMTM) dissimilarity measure and its piecewise constant (PWC) and piecewise linear (PWL) approximations, both of them being approximately invariant to monotonic, even non-linear transformations.*

2. *I have shown that both of the measures inherit the advantageous properties of the MTM measures, they are absolute measures and can be computed efficiently.*

3. *Based on the test results I can state that the discrimination power of the MMTM measures is highly competitive with that of previous techniques, the measures can be used efficiently in problems, where a dissimilarity measure invariant to monotonic transformations is required.*

2.7 Outlook

We have some preliminary results on how the quantization affects the performance of the MTM measure [49], namely, it can be showed theoretically that the k-means clustering based quantization can improve the recognition of patterns on noisy backgrounds. Another interesting problem is the extension of the MMTM measure to multichannel images by finding a proper representation of structure using information from all channels.

Chapter 3

Translation invariant (dis)similarity measures in kernel space

In the previous chapter we have given a brief introduction to (dis)similarity functions and presented the Matching by Monotonic Tone Mapping measure as a novel contribution to the field. In this chapter we go further and examine some properties of (dis)similarity functions used in general machine learning and pattern recognition applications.

3.1 Motivation and Introduction

Some of the measures mentioned and discussed so far operate directly on the vectors being compared (like the Euclidean distance, or the PCC). Other measures, like LBP and the rank correlation based techniques, apply some transformation and measure the (dis)similarity of the transformed vectors using one of the simplest (dis)similarity measures. In numerous problems the transformation of vectors can be beneficial, since the important attributes and characteristics of vectors can be emphasized by a properly chosen transformation (like the rank transformation which extracts the rankings of vector coordinates, providing invariance to monotonic distortions). Formally, a non-linear so called *feature mapping* can be considered in the form ϕ :

$\mathbb{R}^d \to \mathbb{R}^{\backslash}$, to map the vectors into a so called *feature space*. Usually, $n >> d$. The aim of feature mapping is to boost the performance of (dis)similarity measures by providing a better distribution of the data and emphasizing the most important characteristics of the vectors. As a shorthand, it is usually ambiguous to find a proper mapping ϕ, and the optimal transformation can map to a very high (possibly infinite) dimensional space. The high dimensionality of $\mathcal{F}$ makes it expensive both in terms of memory and time to compute and store the high dimensional feature vectors used for comparison instead of the primary vectors. Thus, besides improving discrimination power, efficiency issues need to be addressed, as well.

One can overcome the efficiency problems if proper knowledge on the feature mapping is not required for the comparison of the mapped feature vectors in $\mathcal{F}$. For example, having a dissimilarity measure expressed in terms of inner products, one can consider the dissimilarity in the high dimensional image space by replacing the inner product in the dissimilarity by $\langle \phi(\mathbf{x}), \phi(\mathbf{y}) \rangle$, $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$. If one can find a function (*kernel*) $k : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ which represents an inner product in the image space $\mathcal{F}$ of ϕ without computing $\phi(\mathbf{x})$ and $\phi(\mathbf{y})$ explicitly, that is, satisfying

$$k(\mathbf{x}, \mathbf{y}) = \langle \phi(\mathbf{x}), \phi(\mathbf{y}) \rangle, \ \forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^d, \tag{3.1}$$

then, replacing the inner products of input vectors in the dissimilarity function by kernel evaluations (3.1), the discrimination power of the measure can be enhanced (the data may become better separable in the high dimensional space) and the computational demands may remain acceptable. This replacement is called the *kernel trick* [50] in the field of machine learning. In practice, the kernel trick is performed in the opposite way: there are several functions on $\mathbb{R}^d \times \mathbb{R}^d$ proved to be an inner product after mapping the parameters into a high dimensional space and these functions are used as kernels to improve the discrimination power of (dis)similarity functions in certain applications. Numerous kernel functions are proposed and used in the literature of machine learning; for a thorough overview see [51, 52], some well known and widely used ones are the polynomial kernel [53]

$$k(\mathbf{x}, \mathbf{y}) = (\gamma + \alpha \langle \mathbf{x}, \mathbf{y} \rangle)^q, \tag{3.2}$$

$\gamma, \alpha \in \mathbb{R}$, $q \in \mathbb{N}$, which can be shown to operate in finite dimensional

feature spaces and the Gaussian kernel [53]

$$k(\mathbf{x}, \mathbf{y}) = \exp\left(-\frac{\|\mathbf{x} - \mathbf{y}\|^2}{2\sigma^2}\right), \tag{3.3}$$

$\sigma \in \mathbb{R}^+$, which can be shown to operate in an infinite dimensional feature space.

Regarding (dis)similarity measures, Schölkopf proposed the kernel trick to improve the discrimination power of the Euclidean-distance [50]:

$$D^k_{EUC}(\mathbf{x}, \mathbf{y}) = \left(k(\mathbf{x}, \mathbf{x}) + k(\mathbf{y}, \mathbf{y}) - 2k(\mathbf{x}, \mathbf{y})\right)^{\frac{1}{2}}, \tag{3.4}$$

and it can be readily seen that the normalization of any kernel [54] computes the correlation coefficient in kernel space:

$$S^k_{CC}(\mathbf{x}, \mathbf{y}) = \frac{k(\mathbf{x}, \mathbf{y})}{\sqrt{k(\mathbf{x}, \mathbf{x})k(\mathbf{y}, \mathbf{y})}}. \tag{3.5}$$

Both of these kernelized measures are constructed by simply replacing the inner product in the original formulation of the measures to a kernel function k. Similarly to the Euclidean distance, the kernelized Euclidean distance is not invariant to linear transformation in the kernel space, and just like CC, kernelized CC is invariant to only scaling in the kernel space.

Besides the kernelization of (dis)similarity measures, many of the most commonly used machine learning techniques are built or generalized to use kernel functions, like Support Vector Machines [54], nearest neighbor algorithms [55], semi-naive Bayesian classifiers [56], clustering methods [57] and principal component analysis [58]. As characteristic fields of applications, we can mention numerous topics in bioinformatics (gene and cancer research [59]), face recognition [60], speech recognition [61], computer vision [53], etc.

The main advantage of working with kernels is that they enable calculations in high (possibly infinite) dimensional feature spaces efficiently, however, the operations we can perform in the kernel space efficiently are limited to kernel evaluations. Consequently, the kernelization of (dis)similarity measures that cannot be formulated as linear combinations of inner products is not straightforward. For example, in spite of its popularity, there are no results on the kernelization of the widely used PCC measure. The difficulty

in the kernelization of PCC is its translation invariance, namely, one has to know the mean of the coordinates of the feature map to compute PCC in the kernel space. In this chapter we examine the kernelization of the centered Euclidean distance (cEUC) and that of PCC in general, and for one particular kernel: the widely used polynomial kernel function. We show that under mild conditions, there are infinitely many ways to calculate these translation invariant measures in the kernel space, and there are countably infinitely many ways to do it in terms of kernel evaluations. To show the advantages of the application of centered dissimilarities in kernel space, we have carried out several tests with simulated and real data in kNN based classification scenarios. The results show that adding translation invariance in the kernel space does not deteriorate the performance of the dissimilarity measures, moreover, in certain cases better results are achieved by the proposed improvements.

The chapter is organized as follows. In the first part of Section 3.2 we give some general results on the kernelization of the cEUC and PCC measures. In the second part of the section the results related to kernelization by the polynomial kernels are summarized. In Section 3.3 the experimental tests and corresponding results are discussed, and finally, some conclusions are drawn in Section 3.4 and a brief outlook is given in Section 3.5. In the lack of space we do not discuss all our results in related to translation invariance in kernel space. Some further, minor results can be found our corresponding paper [62].

3.2 Translation invariance in kernel space

As mentioned earlier, in the rest of the chapter we focus on the simplest translation invariant measures, namely, D_{cEUC}, and S_{PCC}, and examine how they can be kernelized with the polynomial kernel. For kernelization, the functions D_{cEUC} and S_{PCC} should be composed as the linear combination of inner products of single data vectors, since centering the vector in the input space and then replacing the inner products with possibly non-linear kernels does not give centering in the feature space due to the

possible non-linearity of ϕ:

$$k(\mathbf{x}-\bar{\mathbf{x}}\mathbf{1}, \mathbf{y}-\bar{\mathbf{y}}\mathbf{1}) = \langle \phi(\mathbf{x}-\bar{\mathbf{x}}\mathbf{1}), \phi(\mathbf{y}-\bar{\mathbf{y}}\mathbf{1}) \rangle \neq \langle \phi(\mathbf{x})-\overline{\phi(\mathbf{x})}\mathbf{1}, \phi(\mathbf{y})-\overline{\phi(\mathbf{y})}\mathbf{1} \rangle. \tag{3.6}$$

The centralized (dis)similarity measures S_{PCC} and D_{cEUC} (introduced in Section 2.2.1) can be expanded by exploiting the bilinearity of the inner product:

$$S_{PCC}(\mathbf{x}, \mathbf{y}) = \frac{\langle \mathbf{x}, \mathbf{y} \rangle - d\bar{\mathbf{x}}\bar{\mathbf{y}}}{((\langle \mathbf{x}, \mathbf{x} \rangle - d\bar{\mathbf{x}}^2)(\langle \mathbf{y}, \mathbf{y} \rangle - d\bar{\mathbf{y}}^2))^{1/2}}. \tag{3.7}$$

and

$$D_{cEUC}(\mathbf{x}, \mathbf{x}) = \Big(\langle \mathbf{x}, \mathbf{x} \rangle + \langle \mathbf{y}, \mathbf{y} \rangle - 2\langle \mathbf{x}, \mathbf{y} \rangle - d\overline{\mathbf{x}}^2 - d\overline{\mathbf{y}}^2 + 2d\overline{\mathbf{x}}\overline{\mathbf{y}} \Big)^{\frac{1}{2}}. \tag{3.8}$$

Using the expansions above, the kernelized variants of S_{CC}, S_{PCC}, D_{EUC} and D_{cEUC} become:

$$S^k_{CC}(\mathbf{x}, \mathbf{y}) = \frac{k(\mathbf{x}, \mathbf{y})}{(k(\mathbf{x}, \mathbf{x}))^{\frac{1}{2}} (k(\mathbf{y}, \mathbf{y}))^{\frac{1}{2}}}, \tag{3.9}$$

$$S^{k,\phi}_{PCC}(\mathbf{x}, \mathbf{y}) = \frac{k(\mathbf{x}, \mathbf{y}) - \dim(\phi)\overline{\phi(\mathbf{x})}\ \overline{\phi(\mathbf{y})}}{(k(\mathbf{x}, \mathbf{x}) - \dim(\phi)\overline{\phi(\mathbf{x})}^2)^{\frac{1}{2}} (k(\mathbf{y}, \mathbf{y}) - \dim(\phi)\overline{\phi(\mathbf{y})}^2)^{\frac{1}{2}}}, \tag{3.10}$$

$$D^k_{EUC}(\mathbf{x}, \mathbf{y}) = \big(k(\mathbf{x}, \mathbf{x}) + k(\mathbf{y}, \mathbf{y}) - 2k(\mathbf{x}, \mathbf{y})\big), \tag{3.11}$$

$$D^{k,\phi}_{cEUC}(\mathbf{x}, \mathbf{y}) = \big(k(\mathbf{x}, \mathbf{x}) + k(\mathbf{y}, \mathbf{y}) - 2k(\mathbf{x}, \mathbf{y}) - \dim(\phi)\overline{\phi(\mathbf{x})}^2 - \dim(\phi)\overline{\phi(\mathbf{y})}^2 + 2\dim(\phi)\overline{\phi(\mathbf{x})}\ \overline{\phi(\mathbf{y})}\big)^{\frac{1}{2}}, \tag{3.12}$$

where $\dim(\phi)$ denotes the dimensionality of the feature space ϕ is mapping into, and

$$\overline{\phi(\mathbf{x})} = \sum_{i=1}^{\dim(\phi)} \phi_i(\mathbf{x}) / \dim(\phi). \tag{3.13}$$

We introduce the notation k and ϕ in superscripts to indicate that a (dis)similarity function is kernelized and the centralization is carried out with the feature mapping ϕ. Note that the kernelization of S_{CC} and D_{EUC}

depends only on the kernel, but the kernelization of the centralized measures S_{PCC} and D_{cEUC} also depends on the feature mapping. The deduction above is symbolic, since in (3.13) the dimensionality of the feature mapping may be infinite and therefore the sum and the denominator may be infinite, as well. However, as we will see later on, the infinite dimensional case can be handled, as well.

The questions we are addressing in the rest of the chapter are the following ones: In general, can centralized dissimilarity measures be kernelized in a uniquivocal way? Can centralized dissimiliarty measures be kernelized by the polynomial kernel and efficiently computed in terms of kernel evaluations?

3.2.1 General results

Generally, the feature space can be any space in which an inner product is defined. For example, for any kernel function one can construct the so called *Reproducing Kernel Hilbert Space* (RKHS) [63], which contains functions of the type $\mathbb{R}^d \to \mathbb{R}$. However, as we have showed in the previous section, for the centralization of translation invariant dissimilarity measures, one has to work with the sum or mean of feature map coordinates, which can hardly be interpreted with feature mapping into such abstract spaces. Therefore, in the rest of the section we focus on feature mappings that map into the space of absolute convergent sequences ℓ^1, or into a finite dimensional vector space $\mathbb{R}^d$, $d \in \mathbb{N}$.

We start with giving some general statements for feature mappings into ℓ^1, particularly, we show that for these mappings, the centralized and non-centralized kernelized measures are the same.

Proposition 3.2.1. *Let $k : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ be a kernel function, which can be given in the form $k(\mathbf{x}, \mathbf{y}) = \langle \phi(\mathbf{x}), \phi(\mathbf{y}) \rangle$, where $\phi : \mathbb{R}^d \to \ell^1$. Then, $D^{k,\phi}_{cEUC}$ and $S^{k,\phi}_{PCC}$ belonging to the feature mapping ϕ are equivalent to D^{k}_{EUC} and S^{k}_{CC}, respectively.*

Proof. The term $\dim(\phi)\ \overline{\phi(\mathbf{x})}\ \overline{\phi(\mathbf{y})}$ in the expanded forms of $D^{k,\phi}_{cEUC}$ (3.12) and $D^{k,\phi}_{PCC}$ (3.10) can be written in a symmetric form:

$$\dim(\phi)\ \overline{\phi(\mathbf{x})}\ \overline{\phi(\mathbf{y})} = \frac{\sum\limits_{i=1}^{\infty} \phi_i(\mathbf{x})}{\sqrt{\dim(\phi)}} \frac{\sum\limits_{i=1}^{\infty} \phi_i(\mathbf{y})}{\sqrt{\dim(\phi)}}. \tag{3.14}$$

Note that the dimensionality of the feature space $\dim(\phi) = \infty$ in this case. Thus, (3.14) can be written as:

$$\dim(\phi)\ \overline{\phi(\mathbf{x})}\ \overline{\phi(\mathbf{y})} = \lim_{n\to\infty} \frac{\sum\limits_i \phi_i(\mathbf{x})}{\sqrt{n}} \frac{\sum\limits_i \phi_i(\mathbf{y})}{\sqrt{n}}. \tag{3.15}$$

Now, since $\phi(\mathbf{x}) \in \ell^1$, $\sum\limits_i |\phi_i(\mathbf{x})| < \infty$, thus, $\sum\limits_i \phi_i(\mathbf{x}) < \infty$ holds, and the limit in (3.15) equals 0. Consequently, $S^{k,\phi}_{PCC}$ reduces to S^k_{CC} and $D^{k,\phi}_{cEUC}$ to D^k_{EUC} in this case. □

For the ease of notations, based on the symmetric form (3.15) we introduce the concept of the centralizer function:

Definition 3.2.1. *Let k be a kernel $k : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$, which can be given in the form $k(\mathbf{x}, \mathbf{y}) = \langle \phi(\mathbf{x}), \phi(\mathbf{y}) \rangle$, where $\phi : \mathbb{R}^d \to \mathbb{R}^n$, $n \in \mathbb{N}$. The function $c^\phi : \mathbb{R}^d \to \mathbb{R}$ defined as:*

$$c^\phi(\mathbf{x}) := \frac{\sum\limits_{i=1}^{n} \phi_i(\mathbf{x})}{\sqrt{n}} \tag{3.16}$$

is called a centralizer function *of the kernel k belonging to the feature mapping ϕ.*

Using the centralizer function the kernelized measures can be rewritten in a simpler and symmetric form:

$$D^{k,\phi}_{cEUC}(\mathbf{x}, \mathbf{y}) = \Big(k(\mathbf{x}, \mathbf{x}) + k(\mathbf{y}, \mathbf{y}) - 2k(\mathbf{x}, \mathbf{y}) - (c^\phi(\mathbf{x}) - c^\phi(\mathbf{y}))^2\Big)^{\frac{1}{2}}, \tag{3.17}$$

$$S^{k,\phi}_{PCC}(\mathbf{x}, \mathbf{y}) = \frac{k(\mathbf{x}, \mathbf{y}) - c^\phi(\mathbf{x})c^\phi(\mathbf{y})}{(k(\mathbf{x}, \mathbf{x}) - c^\phi(\mathbf{x})c^\phi(\mathbf{x}))^{\frac{1}{2}} (k(\mathbf{y}, \mathbf{y}) - c^\phi(\mathbf{y})c^\phi(\mathbf{y}))^{\frac{1}{2}}}. \tag{3.18}$$

The following theorem shows that for a kernel having a finite dimensional feature mapping, at least countably infinitely many different centralizer functions can be constructed.

Theorem 3.2.1. *Let $k : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ be a non-zero kernel with the finite dimensional feature mapping $\phi : \mathbb{R}^d \to \mathbb{R}^n$, for which $k(\mathbf{x}, \mathbf{y}) = \langle \phi(\mathbf{x}), \phi(\mathbf{y}) \rangle$. Given the centralizer function c^ϕ, for any $n' > n$, $n' \in \mathbb{N}$, there exist a feature mapping $\psi : \mathbb{R}^d \to \mathbb{R}^{n'}$, for which $c^\psi(\mathbf{x}) = \frac{\sqrt{n}}{\sqrt{n'}} c^\phi(\mathbf{x})$, $\forall \mathbf{x} \in \mathbb{R}^d$.*

Proof. The proof is constructive, for any $n' > n$ we construct a feature mapping satisfying the conditions. Let $\psi : \mathbb{R}^d \to \mathbb{R}^{n'}$ be given in the following form:

$$\psi_i(\mathbf{x}) = \begin{cases} \phi_i(\mathbf{x}), & i \leq n \\ 0, & i > n. \end{cases} . \tag{3.19}$$

It can be readily seen that ψ is a feature mapping of k, since the value of the inner product $\langle \phi(\mathbf{x}), \phi(\mathbf{y}) \rangle$ equals $\langle \psi(\mathbf{x}), \psi(\mathbf{y}) \rangle$. Moreover, ψ maps into an n' dimensional space, but $\sum_{i=1}^{n'} \psi_i(\mathbf{x}) = \sum_{i=1}^{n} \phi_i(\mathbf{x})$. Therefore, we have:

$$c^\psi(\mathbf{x}) = \frac{\sum_{i=1}^{n'} \psi_i(\mathbf{x})}{\sqrt{n'}} = \frac{\sqrt{n}}{\sqrt{n'}} \frac{\sum_{i=1}^{n} \phi_i(\mathbf{x})}{\sqrt{n}} = \frac{\sqrt{n}}{\sqrt{n'}} c^\phi(\mathbf{x}). \tag{3.20}$$

□

Although we showed that countably infinitely many different centralizer functions exist for a non-zero kernel with a finite dimensional feature mapping, it is still unclear if these different centralizer functions lead to different kernelized (dis)similarity measures. Let ϕ and ψ denote centralizer functions of the kernel k. Two kernelized measures based on ϕ and ψ are considered to be different, if there exist $\mathbf{x}', \mathbf{y}' \in \mathbb{R}^d$, such, that $D^{k,\phi}(\mathbf{x}', \mathbf{y}') \neq D^{k,\psi}(\mathbf{x}', \mathbf{y}')$, or similarly $S^{k,\phi}(\mathbf{x}', \mathbf{y}') \neq S^{k,\psi}(\mathbf{x}', \mathbf{y}')$. In the following theorem we show that under mild conditions the measure cEUC can be kernelized in countably infinite different ways.

Theorem 3.2.2. *Let $k : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ be a kernel with finite dimensional non-zero feature mapping $\phi : \mathbb{R}^d \to \mathbb{R}^n$ for which $k(\mathbf{x}, \mathbf{y}) = \langle \phi(\mathbf{x}), \phi(\mathbf{y}) \rangle$ for any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$. Suppose, there exist $\mathbf{x}', \mathbf{y}' \in \mathbb{R}^d$, such, that*

$$D_{EUC}^{k,\phi}(\mathbf{x}', \mathbf{y}') \neq D_{cEUC}^{k,\phi}(\mathbf{x}', \mathbf{y}'). \tag{3.21}$$

Then, there exist countably infinitely many feature mappings $\{\psi^l\}_l$, $l \in \mathbb{N}$, *such, that for any two* ψ^i, ψ^j, $i, j \in \mathbb{N}$, $i \neq j$ *we have*

$$D^{k,\psi^i}_{cEUC} \not\equiv D^{k,\psi^j}_{cEUC}. \tag{3.22}$$

Proof. The condition $D^{k,\phi}_{EUC}(\mathbf{x}', \mathbf{y}') \neq D^{k,\phi}_{cEUC}(\mathbf{x}', \mathbf{y}')$ implies that there are at least two vectors $\mathbf{x}', \mathbf{y}'$, for which the kernelized centralized and non-centralized measures give different results. What we want to prove is that this property remains true for countably infinitely many feature mappings. Similarly to the construction in the proof of the previous theorem, let $\psi^l : \mathbb{R}^d \to \mathbb{R}^{n+l}$, $l \in \mathbb{N}$ defined by

$$\psi^l_i(\mathbf{x}) = \begin{cases} \phi_i(\mathbf{x}), & i \leq n \\ 0, & i > n. \end{cases} \tag{3.23}$$

Clearly, the functions ψ^l are feature mappings and $c^{\psi^l} = \frac{\sqrt{n}}{\sqrt{n+l}} c^{\phi}$ are centralizer functions of k. Substituting these functions into the formulation of D^k_{cEUC}, we get

$$D^{k,\psi^l}_{cEUC}(\mathbf{x}', \mathbf{y}') = (k(\mathbf{x}', \mathbf{x}') + k(\mathbf{y}', \mathbf{y}') - 2k(\mathbf{x}', \mathbf{y}') - \frac{n}{n+l}(c^{\phi}(\mathbf{x}') - c^{\phi}(\mathbf{y}'))^2)^{\frac{1}{2}}.$$

According to the condition $D^k_{EUC}(\mathbf{x}', \mathbf{y}') \neq D^{k,\phi}_{cEUC}(\mathbf{x}', \mathbf{y}')$, the term $(c^{\phi}(\mathbf{x}') - c^{\phi}(\mathbf{y}'))$ cannot be zero. Consequently, for any $i, j \in \mathbb{N}$, $D^{k,\psi^i}_{cEUC}(\mathbf{x}', \mathbf{y}')$ and $D^{k,\psi^j}_{cEUC}(\mathbf{x}', \mathbf{y}')$ must be different values. □

In the previous theorem we have given some conditions and a construction rule for countable infinitely many ways to kernelize the cEUC measure. In the next theorem we give similar results for the PCC measure.

Theorem 3.2.3. *Let* $k : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ *be a kernel with a finite dimensional non-zero feature mapping* $\phi : \mathbb{R}^d \to \mathbb{R}^n$ *satisfying* $k(\mathbf{x}, \mathbf{y}) = \langle \phi(\mathbf{x}), \phi(\mathbf{y}) \rangle$ *for any* $\mathbf{x}, \mathbf{y} \in \mathbb{R}^m$. *Suppose, there exist* $\mathbf{x}', \mathbf{y}' \in \mathbb{R}^d$, *such, that* $c^{\phi}(\mathbf{x}') \neq c^{\phi}(\mathbf{y}')$, $c^{\phi}(\mathbf{x}') \neq 0$, $c^{\phi}(\mathbf{y}') \neq 0$, $k^2(\mathbf{x}', \mathbf{y}') \neq k(\mathbf{x}', \mathbf{x}')k(\mathbf{y}', \mathbf{y}')$, $D^k_{CC}(\mathbf{x}', \mathbf{y}') < \infty$, $D^{k,\phi}_{PCC}(\mathbf{x}', \mathbf{y}') < \infty$, *and*

$$D^k_{CC}(\mathbf{x}', \mathbf{y}') \neq D^{k,\phi}_{PCC}(\mathbf{x}', \mathbf{y}'). \tag{3.24}$$

Then, there exist countably infinitely many feature mappings $\{\psi^l\}_l$, $l \in \mathbb{N}$ *such, that for any two* ψ^i, ψ^j, $i, j \in \mathbb{N}$, $i \neq j$ *we have*

$$D_{PCC}^{k,\psi^i} \not\equiv D_{PCC}^{k,\psi^j}. \tag{3.25}$$

Proof. Similarly to the proof of the previous theorem, we can construct a sequence of feature mappings and corresponding centralizer functions $\psi^l(\mathbf{x}) = \frac{\sqrt{n}}{\sqrt{n+l}}\phi(\mathbf{x})$. Writing these centralizer functions into the formulation of the S_{PCC}^k measure, one gets

$$S_{PCC}^{k,\psi^l}(\mathbf{x}', \mathbf{y}') = \frac{k(\mathbf{x}', \mathbf{y}') - \frac{n}{n+l}c^\phi(\mathbf{x}')c^\phi(\mathbf{y}')}{(k(\mathbf{x}', \mathbf{x}') - \frac{n}{n+l}c^\phi(\mathbf{x}')c^\phi(\mathbf{x}'))^{\frac{1}{2}}(k(\mathbf{y}', \mathbf{y}') - \frac{n}{n+l}c^\phi(\mathbf{y}')c^\phi(\mathbf{y}'))^{\frac{1}{2}}}. \tag{3.26}$$

Squaring the right hand side of (3.26), introducing the variable $s = \frac{n}{n+l}$ and considering the expression as a function of s, we get

$$r(s) = \frac{\Big(k(\mathbf{x}', \mathbf{y}') - sc^\phi(\mathbf{x}')c^\phi(\mathbf{y}')\Big)^2}{(k(\mathbf{x}', \mathbf{x}') - sc^\phi(\mathbf{x}')c^\phi(\mathbf{x}'))(k(\mathbf{y}', \mathbf{y}') - sc^\phi(\mathbf{y}')c^\phi(\mathbf{y}'))}, \tag{3.27}$$

and expanding the multiplications,

$$r(s) = \frac{s^2c^\phi(\mathbf{x}')^2c^\phi(\mathbf{y}')^2 - s2k(\mathbf{x}', \mathbf{y}')c^\phi(\mathbf{x}')c^\phi(\mathbf{y}') + k(\mathbf{x}', \mathbf{y}')^2}{s^2c^\phi(\mathbf{x}')^2c^\phi(\mathbf{y}')^2 - s(k(\mathbf{x}', \mathbf{x}')c^\phi(\mathbf{y}')^2 + k(\mathbf{y}', \mathbf{y}')c^\phi(\mathbf{x}')^2) + k(\mathbf{x}', \mathbf{x}')k(\mathbf{y}', \mathbf{y}')} \tag{3.28}$$

One can recognize, that $r(s)$ is a rational function of s in the form

$$r(s) = \frac{P(s)}{Q(s)}, \tag{3.29}$$

where P and Q are second order polynomials of s. The coefficients of the highest degree term s^2 are the same in $P(s)$ and $Q(s)$, and according to the assumptions of the theorem, $c^\phi(\mathbf{x}')^2c^\phi(\mathbf{y}')^2$ is not zero, thus, $P(s) \neq 0$ and $Q(s) \neq 0$ holds. On the other hand, due to the assumption $k(\mathbf{x}', \mathbf{y}')^2 \neq k(\mathbf{x}', \mathbf{x}')k(\mathbf{y}, \mathbf{y}')$, the zero order terms are different, thus, we can conclude, that $P(s) \not\equiv Q(s)$. Consequently, $r(s)$ and its derivative have only a finite number of singularities and roots, implying that $S_{PCC}^{k,\psi^l}(\mathbf{x}', \mathbf{y}')$ gives different values for countably infinite $s = \dfrac{n}{n+l}$, $l \in \mathbb{N}$ numbers. □

We close this section with a short summary. In Proposition (3.2.1) we showed that for a kernel with a feature mapping into ℓ^1 the kernelized centered and non-centered dissimilarity measures are the same. In Theorems (3.2.2) and (3.2.3) we showed that if a kernel has a finite dimensional feature mapping, under mild conditions there are countably infinite different ways to calculate the centered kernelized dissimilarity measures. In the next subsections we examine the case of the homogeneous and inhomogeneous polynomial kernels in details.

3.2.2 Homogeneous polynomial kernel

The (inhomogeneous) polynomial kernel defined as

$$k(\mathbf{x}, \mathbf{y}) = (\gamma + \alpha\langle \mathbf{x}, \mathbf{y}\rangle)^q, \tag{3.30}$$

$\gamma, \alpha \in \mathbb{R}$, $q \in \mathbb{N}$, is a widely used kernel in machine learning problems, numerous papers report its successful application in various problems and disciplines (see e.g. [64] for medical research, [65] for face and [61] for speech recognition). The homogeneous polynomial kernel is a special case of the inhomogeneous polynomial kernel at $\gamma = 0$ and $\alpha = 1$, thus, the homogeneous polynomial kernel can be defined as

$$k(\mathbf{x}, \mathbf{y}) = \langle \mathbf{x}, \mathbf{y}\rangle^q. \tag{3.31}$$

Although the inhomogeneous and homogeneous kernels are closely related, we discuss them in separate sections, since it is easier to introduce the main ideas through the simple homogeneous case, and generalize them for the inhomogeneous one, than introducing the complicated formulas for the inhomogeneous kernel from scratch.

First, we construct a feature mapping for the homogeneous polynomial kernel by the heuristics of variable separation. Expanding the inner product and applying the multinomial theorem [66], we get:

$$k(\mathbf{x}, \mathbf{y}) = \langle \mathbf{x}, \mathbf{y}\rangle^q = (\mathbf{x}_1\mathbf{y}_1 + \cdots + \mathbf{x}_m\mathbf{y}_m)^q = \sum_{\boldsymbol{a}_1+\boldsymbol{a}_2+\cdots+\boldsymbol{a}_d=q} \binom{q}{\boldsymbol{a}_1!\boldsymbol{a}_2!\cdots\boldsymbol{a}_d!} (\mathbf{x}_1\mathbf{y}_1)^{\boldsymbol{a}_1}\cdots(\mathbf{x}_d\mathbf{y}_d)^{\boldsymbol{a}_d},$$

where $\boldsymbol{a}_i \geq 0$ for $i = 1, \ldots, d$. For shorter notation, we will use multiindices:

$$\sum_{|\boldsymbol{a}|=q} \binom{q}{\boldsymbol{a}} \mathbf{x}^{\boldsymbol{a}} = \sum_{\boldsymbol{a}_1+\boldsymbol{a}_2+\cdots+\boldsymbol{a}_d=q} \binom{q}{\boldsymbol{a}_1!\boldsymbol{a}_2!\cdots\boldsymbol{a}_d!} \mathbf{x}_1^{\boldsymbol{a}_1} \cdots \mathbf{x}_d^{\boldsymbol{a}_d}$$

with $\boldsymbol{a} = (\boldsymbol{a}_1, \ldots, \boldsymbol{a}_d) \in \mathbb{N}^d$, $|\boldsymbol{a}| = \boldsymbol{a}_1+\boldsymbol{a}_2+\cdots+\boldsymbol{a}_d$, $\mathbf{x} \in \mathbb{R}^d$. By rearranging the equation above, we get:

$$k(\mathbf{x}, \mathbf{y}) = \sum_{|\boldsymbol{a}|=q} \binom{q}{\boldsymbol{a}}^{\frac{1}{2}} \mathbf{x}^{\boldsymbol{a}} \binom{q}{\boldsymbol{a}}^{\frac{1}{2}} \mathbf{y}^{\boldsymbol{a}} = \langle \phi(\mathbf{x}), \phi(\mathbf{y}) \rangle,$$

where

$$\phi_{\boldsymbol{a}}(\mathbf{x}) = \binom{q}{\boldsymbol{a}}^{\frac{1}{2}} \mathbf{x}^{\boldsymbol{a}}. \tag{3.32}$$

For the feature mapping deduced in (3.32), the centralizer function becomes:

$$c^{\phi}(\mathbf{x}) = \frac{\sum\limits_{|\boldsymbol{a}|=q} \phi_{\boldsymbol{a}}(\mathbf{x})}{\sqrt{\dim(\phi)}}, \tag{3.33}$$

where

$$\dim(\phi) = \sum_{i=0}^{q-1} \binom{q-1}{i} \binom{d}{i+1} \tag{3.34}$$

is the dimensionality of the feature mapping, i.e. the number of components after expanding the exponentiation. The feature mapping deduced above can be found in the literature (see e.g. [51]) for the illustration of constructing feature mappings by variable separation. However, as the following results show, more feature mappings can be constructed for the same kernel with the same parameters, and each of them leads to a different centralizer function.

Proposition 3.2.2. *The feature mapping ϕ with coordinates $\phi_{\boldsymbol{a},i} : \mathbb{R}^d \to \mathbb{R}$, $\boldsymbol{a} \in \mathbb{N}^d$, $|\boldsymbol{a}| = q$, $i = 1, \ldots, n_{\boldsymbol{a}}$, $n_{\boldsymbol{a}} \in \mathbb{N}$ defined as*

$$\phi_{\boldsymbol{a},i}(\mathbf{x}) = \beta_{\boldsymbol{a},i} \mathbf{x}^{\boldsymbol{a}}, \ \beta_{\boldsymbol{a},i} \in \mathbb{R}_{>0}, \tag{3.35}$$

is a feature mapping of the kernel $k(\mathbf{x}, \mathbf{y}) = \langle \mathbf{x}, \mathbf{y} \rangle^q$*, if we have:*

$$\sum_{i=1}^{n_{\boldsymbol{a}}} \beta_{\boldsymbol{a},i}^2 = \binom{q}{\boldsymbol{a}} \tag{3.36}$$

for all $\boldsymbol{a} : |\boldsymbol{a}| = q$*.*

Proof. It can be readily seen, that after computing the inner product of the feature maps, we get exactly $k(\mathbf{x}, \mathbf{y})$:

$$\langle \phi(\mathbf{x}), \phi(\mathbf{y}) \rangle = \sum_{|\boldsymbol{a}|=q} \sum_{i=1}^{n_{\boldsymbol{a}}} \beta_{\boldsymbol{a},i} \mathbf{x}^{\boldsymbol{a}} \beta_{\boldsymbol{a},i} \mathbf{y}^{\boldsymbol{a}} = \sum_{|\boldsymbol{a}|=q} \binom{q}{\boldsymbol{a}} \mathbf{x}^{\boldsymbol{a}} \mathbf{y}^{\boldsymbol{a}} = \langle \mathbf{x}, \mathbf{y} \rangle^q. \tag{3.37}$$

□

In other words, the previous proposition says that the only requirement for a function of the form (3.35) to be the feature mapping of the homogeneous polynomial kernel is that the squared sum of coefficients corresponding to a term $\mathbf{x}^{\boldsymbol{a}}$ has to be $\binom{q}{\boldsymbol{a}}$. In Table 3.1 we illustrate the statement for the case $q = 3$, $d = 2$. In column ϕ^a we give the feature mapping arising from the application of the multinomial theorem, and in columns ϕ^b and ϕ^c we give two further feature mappings satisfying the conditions of Proposition 3.2.6. It can be readily seen that each of the feature mappings satisfy the condition $\sum_{i=1}^{n_{\boldsymbol{a}}} \beta_{\boldsymbol{a},i}^2 = \binom{q}{\boldsymbol{a}}$. For example, for the term $\mathbf{x}_1^2 \mathbf{x}_2$, we have $\boldsymbol{a} = (2, 1)$, $\binom{q}{\boldsymbol{a}} = \binom{3}{2!1!} = 3$, correspondingly, for ϕ^a, we have $n_{\boldsymbol{a}} = 1$ and $\beta_{\boldsymbol{a},1}^a = \sqrt{3}$, $\left(\beta_{\boldsymbol{a},1}^a\right)^2 = 3$. In the case of ϕ^b, we have $n_{\boldsymbol{a}} = 2$, $\beta_{\boldsymbol{a},1}^b = \sqrt{2}$, $\beta_{\boldsymbol{a},2}^b = 1$, and $\sum_{i=1}^{n_{\boldsymbol{a}}} \left(\beta_{\boldsymbol{a},i}^b\right)^2 = 3$, finally, in the case of ϕ^c, we have $n_{\boldsymbol{a}} = 3$, $\beta_{\boldsymbol{a},1}^c = \beta_{\boldsymbol{a},2}^c = \beta_{\boldsymbol{a},3}^c = 1$, and again, $\sum_{i=1}^{n_{\boldsymbol{a}}} \left(\beta_{\boldsymbol{a},i}^c\right)^2 = 3$.

In Proposition 3.2.2 we have given a sufficient condition for functions to be finite dimensional feature mappings of the homogeneous polynomial kernel. In the following theorem we show that these feature mappings lead to continuum many different centralizer functions for the homogeneous polynomial kernel.

Table 3.1: Some feature mappings for the homogeneous polynomial kernel. In the last row we give the value of the centraliezer functions for the vector $\mathbf{1}$ to illustrate that the centralizer functions are also different. In the first column the coordinates are indexed by natural numbers, and in the last column the composite indices $\phi_{\boldsymbol{a},i}$, $i = 1, \ldots, n_{\boldsymbol{a}}$ are used

feature mapping	ϕ^a	ϕ^b	ϕ^c	
dimensionality	4	6	8	coordinate
$\phi_1(\mathbf{x})$	$\mathbf{x}_1^3$	$\mathbf{x}_1^3$	$\mathbf{x}_1^3$	$\phi_{(3,0),1}(\mathbf{x})$
$\phi_2(\mathbf{x})$	$\sqrt{3}\mathbf{x}_1^2\mathbf{x}_2$	$\sqrt{2}\mathbf{x}_1^2\mathbf{x}_2$	$\mathbf{x}_1^2\mathbf{x}_2$	$\phi_{(2,1),1}(\mathbf{x})$
$\phi_3(\mathbf{x})$	$\sqrt{3}\mathbf{x}_1\mathbf{x}_2^2$	$\sqrt{2}\mathbf{x}_1\mathbf{x}_2^2$	$\mathbf{x}_1\mathbf{x}_2^2$	$\phi_{(1,2),1}(\mathbf{x})$
$\phi_4(\mathbf{x})$	$\mathbf{x}_2^3$	$\mathbf{x}_2^3$	$\mathbf{x}_2^3$	$\phi_{(0,3),1}(\mathbf{x})$
$\phi_5(\mathbf{x})$		$\mathbf{x}_1^2\mathbf{x}_2$	$\mathbf{x}_1^2\mathbf{x}_2$	$\phi_{(2,1),2}(\mathbf{x})$
$\phi_6(\mathbf{x})$		$\mathbf{x}_1\mathbf{x}_2^2$	$\mathbf{x}_1\mathbf{x}_2^2$	$\phi_{(1,2),2}(\mathbf{x})$
$\phi_7(\mathbf{x})$			$\mathbf{x}_1^2\mathbf{x}_2$	$\phi_{(2,1),3}(\mathbf{x})$
$\phi_8(\mathbf{x})$			$\mathbf{x}_1\mathbf{x}_2^2$	$\phi_{(1,2),3}(\mathbf{x})$
$c^{\phi}(\mathbf{1})$	2.73	2.78	2.82	

Theorem 3.2.4. *Let $k : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$, $d \geq 2$ denote the homogeneous polynomial kernel $k(\mathbf{x}, \mathbf{y}) = \langle \mathbf{x}, \mathbf{y} \rangle^q$. There are continuum many different centralizer functions of k.*

Proof. Without the loss of generality, let ϕ denote a feature mapping of the homogeneous polynomial kernel $k = \langle \mathbf{x}, \mathbf{y} \rangle^q$, in which $n_{\boldsymbol{a}} = 1$, $\beta_{\boldsymbol{a},1} = \sqrt{\binom{q}{\boldsymbol{a}}}$ for all $|\boldsymbol{a}| = q$, except $\boldsymbol{a}' = (q, 0, 0, \ldots, 0)$. For $\boldsymbol{a}'$ we have $n_{\boldsymbol{a}'} = 2$, and

$$\beta_{\boldsymbol{a}',1}^2 + \beta_{\boldsymbol{a}',2}^2 = \binom{q}{\boldsymbol{a}'} = \binom{q}{\boldsymbol{a}_1'! \boldsymbol{a}_2'! \ldots \boldsymbol{a}_d'!} = \frac{q!}{q!} = 1. \tag{3.38}$$

For the dimensionality of ϕ we have $\dim(\phi) = 1 + \sum_{i=0}^{q-1} \binom{q-1}{i} \binom{d}{i+1}$. In other words, ϕ is a feature mapping we constructed at the beginning of the section (the multinomial coefficients are not decomposed to several

coordinates), except the multinomial coefficient corresponding to $\boldsymbol{a}'$, which is decomposed to two coordinates with coefficients $\beta_{\boldsymbol{a}',1}$ and $\beta_{\boldsymbol{a}',2}$. One can readily see that this latter condition can be satisfied by continuum many different configurations of $\beta_{\boldsymbol{a}',1}$ and $\beta_{\boldsymbol{a}',2}$, since the condition is basically the equation of a hypersphere of non-zero radius $\sqrt{\binom{q}{\boldsymbol{a}'}}$, which has continuum many points $(\beta_{\boldsymbol{a}',1}, \beta_{\boldsymbol{a}',2})$. Thus, in the feature mapping ϕ the parameters $\beta_{\boldsymbol{a}',1}$, $\beta_{\boldsymbol{a}',2}$ are not determined totally, one can construct continuum many different feature mappings differing only in $\beta_{\boldsymbol{a}',1}$ and $\beta_{\boldsymbol{a}',2}$. We want to show that there is at least one vector $\mathbf{x}$ for which the centralizer functions corresponding these continuum many feature mappings give continuum many different values. Let $\mathbf{x} = (1, 0, 0, \ldots, 0)$. One can readily see that for the feature mapping defined before, we have $c^{\phi}(\mathbf{x}) = \dfrac{\beta_{\boldsymbol{a}',1} + \beta_{\boldsymbol{a}',2}}{\sqrt{\dim(\phi)}}$. Considering the latter expression as the function of $\beta_{\boldsymbol{a}',1}$ and $\beta_{\boldsymbol{a}',2}$, one can readily see that it is the implicit equation of a line in the 2D Euclidean space: $\beta_{\boldsymbol{a}',1} + \beta_{\boldsymbol{a}',2} = \omega$, $\omega = c^{\phi}(\mathbf{x})\sqrt{\dim(\phi)}$. We want to show that for continuum many different ω values one can find $\beta_{\boldsymbol{a}',1}, \beta_{\boldsymbol{a}',2} \in \mathbb{R}$, satisfying $\beta_{\boldsymbol{a}',1} + \beta_{\boldsymbol{a}',2} = \omega$ as well as $\beta^2_{\boldsymbol{a}',1} + \beta^2_{\boldsymbol{a}',2} = 1$. In other words, we want to show that there are continuum many parallel lines intersecting the circle of radius 1, which is trivially true. □

For the homogeneous polynomial kernel we showed that the number of different centralizer functions is continuum. However, it is still unclear if these continuum many centralizer functions lead to different (dis)similarity measures.

Theorem 3.2.5. *Let $k : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$, $d \geq 2$ denote the homogeneous polynomial kernel $k(\mathbf{x}, \mathbf{y}) = \langle \mathbf{x}, \mathbf{y} \rangle^q$. There are continuum many different ways to kernelize the cEUC and PCC measures using the kernel k.*

Proof. The statement of the theorem is proved, if we manage to find two vectors $\mathbf{x}', \mathbf{y}' \in \mathbb{R}^d$ and continuum many feature mappings ϕ^{ω}, $\omega \in \Omega$ such, that $S^{k,\phi^{\omega}}_{PCC}(\mathbf{x}', \mathbf{y}')$ and $D^{k,\phi^{\omega}}_{cEUC}(\mathbf{x}', \mathbf{y}')$ give continuum many different values for $\omega \in \Omega$. The proof is based on the construction we used in the proof of the previous theorem.

Particularly, let $\mathbf{x}' = (1, 0, 0, \ldots, 0)$ and $\mathbf{y}' = (0, 1, 0, \ldots, 0)$, and let ϕ^{ω}, $\omega \in \Omega \subset \mathbb{R}$ denote a feature mapping of k such, that except the coefficient

corresponding to $\boldsymbol{a}' = (q, 0, 0, \ldots, 0)$, the multinomial coefficients $\binom{q}{a}$ are not decomposed. The coefficient corresponding to $\boldsymbol{a}' = (q, 0, 0, \ldots, 0)$ is decomposed to $n_{\boldsymbol{a}'} = 2$ coordinates with coefficients $\beta_{\boldsymbol{a}',1}$ and $\beta_{\boldsymbol{a}',2}$, satisfying $\beta^2_{\boldsymbol{a}',1} + \beta^2_{\boldsymbol{a}',2} = 1$, $\beta_{\boldsymbol{a}',1} + \beta_{\boldsymbol{a}',2} = \omega$, $\omega \in \Omega$, where Ω is a set of continuum cardinality according to the construction in the proof of the previous theorem. Obviously, each function ϕ^ω is a feature mapping of k. First we focus on the kernelized cEUC measure. Substituting $\mathbf{x}'$ and $\mathbf{y}'$ into the (3.12) and utilizing that $k(\mathbf{x}', \mathbf{x}') = 1$, $k(\mathbf{y}', \mathbf{y}') = 1$, $k(\mathbf{x}', \mathbf{y}') = 0$, $c^{\phi^\omega}(\mathbf{x}') = \dfrac{\omega}{\sqrt{\dim(\phi^\omega)}}$ and $c^{\phi^\omega}(\mathbf{y}') = \dfrac{1}{\sqrt{\dim(\phi^\omega)}}$, we get

$$D^{k,\phi^\omega}_{cEUC}(\mathbf{x}', \mathbf{y}') = \left(2 + \frac{1}{\dim(\phi^\omega)}(\omega - 1)^2\right)^{\frac{1}{2}}, \tag{3.39}$$

and it can be readily seen that this expression is a continuous function of ω, giving continuum many different values for the centralized dissimilarity $D^{k,\phi^\omega}_{cEUC}(\mathbf{x}', \mathbf{y}')$ of $\mathbf{x}'$ and $\mathbf{y}'$ as $\omega \in \Omega$ varies. Thus, there are continuum many ways to centralize the measure cEUC. Similarly, we can substitute $\mathbf{x}'$ and $\mathbf{y}'$ into the kernelized PCC measure (3.10):

$$S^{k,\phi^\omega}_{PCC}(\mathbf{x}', \mathbf{y}') = \frac{-\omega}{(\dim(\phi^\omega)^2 + \omega^2(1 + \dim(\phi^\omega)) - \dim(\phi^\omega))^{\frac{1}{2}}}, \tag{3.40}$$

again, considering $S^{k,\phi^\omega}_{PCC}(\mathbf{x}', \mathbf{y}')$ as the function of ω, it can be readily seen that we have a non-constant function having only a finite number of singularities and roots, thus, it gives continuum many different values as $\omega \in \Omega$ varies. □

The previous theorem shows that there are continuum many ways to kernelize the centralized dissimilarities. In order to make centralization in kernel space useful in practice, we have to find centralizer functions which can be plugged into the formulations (3.17) and (3.18) efficiently: it is desired to find centralizer functions that can be computed in terms of kernel evaluations. As the following theorem shows, there are countably infinite different such centralizer functions for the homogeneous polynomial kernel, and they lead to countably infinite different kernelized cEUC and PCC measures.

Theorem 3.2.6. *The function*

$$c(\mathbf{x}) = \frac{\sqrt{d^q}}{\sqrt{n}} k(\mathbf{x}, \mathbf{1}) \tag{3.41}$$

is a centralizer function of the kernel $k : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$, $k(\mathbf{x}, \mathbf{y}) := \langle \mathbf{x}, \mathbf{y} \rangle^q$ *for all* $n \geq d^q$, $n \in \mathbb{N}$, *and leads to countably infinitely many different kernelized cEUC and PCC measures.*

Proof. For $n = d^q$ a possible feature mapping belonging to c is the function $\phi : \mathbb{R}^d \to \mathbb{R}^{d^q}$, $\phi_{\boldsymbol{a},i}(\mathbf{x}) = \mathbf{x}^{\boldsymbol{a}}$ for all $i = 1, \ldots, \binom{q}{\boldsymbol{a}}$, $|\boldsymbol{a}| = q$, $\boldsymbol{a} \in \mathbb{N}^d$, particularly, all $\beta_{\boldsymbol{a},i}$ coordinates are equal to 1. An illustration for this feature mapping can be found in column ϕ^c of Table 3.1 for the $q = 3$, $d = 2$ case.

To prove the second part of the statement, one can observe that the structure of (3.41) is similar to the one we used in Theorems (3.2.2) and (3.2.3). Since the feature mapping ϕ we just constructed in the $n = d^q$ case is finite dimensional, all we have to do is to find two vectors $\mathbf{x}'$ and $\mathbf{y}'$ satisfying the conditions of Theorems (3.2.2) and (3.2.3). First, we focus on the cEUC measure. Let $\mathbf{x}' = (1, 0, 0, \ldots, 0)$ and $\mathbf{y}' = (0, 0, 0, \ldots, 0)$. Then, we have $k(\mathbf{x}', \mathbf{y}') = 0$, $k(\mathbf{x}', \mathbf{x}') = 1$, $k(\mathbf{y}', \mathbf{y}') = 0$, $c^\phi(\mathbf{x}') = \frac{1}{\sqrt{d^q}}$ and $c^\phi(\mathbf{y}') = 0$, thus

$$D^{k,\phi}_{cEUC}(\mathbf{x}', \mathbf{y}') = \left(1 + \left(\frac{1}{d^q}\right)^2\right)^{\frac{1}{2}} \neq 1 = D^k_{EUC}(\mathbf{x}', \mathbf{y}'). \tag{3.42}$$

The conditions of Theorem (3.2.2) are satisfied, and we can apply Theorem (3.2.2) to construct countably different centralizer functions in the form (3.41). For the PCC measure let $\mathbf{x}'$ be the same as before, but $\mathbf{y}' = (0, 2, 0, \ldots, 0)$. Now, $k(\mathbf{x}', \mathbf{y}') = 0$, $k(\mathbf{y}', \mathbf{y}') = 2^{2d}$, $c^\phi(\mathbf{y}') = \frac{2^d}{\sqrt{d^q}}$, and we have

$$S^{k,\phi}_{PCC}(\mathbf{x}', \mathbf{y}') = \frac{-\frac{2^d}{d^q}}{\left(\left(1 - \frac{1}{d^q}\right)\left(1 - \frac{2^{2d}}{d^q}\right)\right)^{\frac{1}{2}}} \neq 0 = S^{k,\phi}_{CC}(\mathbf{x}', \mathbf{y}'). \tag{3.43}$$

Again, we can apply Theorem (3.2.3) to construct countably infinitely many different feature mappings in the form (3.41). □

3.2.3 Inhomogeneous polynomial kernel

In the rest of the chapter we give similar results for the inhomogeneous polynoial kernel defined by

$$k(\mathbf{x}, \mathbf{y}) = (\gamma + \alpha\langle \mathbf{x}, \mathbf{y}\rangle)^q, \tag{3.44}$$

$q \in \mathbb{N}$, $\alpha \in \mathbb{R}$, $\gamma \in \mathbb{R}$, as we did in the previous section for the homogeneous polynomial kernel. First, we construct a feature mapping by the separation of variables. Expanding the inner product and applying the multinomial theorem with multiindices, we get:

$$k(\mathbf{x}, \mathbf{y}) = (\gamma + \alpha\langle \mathbf{x}, \mathbf{y}\rangle)^q = (\gamma + \alpha\mathbf{x}_1\mathbf{y}_1 + \cdots + \alpha\mathbf{x}_m\mathbf{y}_m)^q = \sum_{|\boldsymbol{a}|\leq d} \alpha^{|\boldsymbol{a}|}\gamma^{q-|\boldsymbol{a}|}\binom{q}{\boldsymbol{a}}\mathbf{x}^{\boldsymbol{a}}\mathbf{y}^{\boldsymbol{a}}. \tag{3.45}$$

Comparing with the homogeneous case, one can observe, that in (3.45) there can be terms in the sum consisting of less than q variable components, and there are also some tunable parameters α and γ in the expansion. After variable separation, we get a feature mapping with $\dim(\phi) = \sum_{j=1}^{q-1}\sum_{i=0}^{j}\binom{j}{i}\binom{q}{i+1}$ coordinates:

$$\phi : \mathbb{R}^d \to \mathbb{R}^{\dim(\phi)} \text{ with} \tag{3.46}$$

$$\phi_{\boldsymbol{a}}(x) = \binom{q}{\boldsymbol{a}}^{\frac{1}{2}} \alpha^{\frac{|\boldsymbol{a}|}{2}}\gamma^{\frac{q-|\boldsymbol{a}|}{2}} x^{\boldsymbol{a}}. \tag{3.47}$$

The centralizer function of the feature mapping based on the multinomial expansion becomes:

$$c^{\phi}(\mathbf{x}) = \frac{\sum_{|\boldsymbol{a}|\leq q}\phi_{\boldsymbol{a}}(\mathbf{x})}{\sqrt{\dim(\phi)}}. \tag{3.48}$$

In the first proposition we give some general conditions on the finite dimensional feature mappings of the inhomogeneous polynomial kernel.

Proposition 3.2.3. *The feature mapping ϕ with coordinates $\phi_{a,i} : \mathbb{R}^d \to \mathbb{R}$, $|\boldsymbol{a}| \leq q$, $i = 1, \ldots, n_{\boldsymbol{a}}$, $n_{\boldsymbol{a}} \in \mathbb{N}$ defined as*

$$\phi_{\boldsymbol{a},i}(\mathbf{x}) = \beta_{\boldsymbol{a},i}\mathbf{x}^{\boldsymbol{a}},\ \beta_{\boldsymbol{a},i} \in \mathbb{R}, \tag{3.49}$$

is a feature mapping of the kernel $k(\mathbf{x}, \mathbf{y}) = (\gamma + \alpha\langle \mathbf{x}, \mathbf{y}\rangle)^q$*, if*

$$\sum_{i=1}^{n_{\boldsymbol{a}}} \beta_{\boldsymbol{a},i}^2 = \binom{q}{\boldsymbol{a}} \alpha^{|\boldsymbol{a}|} \gamma^{q-|\boldsymbol{a}|} \tag{3.50}$$

holds for all $\boldsymbol{a} : |\boldsymbol{a}| \leq d$*.*

Proof. Similarly to the proof of Theorem (3.2.2), after computing the inner product of two mappings $\phi(\mathbf{x})$ and $\phi(\mathbf{y})$, we get the multinomial expansion of the inhomogeneous polynomial kernel. □

For illustration of some feature mappings see Table 3.2: the column of ϕ^a contains the feature mapping arising from the application of the multinomial theorem, ϕ^b and ϕ^c is a demonstration of Proposition 3.2.3. As before, one can readily check that the coefficients in ϕ^a, ϕ^b and ϕ^c satisfy the conditions of Proposition 3.2.3. Again, in the last row we have given the response of the centralizer functions for the vector $\mathbf{1}$. As one can see, not only the feature mappings, but the centralizer functions are also different.

The following theorem states that there are continuum many different centralizer functions for the inhomogeneous polynomial kernel.

Theorem 3.2.7. *Let* $k : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$*,* $d \geq 2$ *denote the inhomogeneous polynomial kernel* $k(\mathbf{x}, \mathbf{y}) = (\gamma + \alpha\langle \mathbf{x}, \mathbf{y}\rangle)^q$*. There are continuum many different centralizer functions of* k*.*

Proof. The proof is analogous to the proof of Theorem 3.2.4. □

Theorem 3.2.8. *Let* $k : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$*,* $d \geq 2$ *denote the inhomogeneous polynomial kernel* $k(\mathbf{x}, \mathbf{y}) = (\gamma + \alpha\langle \mathbf{x}, \mathbf{y}\rangle)^q$*. There are continuum many different ways to kernelize the cEUC and PCC measures using the kernel* k*.*

Proof. The proof is analogous to the proof of Theorem 3.2.5. □

Finally, we show that out of the continuum many centralizer functions, there exist a countably infinite number of centralizer functions that can be computed in terms of kernel evaluations.

Table 3.2: Feature mappings for the inhomogeneous polynomial kernel with $q = 2$, $d = 2$. In the first column the coordinates are indexed by natural numbers, and in the last column the composite indices $\phi_{\boldsymbol{a},i}$, $i = 1, \ldots, n_{\boldsymbol{a}}$ are used

feature mapping	ϕ^a	ϕ^b	ϕ^c	
dimensionality	6	8	9	coordinates
$\phi_1(\mathbf{x})$	$\sqrt{\gamma^2}$	$\sqrt{\gamma^2}$	$\sqrt{\gamma^2}$	$\phi_{(0,0),1}(\mathbf{x})$
$\phi_2(\mathbf{x})$	$\sqrt{2\gamma\alpha}\mathbf{x}_1$	$\sqrt{\gamma\alpha}\mathbf{x}_1$	$\sqrt{\gamma\alpha}\mathbf{x}_1$	$\phi_{(1,0),1}(\mathbf{x})$
$\phi_3(\mathbf{x})$	$\sqrt{2\gamma\alpha}\mathbf{x}_2$	$\sqrt{\gamma\alpha}\mathbf{x}_2$	$\sqrt{\gamma\alpha}\mathbf{x}_2$	$\phi_{(0,1),1}(\mathbf{x})$
$\phi_4(\mathbf{x})$	$\sqrt{\alpha^2}\mathbf{x}_1^2$	$\sqrt{\alpha^2}\mathbf{x}_1^2$	$\sqrt{\alpha^2}\mathbf{x}_1^2$	$\phi_{(2,0),1}(\mathbf{x})$
$\phi_5(\mathbf{x})$	$\sqrt{\alpha^2}\mathbf{x}_2^2$	$\sqrt{\alpha^2}\mathbf{x}_2^2$	$\sqrt{\alpha^2}\mathbf{x}_2^2$	$\phi_{(0,2),1}(\mathbf{x})$
$\phi_6(\mathbf{x})$	$\sqrt{2\alpha^2}\mathbf{x}_1\mathbf{x}_2$	$\sqrt{2\alpha^2}\mathbf{x}_1\mathbf{x}_2$	$\sqrt{\alpha^2}\mathbf{x}_1\mathbf{x}_2$	$\phi_{(1,1),1}(\mathbf{x})$
$\phi_7(\mathbf{x})$		$\sqrt{\gamma\alpha}\mathbf{x}_1$	$\sqrt{\gamma\alpha}\mathbf{x}_1$	$\phi_{(1,0),2}(\mathbf{x})$
$\phi_8(\mathbf{x})$		$\sqrt{\gamma\alpha}\mathbf{x}_2$	$\sqrt{\gamma\alpha}\mathbf{x}_2$	$\phi_{(0,1),2}(\mathbf{x})$
$\phi_9(\mathbf{x})$			$\sqrt{\alpha^2}\mathbf{x}_1\mathbf{x}_2$	$\phi_{(1,1),2}(\mathbf{x})$
$c^{\phi}(\mathbf{1})$	7.82	7.79	7.93	

Proposition 3.2.4. *The function $c : \mathbb{R}^d \to \mathbb{R}$,*

$$c(\mathbf{x}) = \frac{\sqrt{(d+1)^q}}{\sqrt{n}} k(\mathbf{x}, 1) \tag{3.51}$$

is a centralizer function of the kernel $k : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$, $k(\mathbf{x}, \mathbf{y}) := (\gamma + \alpha\langle \mathbf{x}, \mathbf{y} \rangle)^q$, $\gamma \in \mathbb{R}_{>0}$, for all $n \geq (d+1)^q$, $n \in \mathbb{N}$, and these centralizer functions lead to countably infinitely many different kernelized cEUC and PCC measures.

Proof. Similarly to Theorem 3.2.6, for $n = (d+1)^q$, the feature mapping belonging to c is the function $\phi : \mathbb{R}^d \to \mathbb{R}^{(d+1)^q}$, $\phi_{\boldsymbol{a},i}(\mathbf{x}) = \alpha^{|\boldsymbol{a}|}\gamma^{q-|\boldsymbol{a}|}\mathbf{x}^{\boldsymbol{a}}$, for all $i = 1, \ldots, \binom{q}{\boldsymbol{a}}$, $|\boldsymbol{a}| \leq q$, $\boldsymbol{a} \in \mathbb{N}^d$. For $n > (d+1)^q$, analogously to the proof of Theorem 3.2.6, one can easily show that the conditions of Theorems 3.2.2 and 3.2.3 hold for some specific $\mathbf{x}', \mathbf{y}' \in \mathbb{R}^d$ vectors. □

3.3 Tests and Results

In the previous section, we have deduced centralizer functions for the homogeneous and inhomogeneous polynomial kernels, which can be used to centralize the centered Euclidean distance and the Pearson correlation coefficient in the polynomial kernel space, and we also derived centralizer functions which can be computed by kernel evaluations. In order to evaluate the effect of centralization in kernel space, we have carried out several tests in k *Nearest Neighbor* (kNN) classification scenarios using dissimilarity measures based on the previous derivations:

$$D^k_{PCC}(\mathbf{x},\mathbf{y}) = 1 - S^k_{PCC}(\mathbf{x},\mathbf{y}) = \frac{k(\mathbf{x},\mathbf{y}) - c^\phi(\mathbf{x})c^\phi(\mathbf{y})}{(k(\mathbf{x},\mathbf{x}) - c^\phi(\mathbf{x})^2)(k(\mathbf{y},\mathbf{y}) - c^\phi(\mathbf{y})^2)}, \tag{3.52}$$

$$D^k_{cEUC}(\mathbf{x},\mathbf{y}) = k(\mathbf{x},\mathbf{x}) + k(\mathbf{y},\mathbf{y}) - 2k(\mathbf{x},\mathbf{y}) - (c^\phi(\mathbf{x}) - c^\phi(\mathbf{y}))^2, \tag{3.53}$$

respectively, where ϕ is a feature mapping of k. The kernels we used and the corresponding centralizer functions are summarized in Table 3.3: c_i refers to centralization in input space, when no kernelization is applied; c_{hm} denotes centralization in the homogeneous polynomial kernel space for the feature mapping achieved by the multinomial theorem; c_{hk} refers to centralization by kernel evaluations for the homogeneous polynomial kernel. Similarly, c_{im} denotes the centralization function for inhomogeneous polynomial kernels achieved by the multinomial theorem; and c_{ik} refers to centralization in the inhomogeneous kernel space by kernel evaluations.

In the classification scenarios we used own data and some of the commonly cited databases from the UCI repository [67], see Table 3.4 for the description of the datasets. The classification accuracy is measured by leave-one-out cross validation. The kernels as well as the kNN classifier have some free parameters. The proper selection of these parameters (called model selection in the literature) is a crucial point since the parameters may affect the performance of classification heavily. To overcome the problem of model selection, in each test scenario and for each dissimilarity measure, we run grid search in a reasonable part of the parameter space to find the best configuration. The grid of parameters was constructed by the combinations of $k \in \{1,3,5,7\}$, $\alpha \in \{0.2, 0.5, 0.8, 1, 2, 3, 4\}$, $\gamma \in \{-3,-2,-1,0,1,2,3\}$ and $q \in \{1,2,3,4,5\}$. For each combination of kernel, centralizer and

Table 3.3: The centralizer functions used in the empirical analysis

kernel	centralizer
$\langle \mathbf{x}, \mathbf{y} \rangle$	$c_i(\mathbf{x}) = \frac{\sum_{i=1}^{d} \mathbf{x}_i}{\sqrt{d}}$
$\langle \mathbf{x}, \mathbf{y} \rangle^q$	$c_{hm}(\mathbf{x}) = \frac{\sum_{\lvert \boldsymbol{a} \rvert = q} \binom{q}{\boldsymbol{a}}^{\frac{1}{2}} \mathbf{x}^{a}}{\left(\sum_{i=0}^{q-1} \binom{q-1}{i} \binom{d}{i+1} \right)^{\frac{1}{2}}}$
$\langle \mathbf{x}, \mathbf{y} \rangle^q$	$c_{hk}(\mathbf{x}) = \frac{\langle \mathbf{x}, \mathbf{1} \rangle^q}{d^{\frac{q}{2}}}$
$(\gamma + \alpha \langle \mathbf{x}, \mathbf{y} \rangle)^q$	$c_{im}(\mathbf{x}) = \frac{\sum_{\lvert \boldsymbol{a} \rvert \leq q} \binom{q}{\boldsymbol{a}}^{\frac{1}{2}} \alpha^{\frac{\lvert \boldsymbol{a} \rvert}{2}} \gamma^{\frac{d - \lvert \boldsymbol{a} \rvert}{2}} \mathbf{x}^{\boldsymbol{a}}}{\left(\sum_{j=1}^{q-1} \sum_{i=0}^{j} \binom{j}{i} \binom{q}{i+1} \right)^{\frac{1}{2}}}$
$(\gamma + \alpha \langle \mathbf{x}, \mathbf{y} \rangle)^q$	$c_{ik}(\mathbf{x}) = \frac{\langle \mathbf{x}, \mathbf{1} \rangle^q}{(d+1)^{\frac{q}{2}}}$

Table 3.4: Details of the test databases

	training vectors	attributes		classes
		rational	integer	
IRIS (UCI)	150	4	0	3
Liver (UCI)	346	0	6	2
Dermatology (UCI)	358	0	34	6
Microaneurysm (own)	666	5	0	2
Artificial data (own)	200	8	0	2

dissimilarity measure, the accuracy score corresponding to the best model is reported only. For the sake of completeness, the parameters of the best models are also given.

In the following subsections we give the quantitative results for four datasets from the publicly available UCI repository [67]. Then, the results of microaneurysm recognition are presented, and finally, the classification results of an artificial dataset are discussed.

3.3.1 UCI datasets

3.3.1.1 IRIS database

The classification results corresponding to the various kernels, centralizer and dissimilarity functions are summarized in Table 3.5. One can see that centralization in kernel space does not improve the accuracies achieved by correlation based measures, however, the centralization of the kernelized Euclidean distance improves the accuracy scores. Interestingly, comparing the performance of the kernelized and non-kernelized measures, one can see that kernelization improves the accuracy scores in each case.

3.3.1.2 Liver database

The corresponding results are presented in Table 3.6 in a similar way as in Table 3.5. For this dataset the figures are very demonstrative: adding translation invariance in the input space decreases the accuracy for both correlation and Euclidean distance, on the other hand, adding translation invariance in kernel space increases it in both cases. Moreover, kernelized

Table 3.5: Results for the IRIS database

kernel	centralizer	dissim.	acc.	α	γ	q	k
	none	cEUC	0.9666	-	-	-	5
		PCC	0.9800	-	-	-	3
	c_i	cEUC	0.9733	-	-	-	7
		PCC	0.9733	-	-	-	3
$\langle \mathbf{x}, \mathbf{y} \rangle^q$	none	cEUC	0.9666	-	-	1	5
		PCC	0.9800	-	-	2	3
	c_i	cEUC	0.9733	-	-	2	7
		PCC	0.9733	-	-	2	3
	c_{hm}	cEUC	0.9800	-	-	1	7
		PCC	0.9800	-	-	3	3
	c_{hk}	cEUC	0.9800	-	-	2	1
		PCC	**0.9866**	-	-	4	3
$(\gamma + \alpha \langle \mathbf{x}, \mathbf{y} \rangle)^q$	none	cEUC	0.9733	0.2	2	2	7
		PCC	0.9866	0.8	1	2	3
	c_i	cEUC	0.9733	3	1	3	7
		PCC	0.9733	3	0	5	5
	c_{im}	cEUC	**0.9866**	1	1	2	3
		PCC	**0.9866**	1	0	1	7
	c_{ik}	cEUC	**0.9866**	1	2	1	7
		PCC	**0.9866**	0.5	2	2	3

PCC gives the highest accuracy with the centralizer function based on fast kernel evaluations.

3.3.1.3 Dermatology database

The results for this dataset are summarized in Table 3.7. Except the case of the homogeneous polynomial kernel with multinomial centralization, translation invariance does not decrease accuracy. Again, the best results are achieved by PCC in kernel space using the proposed kernel evaluation based centralization.

3.3.2 Microaneurysm database

In this subsection the test results achieved on image data are presented and discussed. The proposed techniques are applied to distinguish microaneurysms and background tissue on fundus images, which is a key problem in the diagnosis of diabetic retinopathy [68]. To have an impression for the appearance of this lesion, see Figure 3.1. In the experiments we have

Table 3.6: Results for the Liver database

kernel	centralizer	dissim.	acc.	α	γ	q	k
	none	cEUC	0.6878	-	-	-	7
		PCC	0.6907	-	-	-	7
	c_i	cEUC	0.6791	-	-	-	5
		PCC	0.6734	-	-	-	7
$\langle \mathbf{x}, \mathbf{y} \rangle^q$	none	cEUC	0.6878	-	-	1	7
		PCC	0.6907	-	-	2	7
	c_i	cEUC	0.6791	-	-	1	5
		PCC	0.6734	-	-	2	7
	c_{hm}	cEUC	0.6907	-	-	2	5
		PCC	0.6965	-	-	2	7
	c_{hk}	cEUC	0.6791	-	-	3	7
		PCC	0.6820	-	-	3	7
$(\gamma + \alpha \langle \mathbf{x}, \mathbf{y} \rangle)^q$	none	cEUC	0.6878	2	2	1	7
		PCC	0.6907	2	2	2	7
	c_i	cEUC	0.6791	2	2	1	5
		PCC	0.6763	0.2	-2	2	7
	c_{im}	cEUC	0.6907	2	2	1	7
		PCC	**0.6994**	0.8	2	3	7
	c_{ik}	cEUC	0.6907	0.5	2	2	7
		PCC	**0.6994**	1	2	5	7

used the publicly available database of the Retinopathy Online Challenge (ROC) [69], which contains manually selected microaneurysms in 50 images. Unlike in the UCI classification problems, for the ROC dataset we have to define the feature vectors on our own. We have considered 5 local features to characterize the local neighborhood of microaneurysms: the features at pixel $\mathbf{p}$ are the lowest intensity values in rings around $\mathbf{p}$ with inner and outer radii $(0, 2)$, $(2, 4)$, $(4, 6)$, $(6, 8)$, $(8, 10)$, respectively. These features can amplify the most conspicuous properties of microaenurysms: since the center of a microaneurysm is relatively dark, one can expect that the first feature has the lowest value, and values of the features increase with the radii of the regions characterized by them. On the contrary, background regions have quasi-constant feature vectors, while other anatomical parts (like vessels) are expected to have feature vectors without systematic structure. Based on the 5 dimensional feature vectors we have prepared a classification dataset of 333 microaneurysms and 333 randomly selected background regions.

Caused by various acquisition circumstances, the brightness of retinal

Table 3.7: Results for the Dermatology database

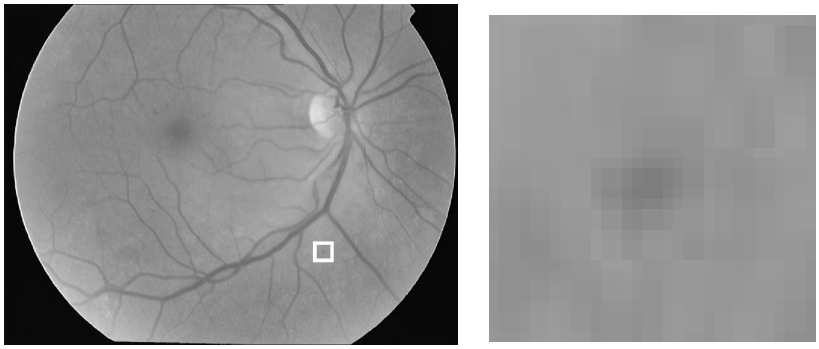
kernel	centralizer	dissim.	acc.	α	γ	q	k
	none	cEUC	0.9022	-	-	-	3
		PCC	0.9664	-	-	-	1
	c_i	cEUC	0.9078	-	-	-	1
		PCC	0.9692	-	-	-	1
$\langle \mathbf{x}, \mathbf{y} \rangle^q$	none	cEUC	0.9022	-	-	1	3
		PCC	0.9664	-	-	2	7
	c_i	cEUC	0.9078	-	-	1	1
		PCC	0.9692	-	-	2	1
	c_{hm}	cEUC	0.8994	-	-	1	1
		PCC	0.9664	-	-	2	1
	c_{hk}	cEUC	0.9078	-	-	1	1
		PCC	0.9692	-	-	1	1
$(\gamma + \alpha \langle \mathbf{x}, \mathbf{y} \rangle)^q$	none	cEUC	0.9022	2	2	1	3
		PCC	0.9664	2	2	2	1
	c_i	cEUC	0.9078	2	2	1	1
		PCC	0.9692	2	2	2	1
	c_{im}	cEUC	0.9022	1	2	2	1
		PCC	**0.9692**	4	2	1	1
	c_{ik}	cEUC	0.9078	1	2	1	1
		PCC	**0.9692**	0.2	3	1	7

images varies. Therefore, computing simply the Euclidean-distance of the input vectors is not sufficient for accurate classification. One can expect that the normalization and/or centralization of the vectors may enhance the classification results.

We have carried out the same tests as in the previous sections and summarized the results in Table 3.8. As the results show, the highest scores are achieved by using centralization in the homogeneous polynomial kernel space. The reason for this behaviour is that – due to the product features of polynomial kernelization – the inhomogeneities are amplified, while centralization makes the measurement independent from the brightness of the images.

3.3.3 Artificial data

In order to measure the performance of the proposed techniques as the function of noise, we carried out the classification of alternating and homogeneous patterns. Based on the vectors $\mathbf{x} = (1, 2, 1, 2, 1, 2, 1, 2)$ and



(a) (b)

Figure 3.1: A standard fundus image with a microaneurysm in the middle of the black circle marker (a); the microaneurysm zoomed-in (b)

Table 3.8: Results for the Microaneurysm database

kernel	centralizer	dissim.	acc.	α	γ	q	k
	none	cEUC	0.7417	-	-	-	7
		PCC	0.7717	-	-	-	7
	c_i	cEUC	0.7732	-	-	-	7
		PCC	0.6141	-	-	-	5
$\langle \mathbf{x}, \mathbf{y} \rangle^q$	none	cEUC	0.7462	-	-	3	3
		PCC	0.7717	-	-	2	7
	c_i	cEUC	0.7852	-	-	5	7
		PCC	0.6141	-	-	2	5
	c_{hm}	cEUC	0.7417	-	-	3	3
		PCC	0.7882	-	-	3	7
	c_{hk}	cEUC	**0.8048**	-	-	5	5
		PCC	0.6816	-	-	5	7
$(\gamma + \alpha \langle \mathbf{x}, \mathbf{y} \rangle)^q$	none	cEUC	0.7462	3	3	3	3
		PCC	0.7777	0.5	3	2	7
	c_i	cEUC	0.7942	0.8	-2	5	7
		PCC	0.7957	0.5	3	2	7
	c_{im}	cEUC	0.7492	0.8	3	3	5
		PCC	0.7822	0.8	3	2	7
	c_{ik}	cEUC	0.7552	1	2	1	5
		PCC	0.7792	0.2	2	3	7

$\mathbf{y} = (2, 2, 2, 2, 2, 2, 2, 2)$, we have generated 100 samples for the alternating

Table 3.9: Results for the artificial data

k	cent.	diss./σ	0.1	0.3	0.5	0.7	0.9	1.1	1.3	1.5	1.7
	none	cEUC	.96	.87	.83	.775	.755	.685	.665	.6	.59
		PCC	.985	.905	.81	.765	.69	.66	.64	.605	.59
	c_i	cEUC	.955	.855	.78	.7	.62	.575	.56	.54	.52
		PCC	.93	.805	.725	.645	.61	.55	.545	.525	.505
$\langle \mathbf{x}, \mathbf{y} \rangle^d$	none	Euc.	.97	.87	.83	.775	.755	.685	.665	.625	.605
		PCC	.985	.905	.81	.765	.69	.66	.64	.605	.59
	c_i	cEUC	.955	.855	.78	.72	.68	.64	.585	.585	.555
		PCC	.93	.805	.725	.645	.61	.55	.545	.525	.505
	c_{hm}	cEUC	.96	.895	.835	.765	.75	.705	.66	.63	.6
		PCC	.985	.895	.84	.77	.725	.685	.67	.63	.625
	c_{hk}	cEUC	.955	.855	.8	.725	.695	.66	.645	.625	.595
		PCC	.955	.85	.79	.725	.685	.66	.64	.6	.575
$(\gamma + \alpha \langle \mathbf{x}, \mathbf{y} \rangle)^d$	none	cEUC	.975	**.93**	.84	.775	.755	.72	.675	.655	.635
		PCC	.985	.915	**.87**	.81	.75	.7	.65	.64	.625
	c_i	cEUC	.965	.865	.795	.74	.7	.655	.615	.6	.58
		PCC	.955	.86	.805	.715	.66	.635	.59	.56	.565
	c_{im}	cEUC	.985	.9	.82	.785	.755	**.735**	.68	**.66**	**.645**
		PCC	.985	.92	.86	.805	.76	.69	.665	.65	.63
	c_{ik}	cEUC	.97	.915	.84	.78	**.765**	.725	.675	.645	**.645**
		PCC	**.99**	.92	**.87**	**.815**	.755	.71	**.685**	.65	.635

class and 100 samples for the homogeneous class, by

$$\mathbf{x}_\sigma^i = u\mathbf{x} + v\mathbf{1} + s, \quad i = 1, \ldots, 100, \tag{3.54}$$

$$\mathbf{y}_\sigma^i = u\mathbf{y} + v\mathbf{1} + s, \quad i = 1, \ldots, 100, \tag{3.55}$$

where $u, v \in \mathcal{U}[-2, 2]$ and $s \in \mathcal{G}_8[0, \sigma]$, that is, the patterns are scaled and shifted by uniform random numbers from the range $[-2, 2]$, and Gaussian noise with standard deviation σ is added to each of their coordinates. The noise level σ was varied from 0.1 to 1.7. For each level we carried out similar tests as in the previous test scenarios and summarized the results in Table 3.9. It can be seen that except the case $\sigma = 0.3$, the proposed method (i.e. centering in the kernel space) outperforms those ones which do not use centering or do it in the input space. One can also observe that in four cases the highest accuracy scores are achieved by the centralizer function based on kernel evaluations.

3.4 Discussion and Conclusions

In this chapter, we have examined the kernelization of the centralized Euclidean distance and the Pearson correlation coefficient. Due to the structure of the cEUC and PCC measures, the kernelization depends on the feature mapping we associate with the kernel. Our contributions to the field can be summarized as follows. We showed that for feature mappings into ℓ^1, the centralized and non-centralized measures are the same in kernel space. For finite dimensional feature mappings we found that under mild conditions, the kernelization of the centralized measures can be carried out in countably infinitely many different ways, each of them leading to a different (dis)similarity measure. We have examined the homogeneous and inhomogeneous polynomial kernels in details and gave a characterization of their finite dimensional feature mappings. We found that for these kernels continuum many different centralizer functions can be constructed and there are countably infinitely many that can be computed in terms of kernel evaluations, which is highly desired in practice. We also showed that these countably infinitely many centralizer functions lead to countably infinitely many different kernelized cEUC and PCC measures. Considering the test results, we can conclude that adding translation invariance in kernel spaces does not deteriorate the classification results, in general. When invariances are present in the input space, not only centralization in the input space, but centralization in the feature space can also improve the discrimination power of the dissimilarity measures.

Based on the derivations and test results, we can formulate the second thesis of the dissertation as follows.

Thesis 2.

1. *Having a kernel with an associated feature mapping into ℓ^1, the kernelized translation invariant (dis)similarity measures cEUC and PCC are equivalent to the EUC and CC measures, respectively.*

2. *Having a kernel with a finite dimensional feature mapping into $\mathbb{R}^n$, under some mild conditions the cEUC and PCC measures can be kernelized at least in countably infinitely many different ways.*

3. *Using the homogeneous or inhomogeneous polynomial kernels, the*

translation invariant measures cEUC and PCC can be kernelized in continuum many different ways,

4. *out of which there exist at least countably infinitely many ones that can be computed in terms of kernel evaluations.*

3.5 Outlook

Centralization in the kernel space using feature mappings that can be computed in terms of kernel evaluations is a novel idea. Consequently, there are many questions arisen: For the homogeneous and inhomogeneous kernels are there further centralizer functions that can be computed in terms of kernel evaluations? Do the kernelizable centralizer functions constitute some mathematical structure, like a group or vector space? Are there further (dis)similarity measures or mathematical techniques that can be kernelized using the polynomial kernels and the developed centralizer functions which can be computed in terms of kernel evaluations? Are there further kernels with centralizer functions that can be computed in terms of kernel evaluations? These exciting questions point out a new direction in the research of kernels and kernel methods, and the test results suggest that it may be worth to study.

Chapter 4

Segmentation of retinal vessels by template matching and contour reconstruction

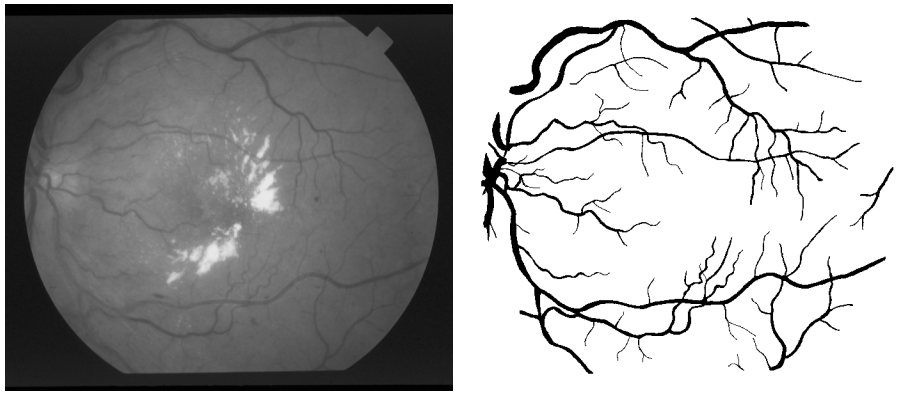
In the previous chapters we have given some novel theoretical results on (dis)similarity measures and their properties. In this chapter we go to the field of applications and describe a method for the segmentation of vessels in retinal images. The proposed method is based on template matching and as such, the proper selection and use of (dis)similarity measures is a crucial point.

4.1 Motivation and Introduction

Retinopathies are indicators of various diseases of the eye and the body. Since the examination of the retina enables in vivo screening, the automated processing of retinal images is a widely researched area in the field of medical image analysis. For several reasons the automated segmentation of the vasculature (see Figure 4.1 for the illustration of the problem) is a crucial task. On the one hand, the first sign of diabetic retinopathy is the presence of tiny capillary dilations called microaneurysms [70], appearing usually

near thin vessels. Since microaneurysms have similar visual features as high-curvature and junction points of thin vessels, the vast majority of microaneurysm detection algorithms ignore the pixels of the vasculature [71, 72], consequently, the accurate segmentation of the vascular system aids the detection of microaneurysms indirectly. The vascular system is also masked out during the segmentation of other retinopathies, like drusen [73]. Besides, the characteristic shape of the vascular system can aid the localization of other, occasionally degenerated anatomical parts like the optic disc [74], or macula [75]. Recent studies have pointed out that the geometrical features of the vascular system are in correlation with various diseases. The authors of [76] analyse the temporal changes of retinal vessel width and tortuosity related to retinopathies of prematurity. In [77] the relation of tortuosity and the Fabry disease is discussed, while in [78] an automated system for tortuosity diagnosis is described. In [79], the associations between peripheral vascular disease and the abnormal arteriolar diameters at bifurcations of the vasculature are investigated, and automated methods for the measurement and characterization of bifurcations is described in [80]. The novel study [81] discusses the associations between the changes of retinal vascular geometry and the progression of retinopathies. Age related rarefraction in the fractal dimensions of the vascular system and the correspondence between the fractal dimension and the risk of stroke is discussed in [82, 83]. According to the thorough and excellent overview of image processing issues related to retinal images in [84], *there is substantial interest in automatically segmenting the vasculature and measuring its properties.*

Recently, we have developed a robust and highly accurate technique [85] for the segmentation of the vasculature in retinal images. The proposed method has two main steps. First, a rough but high *precision* segmentation is carried out in a template matching framework. In this context the term precision refers to the widely used measure of performance in the field of machine learning, and *high precision* means that the pixels identified as vessel are likely to be true positive matches. The second step reconstructs the vessel walls by an iterative region growing technique to improve the quality of the segmentation. The method was evaluated in two *de facto* standard databases and we found that the proposed technique outperforms the previous methods in terms of segmentation accuracy. In the rest of the chapter the proposed segmentation technique is described in details.



(a) (b)

Figure 4.1: Illustration of vessel segmentation: a retinal images (a) and the manual annotation of its vasculature (b), as the target of segmentation

Particularly, in Section 4.2 we give a brief overview of previous techniques developed for the segmentation of retinal vessels, in Section 4.3 the proposed method is described, in Section 4.4 the tests and results are discussed, some conclusions are drawn in Section 4.5 and finally, a brief outlook is given in Section 4.6.

4.2 Overview of related work

One of the earliest result on the segmentation of the vascular system comes from 1989, where the use of matched line segment filters with Gaussian profiles was proposed [86]. In the first half of the 90's, several papers were published on the segmentation of the vasculature in angiograms, in [87] a recursive tracking algorithm is described, and the vessel profiles are modelled by Gaussian functions and an adaptive tracking technique is used in [88]. Although these methods can be adapted for the segmentation of the vascular system in regular retinal images, many further methods have been proposed in the upcoming two decades.

A neural network based method is described in [89] for the classification of rectangular regions of the fundus image, the pixelwise neural network

based classification is preceded by principal component analysis in [90] and fuzzy c-means clustering is applied along vessel profiles for segmentation in [91]. A method based on the tracking of Gaussian and Kalman filter responses is proposed in [92] and dual edge tracking in the Canny edge map is applied in [93].

The work [94] has boosted the development of vessel segmentation techniques by publishing a database (STARE) containing 20 fundus images and two manual annotations for each image. Later on, another database called DRIVE has been made available [95], and in a corresponding paper the authors gave a k-nearest neighbor classification based solution in [96]. Although these two databases can be considered as the *de facto* standards of evaluation, it is worth to notice that they contain images of much lower resolution than the ones used in today's practice.

In [97], a threshold probing approach is described, while the authors of [98] and [99] extended the multiscale analysis by some tracking features. In [100], 2D Gabor filter responses are used as features for supervised multivariate Gaussian mixture model based classification. Various matched filtering based techniques are proposed in [101, 102] and the vasculature is modelled by Hermite polynomials in [103, 104]. Unsupervised texture classification is proposed in [105] and support vector machine (SVM) based classification using line operator features is described in [106]. Specialized snake algorithms and active contour models are proposed in [107] and [108]. The divergence of gradient vector fields is utilized for segmentation in [109]. Some authors [110, 111] apply morphological operators utilizing that vessels are usually darker than their background.

Recently, the authors of [70] used gray-level features and grayscale invariant-moments with SVM classification. In [112] and [113] probabilistic Bayesian methods are extended by spatial constraints. The authors of [114] use RANSAC-fitting of ideal vessel templates, while [115] integrates the extraction of the optic disc with the segmentation of the vasculature to increase efficiency. In [116], AdaBoost is used with several features, including Frangi-vesselness [117], while the authors of [118] apply ensemble classification techniques for segmentation. Novel multi-scale approaches are described in [119, 120]. In [121], an SVM based method is described with connectivity constraints on the segmentation. The authors of [122] recently introduced a novel publicly available database containing high

resolution fundus images (HRF), and proposed a multiconcavity based technique for the segmentation of vessels. Surveys and overviews of retinal vessel segmentation algorithms can be found in [95] and [123]. A thorough overview of the current issues of retinal image processing (including the segmentation of vessels) can be found in the study [84].

4.3 The proposed method

The majority of the techniques described in the previous section follow the general approach of statistical learning and segmentation: extract some numerical features from the local neighborhood of pixels; train some binary supervised classifier to differentiate vessel pixels from the background and use this classifier for the segmentation of unseen input images. These techniques implicitly assume that all the vessel pixels can be separated from the background with the very same set of features. Our assumption is that better results can be achieved if the segmentation of the entire vasculature is divided to two main tasks. First, one has to detect the vessel pixels near the center-lines of the vessels: these pixels have characteristic visual features in their local neighborhoods, like a kind of symmetry due to residing near the center of the more-or-less symmetric vessel profile. In a second step, one has to refine the previous segmentation by reconstructing the vessel walls. The segmentation of the pixels near the vessel walls is more ambiguous, since the visual features of these pixels are less characteristic than that of the pixels near the center-lines.

4.3.1 The proposed template matching framework

The main goal of template matching is to yield a highly precise segmentation of vessel pixels. The vessel contours are going to be reconstructed from these pixels by an iterative region growing technique described in the next section. During the development of the method we have taken into account the following considerations:

1. Vessels have highly varying profiles – the intensity of vessels changes with the amount of oxygen transported by the blood flowing in them (veins are darker than arteries); images are usually taken under varying illumination conditions and imaging devices have varying

optical transfer functions affecting the level of noise and contrast. Even though vessels are relatively simple structures in retinal images (dark stripes on a bright background), these issues lead to many possible shapes of vessel profiles.

2. Vessels vary smoothly – shifting a vessel segment a little along the vasculature it is still highly similar to the underlying image content. Having a template representing one specific type of vessel segments accurately, islands of vessel pixels can be detected by extracting the pixels where the similarity of the template and the image content is sufficiently high.

According to the second observation, a template matching framework may be suitable to identify vessel pixels accurately. However, when the template matching framework is developed, we have to decide how to handle the issues related to the variability of vessel profiles. There are basically three choices:

1. Handling at the level of the image – many authors use some kind of preprocessing step to normalize the intensity variations in retinal images. Global and local intensity normalization techniques (like histogram equalization, background subtraction, etc. [124]) can reduce the intensity variations caused by the varying illumination conditions or large bright lesions but one can hardly expect that these simple techniques could transform each vessel to have a specific, for example sinusoidal profile (which is implicitly assumed in many techniques using Gabor filters to model vessels).

2. Handling at the level of similarity functions – there are many advanced similarity functions being invariant against even non-linear intensity transforms (like the widely used mutual information [125] or the MMTM measure introduced Chapter 2). When these similarity functions are used, there is no need for global or local normalization, since one can suspect that the similarity function focuses on the structural similarity of the template and the image content independently from slight variations of intensity profiles.

3. Handling at the level of templates – the third way is to handle the variability of vessel profiles at the level of templates, namely, one

can use a large set of templates covering almost all possible vessel shapes and a relatively simple similarity function to measure the similarity of the image content and the individual templates. If any of the individual similarity scores is high enough, the image content can be considered to contain a vessel. The drawback of this technique is that dealing with a relatively large set of templates increases the computing demands of the template matching.

We have a strong motivation to follow the third way: one can suspect that vessels with relatively sharp walls can be recognized better than vessels fading into the background smoothly, even though they have the same width. In order to reach as high accuracy as possible, we have to approach the segmentation of vessel segments with different vessel profiles in different ways. This leads us to choose the third way, namely, using a large number of templates with various profiles and a simple (dis)similarity measure to characterize the similarity of the template and the underlying image content.

4.3.2 The introduction of weak segmentation operators

Along the considerations regarding the template matching based segmentation of vessels, we introduce *weak segmentation operators*, each of them corresponding to a particular kind of vessel template. The term "weak" refers to the expectation that one operator does not extract the entire vasculature, rather, one operator extracts only patches of vessels being similar to the template corresponding to the operator. A large set of these weak segmentation operators (each of them corresponding to a particular vessel structure) is applied to the retinal image, and the resulting patches are fused to get a highly precise segmentation of vessels.

In order to realize the weak segmentation operators, we need to find an appropriate (dis)similarity function; some parametrizable vessel segment templates; and a way to binarize the outcome of template matching which is expected to be an image containing the similarity scores of a particular template and the image content.

4.3.2.1 Generalized Gabor function based templates

The use of convolutional Gabor filters is a commonly accepted way to emphasize linear structures in digital images. In its general formulation a Gabor filter is composed as the product of a complex wave and a 2D Gaussian. For linear structures being darker than their background:

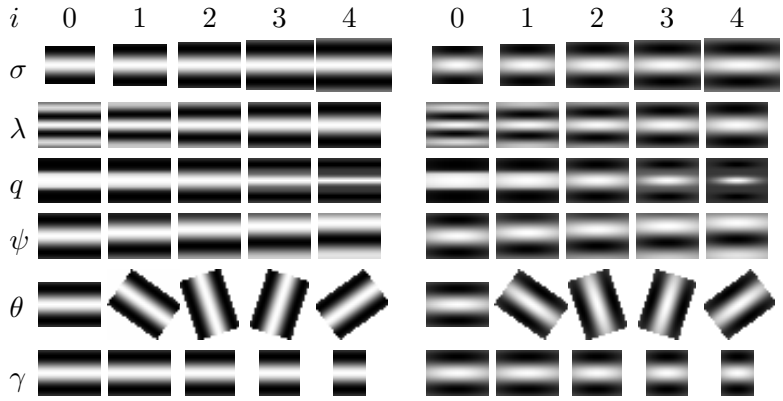
$$G(\sigma, \theta, \lambda, \gamma | r, c) = \exp\left(-\frac{\hat{r}^2 + \gamma^2 \hat{c}^2}{2\sigma^2}\right) \cos\left(2\pi \frac{\hat{r}}{\lambda} + \pi\right), \tag{4.1}$$

$$\hat{r} = r \cos\theta + c \sin\theta, \tag{4.2}$$

$$\hat{c} = -r \sin\theta + c \cos\theta, \tag{4.3}$$

where σ denotes the standard deviation of the Gaussian controlling the spatial decay of the filter, θ and λ denote the orientation and wavelength of the linear structure emphasized by the filter, and γ is the aspect ratio of the filter. In practice, the spatial extents of the filter along the local coordinate axis corresponding to coordinate r should be limited to only one valley of the cosine wave. Accordingly, the filter is considered to contain a discrete point (r, c) if $(\hat{r}, \hat{c}) \in [-\sigma_r, \ldots, \sigma_r] \times [-\sigma_c, \ldots, \sigma_c]$ holds, where $\sigma_r = \sigma$, $\sigma_c = \sigma/\gamma$, and we suppose that for the given wavelength parameter λ the spatial decay σ is selected to let approximately one valley appear in the filter. This choice of spatial extents results filters having approximately sinusoidal profiles with some exponential decay near the boundaries of the filter making them smoothly fade into zero.

Since linear structures can have arbitrary orientations, Gabor filters are usually used in matched filtering frameworks. Particularly, N distinct filters are generated by fixing the parameters σ, λ, γ, and varying θ in the range $[0, \pi]$. Each of the convolutional filters is applied to a given pixel (i, j) of the input image and the response of matched filtering is defined as the maximum of the responses of the individual filters at pixel (i, j). Clearly, the discretized Gabor filter can be considered as a template, and the matched filtering approach can be turned into an orientation independent template matching technique if the convolution operation is replaced by a similarity function. However, a given Gabor function based template can represent only one specific type of vessel segments accurately: having approximately sinusoidal profiles and fading into the background in both directions $\pm\theta$. In order to enable the representation of a wider range



(a) $\beta = 0$ (b) $\beta = 1$

Figure 4.2: Templates based on various parametrizations of the function $H(\sigma, \theta, \lambda, \gamma, q, \beta)$. In both cases, the settings $\sigma = 10$, $\lambda = 16$, $\theta = 0$, $\gamma = 0.7$, $q = 1$ are varied by changing one of the parameters: $\sigma_i = 8+i$, $\lambda_i = 8+2i$, $p_i = 1^{-i-2}$, $\psi_i = i\frac{\pi}{5}$, $\theta_i = i\frac{\pi}{5}$, $\gamma = 0.7+0.2i$

of vessel structures, some further shape parameters and the generalized Gabor function H are introduced. Particularly, in the rest of the paper, we use the templates based on the function

$$H(\sigma, \theta, \lambda, \gamma, q, \beta | r, c) = \exp\left(-\frac{\hat{r}^2 + \beta\gamma^2\hat{c}^2}{2\sigma^2}\right) A(r), \tag{4.4}$$

$$A(r) = sgn\left(\cos\left(2\pi\frac{\hat{r}}{\lambda} + \pi\right)\right)\left|\cos\left(2\pi\frac{\hat{r}}{\lambda} + \psi\right)\right|^q, \tag{4.5}$$

where $\beta \in \{0, 1\}$ and $q \in \mathbb{R}^+$ holds. First, the role of the binary parameter β used to disable the term $\gamma^2\hat{c}^2$ in the exponent is discussed. Simple Gabor function based templates have an exponential decay in every spatial direction, including $\pm\theta$ and $\theta \pm \pi/2$, so they can represent only those parts of the vascular system properly, where the intensities are increasing along the vessels in both tangential directions $\pm\theta$. By setting $\beta = 0$ and disabling the term $\gamma^2\hat{c}^2$, the exponential decay in the directions $\pm\theta$ is removed, thus, the template can represent those parts of the vascular system, where the profiles of vessels do not change along the vasculature significantly. We highlight that when the exponential decay in directions $\pm\theta$ is disabled by

setting $\beta = 0$, the parameter γ still affects the template by controlling its spatial extent as described before. The role of the exponential parameter q is to enable non-linear deviations from the sinusoidal profile. Clearly, the templates based on H can represent a wider variety of vessel segments than templates based on G, since G is a special case of H, by choosing $q = 1$ and $\beta = 1$. Some templates corresponding to specific parameter settings are visualized in Figure 4.2.

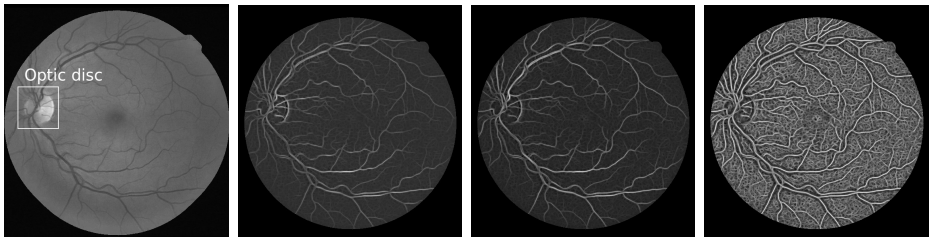
4.3.2.2 The similarity measure

Since the possible vessel structures are assumed to be represented by a large set of templates, the similarity function does not have to handle the possible non-linear deviations of one template and the real vessels. Accordingly, there are only a few requirements against the desired similarity function. The function should be an absolute measure to make the binarization of the results easier; and the function should be invariant against linear intensity transforms to make the template matching independent from the brightness of the background and the variance of the image content covered by the template. Probably the simplest and fastest similarity function satisfying these conditions is the Pearson correlation coefficient (PCC).

In order to qualitatively validate the choice of PCC as a similarity function, we have performed template matching by a specific set of orientation invariant Gabor templates using inner product (IP), simple correlation coefficient (CC) and PCC as similarity function and compared the results in Figure 4.3. As one can see in the subfigures, both IP and CC are affected by the brightness of the image content, therefore, the border of the optic disc gives similar responses as relatively thin vessels. On the contrary, PCC gives low responses at the border of the optic disc, although the noise of the background is definitely larger. Nevertheless, the PCC responses at vessel pixels are still higher than the responses on the background, thus, we can expect that even a simple binarization technique can differentiate vessels from the background when PCC is used.

4.3.2.3 Binarization of the outcome of template matching

For a given combination of parameters σ, λ, γ, q, β one can generate an orientation invariant set of generalized Gabor function based templates,



(a) Input image (b) IP (c) CC (d) PCC

Figure 4.3: The extended green channel image I_G and the result of template matching with various similarity functions for the parameter setting $\sigma = 3$, $\lambda = 8$, $\psi = \pi$, $\gamma = 0.5$, $\theta \in \{0, \pi/30, \ldots, \pi - \pi/30\}$

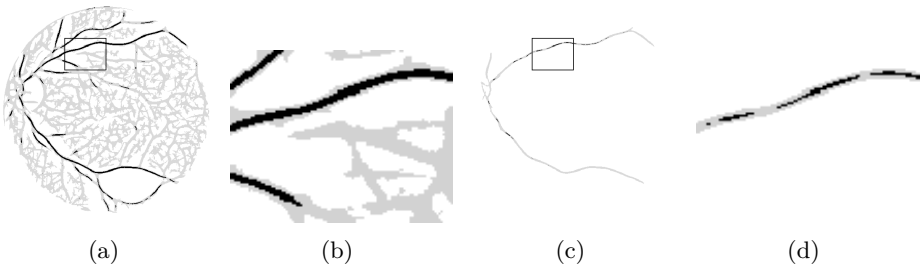
carry out template matching with PCC as a similarity function and get an intensity image, where high scores are related to vessel structures. In order to extract patches of vessels being similar to the vessel templates represented by the parameters, one has to binarize the outcome of template matching.

Since the binarization of images is one of the basic problems of image processing, there are countless methods published in the literature. For example, a thorough overview of more than 50 image thresholding techniques (histogram based methods, clustering techniques, entropy based approaches, fuzzy techniques, higher order probability based techniques and local methods) can be found in [126]. In the choice of the method we have taken into account the following considerations:

1. The method should have some tunable parameters to enable the training of binarization and the tailor-made segmentation of the various (possibly underrepresented) vessel structures.

2. In order to enable the fast numeric training and optimization of parameters, the method should be fast.

Observing some analogies with edge detection (where the image of edge magnitudes has to be binarized by tracing the ridges of the edge map), we have chosen a method similar to the hysteresis thresholding technique applied in the well-known and widely used Canny edge detector. First,



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

Figure 4.4: Seed points (black) and foreground pixels (gray) after segmentation by weak segmentation operators $W(4, 12, 0.5, 1, 0, 1, 0.7, 0.6)$ (a)-(b) and $W(4, 12, 0.5, 1, 0, 1, 0.5, 0.4)$ (c)-(d)

the image containing similarity scores is thresholded at level t_h and the foreground pixels are considered to be the seed points of an iterative region growing method. Then, pixels having intensities larger than t_l and being connected to seed points are added to the set of seed points recursively. When no further pixels can be added to the set of seed points, the operation stops and the final set of seed points is considered as the foreground of binarization. With properly selected t_h and t_l thresholds one can extract vessel pixels having local environments similar to the template described by the parameters σ, λ, γ, q, β and function H. In Figure 4.4, we have visualized the seed points and the result of hysteresis thresholding applied to the outcome of template matching. As one can observe, some settings of t_h and t_l lead to highly precise segmentations in the sense that only real vessel pixels are segmented as foreground (as desired), while other combinations lead to noisy, practically useless results.

The proposed binarization technique can be considered as a special case of the widely used fuzzy segmentation [127] techniques, with a Heaviside membership function and 4-neighborhood connectivity term. We emphasize that one could use the advanced form of this segmentation technique (smoother membership functions and connectivity terms), as well. However, the general form of fuzzy segmentation has more free parameters. As we point out in the following sections, the t_h and t_l parameters of the binarization technique are trained using manually annotated training databases.

Since the parameters of binarization may change as the parameters of the templates are varied, a large number of training operations have to be carried out. Thus, adding further free parameters to the binarization technique would highly deteriorate the performance of the training process. Since the structure of vessels is pretty simple, we assume that a binarization technique with two free parameters should be flexible enough to extract one specific type of vessel pixels precisely.

4.3.2.4 The weak segmentation operators

As a composition of template matching by generalized Gabor function based templates and hysteresis thresholding, we introduce weak segmentation operators $W(\sigma, \lambda, \gamma, q, \beta, t_h, t_l)$ by the following operation: W applies orientation independent template matching using the templates specified by $H(\sigma, i\pi/N, \lambda, \gamma, q, \beta)$, $i = 0, \ldots, N$, PCC as similarity function, and performs hysteresis thresholding on the outcome of template matching with thresholds t_h and t_l. As a technical detail we mention that the empirical choice $N = 30$ was found to give satisfactory results, thus, the angular resolution we used is 6°. Due to the limited spatial resolution of images, increasing the resolution does not improve the quality of segmentation remarkably, however, decreasing the resolution below $N = 20$ deteriorates the outcome of template matching with relatively small templates noticably. Nevertheless, small changes of the resolution do not affect the overall performance.

As a possible future improvement of the proposed method, the angular resolution of template matching could be selected by taking into consideration the size of the templates, that is, template matching with relatively large templates could be carried out at lower angular resolutions to decrease the computational demands of the method.

4.3.3 Training of weak segmentation operators

The parameters of the weak segmentation operators can be divided into two groups. Parameters λ, γ, q and β are responsible for the intensity distribution of the template, thus, these parameters determine the shape of the vessel structure recognized by the operator. Later on, we intend to use a large set of weak segmentation operators by specifying the shape

descriptor parameters explicitly to make the templates corresponding to the operators cover the possible vessel structures (by making the parameters cover the reasonable range of the parameter space uniformly). Contrarily, the parameters σ, t_h, t_l are related to the nature of the problem: the level of noise in the images, the contrast of the vessel borders, etc. For a given setting of shape descriptors λ, γ, q, β the proper values of σ, t_h, t_l have to be trained.

4.3.3.1 Training of one weak segmentation operator

In order to simplify the notations, we introduce two multidimensional parameters for the weak segmentation operators, namely $\mathbf{s} = (\lambda, \gamma, q, \beta)$ is related to the shape of the template, $\mathbf{n} = (\sigma, t_0, t_1)$ is related to the nature of the problem and the notation of a weak segmentation operator is simplified to $W(\mathbf{s}, \mathbf{n})$.

Given an explicit setting of the shape descriptor parameter $\mathbf{s}$, the goal of training is to find the parameter $\hat{\mathbf{n}}$ leading to a highly precise segmentation of vessel pixels. The training goes on by following a simulated annealing [128] based stochastic optimization process respecting the maximization of a target function measuring how precise the segmentation is. In the rest of the paragraph we give the outline of training by simulated annealing. Let $\mathbf{n}^{(k)}$ denote the parameter combination at the kth iteration of the optimization process and $T(\mathbf{n}^{(k)})$ the corresponding value of the target function. For a random combination $\mathbf{n}^{(k+1)}$ the operator $W(\mathbf{s}, \mathbf{n}^{(k+1)})$ is applied to the images of the training database, the segmentation results are evaluated and the target function $T(\mathbf{n}^{(k+1)})$ is determined. The combination $\mathbf{n}^{(k+1)}$ is accepted if the target function has increased ($T(\mathbf{n}^{(k+1)}) > T(\mathbf{n}^{(k)})$), or

$$\exp\left(\frac{T(\mathbf{n}^{(k+1)}) - T(\mathbf{n}^{(k)})}{Z}\right) > r \tag{4.6}$$

holds, where Z denotes the temperature of the system being decreased as the iterations go on and r denotes a uniformly distributed random number in the range $[0, 1]$. Otherwise, the update of the parameter is reverted: $\mathbf{n}^{(k+1)} = \mathbf{n}^{(k)}$. It is worth noting that any other randomized optimization method could be used, as well.

We emphasize again that we do not expect one weak segmentation operator to segment all vessel pixels, rather, we expect one operator to find

vessel pixels surely belonging to the foreground. Accordingly, the target function we have chosen to navigate towards the optimal parameter $\hat{n}$ for a given s is based on the *precision* (referring to the performance measure in binary classification), also known as positive predictive value (PPV) of the segmentation: let TP denote the number of correctly classified vessel pixels in the database and AP the overall number of pixels classified as vessel; then, the positive predictive value is defined as $PPV = TP/AP$. Higher PPV means more real vessel pixels and less background pixels classified as foreground. Since segmentation always has errors, the desired highest PPV values are usually reached when TP is small and $TP \simeq AP$. In order to extract more pixels, the conditions are slightly weakened: small errors are allowed by introducing ϵ-*insensitive* PPV as the target function of optimization. Particularly,

$$T(\mathbf{n}) = \begin{cases} PPV, & \text{if } PPV < 1 - \epsilon, \\ TP, & \text{otherwise.} \end{cases} \tag{4.7}$$

Maximizing the target function means to find the parameter setting $\hat{\mathbf{n}}$ that provides $PPV > 1 - \epsilon$ and TP is as large as possible. Obviously, the target function leads to misclassified foreground pixels, however, we expect that these errors do not interfere when several weak segmentation operators with various parameters are used, and can be removed by simple image processing techniques in a later step. If the condition $PPV > 1 - \epsilon$ cannot be satisfied for a given shape descriptor parameter $\mathbf{s}$, the operator with parameters $\mathbf{s}$ is discarded. This can be interpreted as the setting $\mathbf{s} = (\lambda, \gamma, q, \beta)$ describes templates that cannot be distinguished from the background and noise precisely enough.

The method described in this paragraph can be used to train the parameter $\mathbf{n}$ of one specific weak segmentation operator $W(\mathbf{s}, \mathbf{n})$ for a given shape descriptor $\mathbf{s}$. In order to gain a highly precise segmentation of the entire vasculature, we have to train many weak segmentation operators with various settings of $\mathbf{s}$ related to the various types of vessel structures and unify the results of the individual operators.

We introduce the notation $\mathcal{S} = \Lambda \times \Gamma \times \mathcal{Q} \times \mathcal{B}$ for the set of reasonable parameter combinations of shape descriptors and $\mathcal{W}(\mathcal{S})$ to denote the set of weak segmentation operators $W(\mathbf{s}, \mathbf{n})$, where $\mathbf{s} \in \mathcal{S}$ holds and $\mathbf{n}$ is trained

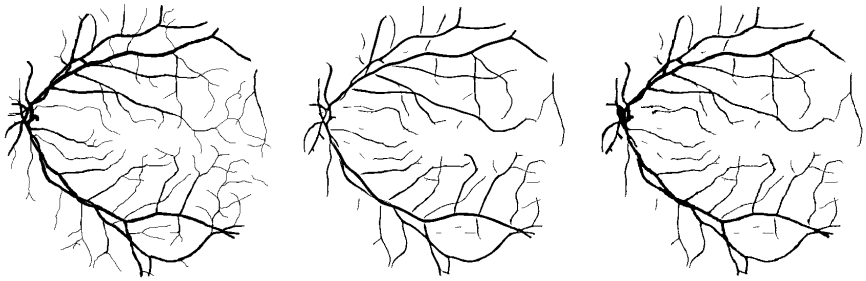
by the method described before. For example, using the parameters

$$\begin{aligned}
\Lambda &= \{6, 7, \ldots, 24\}, \\
\Gamma &= \{0.4, 0.5, \ldots, 1.2\}, \\
\mathcal{Q} &= \{0.166, 0.2, 0.25, 0.33, 0.5, 1, 2, 3, 4, 5, 6\}, \\
\mathcal{B} &= \{0, 1\},
\end{aligned} \tag{4.8}$$

and $\epsilon = 0.05$, the cardinality of the set $\mathcal{S}$ becomes 3762. After the training of weak segmentation operators on the 20 training images of DRIVE, the cardinality of $\mathcal{W}(\mathcal{S})$ becomes 3340, showing that there were approximately 400 combinations of shape descriptors that cannot represent vessel structures with sufficiently high PPV scores. The smallest wavelength $\lambda = 6$ pixels appearing in Λ is selected by the following consideration: if the wavelength of a sinusoidal wave is 6 pixels, the width of one valley becomes 3 pixels, that is, the cross section of the template contains one dark pixel in the middle having one bright pixel on both of its sides. Clearly, a template having a profile like this represents vessels of width 1 pixel. As a general rule, we can state that independently from the resolution of the images, the use of weak segmentation operators having wavelengths smaller than 6 pixels is worthless. Similar considerations were used to select the maximum wavelength $\lambda = 24$ pixels, taking into account the width of the thickest vessels appearing in the DRIVE database. The smallest parameter in the set Γ corresponds to vessel segment templates having more than two times larger length than width. Considering the nature of vessels, it is unlikely that much longer straight vessel segments would appear. The largest parameter $\gamma = 1.2$ in the set Γ defines vessel segment templates having slightly larger width than length. If the γ is further increased, the templates do not represent line segments any more. Finally, due to the non-linear effect of the exponential parameter q, and the fast convergence of x^q, $x \in [0, 1]$ as $q \to 0$ or $q \to \infty$, the settings $q \in [0, 0.166[$ or $q > 6$ lead to highly similar templates as the ones $q = 0.166$ or $q = 6$, respectively.

4.3.3.2 Finding the optimal set of weak segmentation operators

Depending on the cardinality of set $\mathcal{S}$, $W(\mathcal{S})$ can contain thousands of weak segmentation operators, making the proposed method computationally intractable in practice. In order to reduce the computational demands and



(a) (b) (c)

Figure 4.5: Manual segmentation (ground truth) (a); fused outcome of the operators $\mathcal{W}^{red}(\mathcal{S})$ (b); the outcome of contour reconstruction (c)

achieve better segmentation results, the raw set $\mathcal{W}(\mathcal{S})$ of weak segmentation operators is reduced. Computational demands can be decreased by removing the subset $\mathcal{O} \subset \mathcal{W}(\mathcal{S})$ if the pixels classified by the operators in $\mathcal{O}$ as vessel are classified by the operators in $\mathcal{W}(\mathcal{S})/\mathcal{O}$ in the same way. Better segmentation can be achieved by removing the subset $\mathcal{O} \subset \mathcal{W}(\mathcal{S})$ that produces interfering true positive but less interfering false positive matches with the operators in $\mathcal{W}(\mathcal{S})/\mathcal{O}$.

Examining all subsets of $\mathcal{W}(S)$ is computationally infeasible. Suboptimal but acceptable results can be achieved by applying the commonly used backward feature subset selection (BFSS) method to the set $\mathcal{W}(\mathcal{S})$ with a properly chosen target function. Namely, let $\mathcal{W}' = \mathcal{W}(S)$ initially. Then, we repeat the following step while the target function increases: for each $W \in \mathcal{W}'$ the images of the training database are segmented by the weak segmentation operators in the set $\mathcal{W}'/\{W\}$, the binary outcomes are unified by applying pixel-wise "or" operation and the target function is computed; if the highest value of the target function is reached at $\mathcal{W}'/\{W^*\}$, let $\mathcal{W}' = \mathcal{W}'/\{W^*\}$. The target function we have chosen for the BFSS method is the accuracy of the segmentation results: the ratio of correctly classified pixels and all pixels of the training database. After the iteration has finished, the set $\mathcal{W}'$ contains the reduced set of weak segmentation operators, denoted by $\mathcal{W}^{red}(\mathcal{S})$ for better readability.

For illustration, we have applied the proposed BFSS technique to the set $\mathcal{W}(\mathcal{S})$ of 3340 operators introduced in the previous paragraph. In the resulting set $\mathcal{W}^{red}(\mathcal{S})$ the number of operators decreased to 384 having tractable computation demands.

4.3.3.3 Combining the outcomes of weak segmentation operators

So far, we have trained and selected the computationally tractable $\mathcal{W}^{red}(\mathcal{S})$ set of weak segmentation operators. After the preprocessing of an unseen input image, the operators in $\mathcal{W}^{red}(\mathcal{S})$ are applied to the input image, yielding $|\mathcal{W}^{red}(S)|$ binary images, each of them corresponding to a specific kind of vessel structure. In order to yield one binary image as the outcome of Step 2, the results of the individual operators are fused by applying a pixel-wise "or" operation. However, since the outcomes of the operators may contain some foreground pixels related to noise, some shape based filtering is applied to the connected components of the binarized images. Particularly, the connected foreground components in the outcomes of the individual operators are extracted (the objects in Figure 4.4(c) show examples for these components), and the sufficiently elongated components are added to the fused result by pixel-wise "or" operation. The elongatedness can be measured by the circularity of the component defined by $4\pi A/P^2$, where A and P denote the area and the perimeter of the object. Obviously, the more elongated a component is, the smaller its circularity becomes, thus, an object is added to the outcome of Step 2 if its circularity is smaller than the threshold t^{circ}. We have found empirically that the choice $t^{circ} = 0.3$ gives satisfactory results, independently from the source of images being segmented. We also note that small variations of this threshold do not change the results at all.

To illustrate the fusion of the individual binary segmentations and enable the comparison with the ground truth annotation, we have used the set $\mathcal{W}^{red}(\mathcal{S})$ containing 384 operators (being introduced in the previous paragraph) to apply Step 2 of the proposed method to a test image and visualized the binary outcome and the ground truth annotation in figures 4.5(b) and 4.5(a), respectively.

4.3.3.4 A note on the combination of features

The method described in this sebsection uses weak segmentation operators to extract the pixels corresponding to various types of vessel profiles and fuses the binary results to produce a highly precise segmentation. Contrarily to the previous methods published in the literature, the pixels corresponding to different types of vessel profiles are segmented independently. This behaviour of the proposed method enables the accurate representation of even underrepresented vessel structures: if there is a template similar to a vessel structure, there will be a dedicated weak segmentation operator, thus, a dedicated segmentation model.

Clearly, there are many ways to fuse the results of template matching, even before the binarization of the images. In the following paragraph we overview and discuss some of the obvious possibilities.

1. One could combine the outcomes of template matching operations for example by taking the pixel-wise maximum of the similarity scores and binarize the resulting image. However, the similarity scores corresponding to templates of different sizes and intensity distributions are not comparable directly. Due to the higher level of relative noise, the similarity scores corresponding to thin vessel templates are generally lower than that of thick vessel templates. Thus, this kind of fusion would smooth out the fine details and the interpretation and accurate segmentation of the fused image would be extremely difficult.

2. The outcomes of the various template matching operations could be combined into higher dimensional feature vectors and advanced machine learning techniques (like SVM, neural networks) could be used to transform the segmentation problem into a binary classification problem. The drawback of this approach is that the binary classification techniques tend to produce smooth classification surfaces to avoid overfitting, and this smoothness usually represses the underrepresented structures. Thus, the segmentation of relatively rare vessel structures can hardly be guaranteed to be accurate.

3. In order to keep the dedicated modeling of underrepresented strucutres, one could introduce class labels for the various vessel

profiles and use multiclass classification techniques to train the segmentation models. In this way the underrepresented structures would still have a dedicated model, however, it would become difficult to classify the vessel profiles of the training database accurately and the training of multiclass classification techniques is computationally expensive.

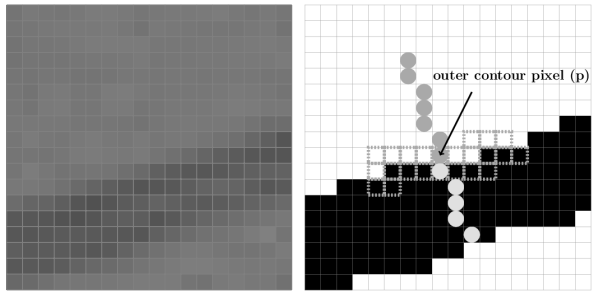
As a summary, the method is based on the natural assumption, that the best feature for the precise segmentation of one specific type of vessel is the outcome of template matching by the corresponding vessel template. We emphasize again, that this step is not intended to extract all the vessel pixels, the main goal is to get a highly precise initial segmentation. According to the discussion above, the independent segmentation of various vessel profiles is not a drawback, rather, this is how rare and underrepresented structures can be modeled properly.

4.3.4 Reconstruction of vessel contours

So far we have introduced some simple weak segmentation operators and a fusion technique to yield highly precise segmentations in the sense that the majority of foreground pixels in the fused outcomes of template matching are real vessel pixels. Since the individual weak segmentation operators extract mainly the center lines of vessels, we cannot accept the fused result as final segmentation: many false negative classifications appear along the borders of vessels. In the current step we focus on the reconstruction of vessel contours to gain a highly accurate segmentation of the entire vasculature.

Considering the intensity profile of a vessel, relatively small intensities appear in the middle and increase towards the vessel walls. Ideally, the contour pixels of a segmentation should have the same relative intensity respecting the minimum intensity of the vessel profile and the mean intensity of the background. Since the sensitivity of the visual perception of human observers varies with the wavelength of the stimulus [129] and the images have finite resolution, we assume that this is not the case in practice. We expect that the position of true vessel contours varies with the width and contrast of vessels.

In order to validate these expectations, we introduce three measures to



(a) Vessel segment (b) Corresponding manual segmentation

Figure 4.6: Measuring the width of a vessel

characterize the contour pixels of segmented vessels: the width (w) of the segmented vessel passing by, the dynamic range (d) of the vessels and the relative intensity (r) at the contour respecting the minimum intensity of the vessel profile and the mean intensity of the background.

4.3.4.1 Measuring the width, dynamic range and relative intensity at contour points

It is worth noting that there are some recent studies focusing on measuring the width of vessels in retinal images [104]. These methods usually fit a parametric surface model to the vessel and estimate its width by the analytic parameters of the fitted model. Contrarily, what we need is to estimate the width of *already segmented* vessels at their inner and outer contour points. The method described below gives an extremely fast solution with satisfactory accuracy.

Let **p** denote an inner or outer contour pixel of a segmented vessel (for illustration we have visualized a real vessel segment and the corresponding manual annotation in figures 4.6(a) and 4.6(b), respectively). In order to find the width of the vessel being tangential to this pixel, we have to get a robust estimation of its normal vector. Sobel operators are applied to the contour pixels of the binary image in a close neighborhood and the mean of the gradient directions is computed. We have found empirically that a disk

shaped neighborhood with the radius of 5 pixels gives satisfactory results. In Figure 4.6(b) pixels with dotted border indicate the region, where the gradient directions are averaged. Let $\alpha(\mathbf{p})$ denote the mean of the gradient directions. A line segment is fitted to $\mathbf{p}$ at angle $\alpha(\mathbf{p})$ (gray dots in Figure 4.6(b)), and the pixels where the line segment intersects the inner and outer contours of the other side of the vessel are located. Then, if $\mathbf{p}$ is an inner contour pixel, the width of the vessel tangential to $\mathbf{p}$ is estimated as the Euclidean distance of $\mathbf{p}$ and the outer contour pixel intersected by the line segment at the other side of the vessel. Similarly, if $\mathbf{p}$ is an outer contour pixel, the width of the vessel is estimated as the Euclidean distance of pixel $\mathbf{p}$ and the inner contour pixel intersected by the line segment at the other side of the vessel. In order to make the estimations smooth along the vasculature, the width $w(\mathbf{p})$ of the vessel tangential to $\mathbf{p}$ is defined as the mean of the estimated widths in the close neighborhood.

Measuring the dynamic range of a vessel tangential to $\mathbf{p}$ is carried out similarly to the measurement of width: the line segment orthogonal to the vessel walls is divided in two regions: the intersection of the line segment and the vessel ($\mathcal{L}_V(\mathbf{p})$ - denoted by light gray dots in Figure 4.6(b)), and the intersection of the line segment and the background ($\mathcal{L}_B(\mathbf{p})$ - denoted by dark gray dots in Figure 4.6(b)). The dynamic range is defined as the difference of the minimum intensity $min_{\mathcal{L}_V(\mathbf{p})}$ covered by region $\mathcal{L}_V(\mathbf{p})$ and the mean intensity $mean_{\mathcal{L}_B(\mathbf{p})}$ covered by region $\mathcal{L}_B(\mathbf{p})$:

$$d(\mathbf{p}) = \max\left\{mean_{\mathcal{L}_B(\mathbf{p})} - min_{\mathcal{L}_V(\mathbf{p})}, 0\right\} \tag{4.9}$$

Finally, the relative intensity of the contour pixel respecting the mean intensity of the background and the minimum intensity of the vessel profile is defined as:

$$r(\mathbf{p}) = \max\left\{\min\left\{\frac{I_G[\mathbf{p}] - min_{\mathcal{L}_V(\mathbf{p})}}{mean_{\mathcal{L}_B(\mathbf{p})} - min_{\mathcal{L}_V(\mathbf{p})}}, 1\right\}, 0\right\}. \tag{4.10}$$

The value $r(\mathbf{p}) \in [0, 1]$ represents the normalized relative intensity of the contour pixel $\mathbf{p}$ within the range $[min_{\mathcal{L}_V(\mathbf{p})}, mean_{\mathcal{L}_B}(\mathbf{p})]$, thus, $r(\mathbf{p})$ characterizes the position of the contour on the normal of the vessel taking into account only the intensity relations.

4.3.4.2 Characterization of relative intensities at the contour points

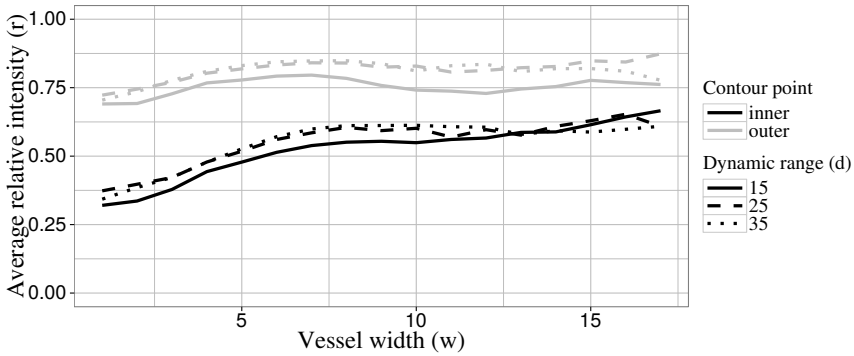
As mentioned before, we assume that the relative intensities depend on the dynamic range and width of vessels. Accordingly, we have computed the means of relative intensities at the inner and outer contour points of vessels having specific width (w) and dynamic range (d) properties in the manually annotated training databases DRIVE and STARE; arranged the values to form two surfaces $r_I(w, d)$ and $r_O(w, d)$ corresponding to the relative intensities at inner and outer contours, respectively; applied some smoothing with a small averaging kernel; and plotted some geodesics in Figure 4.7.

It can be observed, that the shape and distance of the curves is highly similar, independently from the scale of the images (the resolutions of the images in STARE are approximately 1.3 those of DRIVE) and the dynamic range: for thin vessels the relative intensities are small and increase until the width of approximately 8 pixels is reached. Then, a slower but coherent convergence of the curves can be observed as the width of the vessels increases towards the maximum. Another interesting property is that independently from the dynamic range or database, the distance of the curves decreases as the width of the vessels increases. The most characteristic difference of the curves extracted from various databases is their vertical position. The height of the curves extracted from STARE are remarkably higher than that of DRIVE. Based on these observations we can assert that

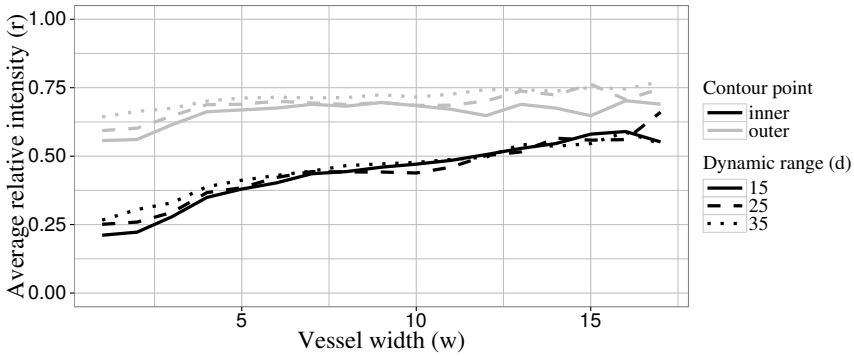
1. the position of vessel contours (represented by their relative intensities) varies with the width of the vessels and slightly varies with the dynamic range of vessel profile;

2. in order to reach highly accurate segmentation results by the reconstruction of vessel contours, one must take into account the information about the position of the contours represented by the relative intensity surfaces.

In order to reconstruct the characteristics of contours measured in the training database, the contours of the binary outcome of template matching have to be adjusted. Moreover, it can happen that several layers



(a) Some relative intensity curves in the DRIVE database



(b) Some relative intensity curves in the STARE database

Figure 4.7: Relative intensity curves measured on various databases

of outer contour pixels have to be added to the foreground to reach the desired similarity of relative intensities extracted from the training database and measured at the contour pixels of the binary segmentation. For the sake of readability the description of the proposed contour reconstruction technique is separated in two subsections: first, the iterative adjustment of one layer of contour pixels is described, then, the scheduling of the iterative adjustment involving one layer of contour pixels is discussed.

4.3.4.3 The iterative solution for one layer of contour pixels

Probably the most obvious and simple way to recover the desired relative intensities at the contour pixels of a binary segmentation is defining some target function measuring how the surfaces $r_I(w,d)$ and $r_O(w,d)$ fit the current segmentation, and adjusting the binary labels of contour pixels randomly towards the extremum of the target function.

Let $\mathcal{C}$ denote the set of contour pixels being involved in the iterative optimization procedure. The core of the iteration is a simple, pixelwise operation described below. For a pixel $\mathbf{p} \in \mathcal{C}$ the score $s_0(\mathbf{p}) = s(\mathbf{p})$ is computed; the binary label at pixel $\mathbf{p}$ is negated if the modification does not make a connected vessel component disconnected; and the score of the modified binary image is computed as $s_1(\mathbf{p}) = s(\mathbf{p})$. The new label at pixel $\mathbf{p}$ is kept, if $s_0(\mathbf{p}) > s_1(\mathbf{p})$, otherwise the modification is reverted. This simple operation is repeated for the contour pixels in $\mathcal{C}$ subsequently. The stopping condition is either convergence (when no more pixels are changed), or the method is stopped after the local operation is applied 30 times for all pixels of $\mathcal{C}$. As we describe below, the score function $s(\mathbf{p})$ depends on the relative intensity surfaces extracted from the manually annotated databases. Correspondingly, the convergence of the method can hardly be prooved rigorously, since erroneous annotations may lead to relative intensity surfaces of highly different characteristics, and in extreme cases the optimization could fall into a loop. In practice, we found that the relative intensity surfaces extracted from various databases do not lead to such loops, and the method converges. As a fall-back case, the upper limit of 30 repetitions was selected empirically, based on the maximum number of iterations required for convergence.

The score function is defined as follows. Let $r(\mathbf{p})$, $w(\mathbf{p})$ and $d(\mathbf{p})$ denote the measured relative intensity, width and dynamic range of the vessel passing by pixel $\mathbf{p}$, respectively, and let $\mathbf{p}$ be an inner contour point. We can use the measured $r_I(w,d)$ surface to estimate the ideal relative intensity at the contour of a vessel having width $w(\mathbf{p})$ and dynamic range $d(\mathbf{p})$. However, since the measured surface is known only at some discrete coordinates, while $w(\mathbf{p})$ and $d(\mathbf{p})$ can take any real value, bilinear interpolation is applied to evaluate the function $r_I(w,d)$ and determine the expected relative intensity $\hat{r}(\mathbf{p}) = r_I(w(\mathbf{p}), d(\mathbf{p}))$ at pixel $\mathbf{p}$. If $\mathbf{p}$ is an outer contour pixel, the desired ideal intensity is derived analogously from the surface $r_O(w,d)$. Let the

term

$$e(\mathbf{p}) = \begin{cases} \max\{0, r(\mathbf{p}) - \hat{r}(\mathbf{p})\}, \text{if } \mathbf{p} \text{ is inner contour point,} \\ \max\{0, \hat{r}(\mathbf{p}) - r(\mathbf{p})\}, \text{if } \mathbf{p} \text{ is outer contour point,} \end{cases} \tag{4.11}$$

measure the error of the estimated and desired relative intensities. This error term is part of the score function evaluated before and after the binary label of $\mathbf{p}$ is changed. If $\mathbf{p}$ is an inner contour pixel and its relative intensity is smaller than the desired relative intensity, it means that the pixel is closer to the center of the vessel than the inner contour pixels of vessels having similar width and dynamic range in the training database. In this case the error becomes zero, since the pixel should not change its label during the optimization process. However, changing the label of $\mathbf{p}$ to the opposite, it becomes an outer contour pixel, with a large error, thus, the error term suppresses to keep the original foreground label of $\mathbf{p}$. The error term behaves similarly if $\mathbf{p}$ is an outer contour pixel and its relative intensity is larger than the desired one. If $\mathbf{p}$ is an inner contour pixel and $r(\mathbf{p})$ is larger than the ideal, the error term measures the distance of $r(\mathbf{p})$ from the ideal $\hat{r}(\mathbf{p})$ intensity. Changing its label to the opposite, the error term measures the distance of $r(\mathbf{p})$ from the ideal relative intensity of outer contour pixels of vessels with similar width and dynamic range. The smaller value of the error term indicates the true label of $\mathbf{p}$. The error term has similar behaviour if $\mathbf{p}$ is an outer contour pixel and $r(\mathbf{p})$ is smaller than the ideal. Consequently, the error term $e(\mathbf{p})$ suppresses a label configuration providing relative intensities being similar to the ideal one. In order to avoid cracked edges, and ensure local smoothness, another term is added to $e(\mathbf{p})$ measuring the similarity of labels in a local neighborhood. Particularly, the smoothness term becomes $w^{nb}(1 - n(\mathbf{p})/8)$, where $n(\mathbf{p})$ measures the number of 8-neighbors of $\mathbf{p}$ having the same binary label as $\mathbf{p}$, and $w^{nb} \in \mathbb{R}$ is a control parameter responsible for the strength of smoothing. Then, the score function $s(\mathbf{p})$ becomes:

$$s(\mathbf{p}) = \frac{e(\mathbf{p})}{r_O(w(\mathbf{p}), d(\mathbf{p})) - r_I(w(\mathbf{p}), d(\mathbf{p}))} + w^{nb}\left(1 - \frac{n(\mathbf{p})}{8}\right). \tag{4.12}$$

The error term is normalized by the distance of the relative intensity surfaces of inner and outer contours to prevent the score function from

preferring thick vessels, where the distance of the curves and consequently the errors of the reconstruction are smaller.

Based on this score function, the iterative optimization process repeats the previously described pixelwise operation for all contour pixels in $\mathcal{C}$ until convergence or up to 30 repetitions for all pixels of $\mathcal{C}$. However, filling $\mathcal{C}$ with inner and outer contour pixels leads to ambiguous results, since the edges of the vasculature may become leaky if an inner contour point is removed, and a neighboring outer contour point is added to the foreground in the same time. In order to avoid these situations, the proposed iterative optimization technique is applied to the layers of inner and outer contour pixels subsequently as described in the next subsection.

4.3.4.4 The contour reconstruction process

The strategy of the proposed contour reconstruction technique is summarized in the following points:

1. We can expect that the adjustment of vessels having extremely small dynamic ranges is ambiguous, since they can hardly be differentiated from the background. Accordingly, a contour pixel is not adjusted if the dynamic range of the vessel passing by is less than t^{dr}.

2. In the first phase, $\mathcal{C}$ contains the inner contour pixels of the vasculature and the iterative optimization described in the previous subsection is applied until convergence. Since the weak segmentation operators introduced before tend to extract the center lines of the vessels, we expect this step to carry out only small adjustments on the contour of the segmented vasculature, and do not expect that more adjustment of inner contour pixels is required. Thus, the next steps involve outer contour pixels only. Two technical details are also noted here:

   (a) The neighborhood term in the score function $s(\mathbf{p})$ is disabled if $\mathbf{p}$ is an inner contour pixel and the width of the vessel passing by is smaller than 2 pixels, since the local 8-neighborhood of the inner contour pixel $\mathbf{p}$ may contain pixels of the outer contour of the opposite side of the vessel and lead to undetermined behaviour.

(b) When a vessel runs into and ends at the border of a bright lesion, or fades into the background, the estimated normal of the vessel at the end point becomes parallel with the vessel and the brightness of the background can suppress the removal of inner contour pixels, shortening the extracted vessels. This undesired behaviour is prevented by extracting the end points of the segmented vasculature and forbidding the removal of these foreground pixels during the adjustment of inner contours.

3. In the second phase, $\mathcal{C}$ is filled with the outer contour pixels of the binary segmentation having already adjusted inner contours, and the iterative optimization technique described in the previous section is applied until convergence. We expect this step to increase the width of vessels by adding outer contour pixels to the foreground of the segmentation, if required. In the case of thick vessels the addition of one layer of outer contour pixels can rarely reconstruct the real width of vessels. Thus, the adjustment of outer contour pixels is repeated layer-by-layer until the set of outer contour pixels becomes unchanged.

The proposed approach for the reconstruction of vessel contours has basically two free parameters: the control parameter of the local smoothness term w^{nb} and the threshold of the dynamic range t^{dr}. The values of these parameters are determined by training on the images of a manually annotated training database for the highest accuracy. For the illustration of the proposed technique, Figure 4.5(c) presents the test image with reconstructed contours. The free parameters we used are the ones trained on DRIVE: $w^{nb} = 3.145$ and $t^{dr} = 10$. As one can observe, the widths of thick vessels became similar to that of the manual annotation, while the widths of relatively thin vessels have not changed significantly, as expected.

A common weakness of template based vessel segmentation techniques is that the similarities near junction points of the vasculature (especially near the junction points of arteries having a bright stripe in the middle) are usually low, resulting fragmented segmentations (see the junction points in Figure 4.8(a)). As a limitation of the proposed method, these structures cannot be modelled accurately enough with the Gabor function based templates we use. Nevertheless, the contour reconstruction technique

(a) (b)

Figure 4.8: Illustration of the iterative contour reconstruction process: the binary segmentation after Step 2 (a), the binary segmentation after contour reconstruction (b). Some examples for the reconstruction of junction points are emphasized by grey rectangles

implicitly attempts to reconstruct junctions: the relative intensities at the outer contour pixels of the erroneous end points of vessels are close to zero, since the estimated normals are parallel with the vessels. Consequently, outer contour pixels are added near the end points, the vessels are traced towards the junctions. As Figure 4.8(b) shows, many of the junctions become reconstructed by the proposed method in this way. In order to make the entire vasculature connected, the reconstruction of all junctions would be required. This could be accomplished in a Hidden Markov Model based framework (similar to the one in [130]) by optimizing an energy function consisting of a connectivity term and a probability term describing how well the added vessel pixels fit the distribution of nearby vessels.

4.4 Tests and Results

The detailed description of the testing methodology is extremely important to make the results comparable with the state-of-the-art approaches. Although the databases and techniques used for evaluation vary, we try to follow a protocol which makes the proposed method comparable with the majority of previous techniques within the available space.

4.4.1 Databases

The majority of vessel segmentation methods are evaluated on two publicly available databases. The DRIVE database [95] consists of 40 RGB fundus images of size 565×584, divided to training and test sets of equal size, both sets containing healthy and abnormal images, as well. The database contains the manual annotations of two human observers, and the annotations of observer 1 are considered to be the ground truth in the literature. The STARE database [94] contains 10 healthy and 10 abnormal RGB fundus images of size 700×605 pixels. Similarly to DRIVE, database STARE contains the manual annotations of two experts, and the annotations of observer 1 are considered to be ground truth segmentations. In many recent papers the vessel segmentation techniques are evaluated using only DRIVE and STARE [131, 112, 132, 119, 122, 70], furthermore, there are some papers where only database DRIVE is used for evaluation [116]. Thus, we can consider these databases as the commonly accepted way for the evaluation and comparison of vessel segmentation techniques.

4.4.2 Measuring segmentation performance

The segmentation performance is characterized by the commonly accepted measures of binary classification. The test images are segmented by the proposed method, and the numbers of correctly classified vessel pixels (TP), incorrectly classified vessel pixels (FN), correctly classified background pixels (TN) and incorrectly classified background pixels (FP) are determined in the region covered by the ROI. Sensitivity $SE = TP/(TP + FN)$ and specificity $SP = TN/(TN + FP)$ measure the proportion of correctly classified foreground and background pixels to all foreground and background pixels, respectively. For both measures, higher values indicate better results. Positive predictive value $PPV = TP/(TP + FP)$ measures the proportion of correctly classified vessel pixels to all pixels classified as vessel. Negative predictive value $NPV = TN/(TN + FN)$ measures the same for background pixels. Again, higher values indicate better performance. Accuracy $ACC = (TP + TN)/(TP + FP + TN + FN)$ measures the proportion of correctly classified pixels to all pixels covered by the ROI. Most of the authors use ACC as the primary measure to characterize the performance of vessel segmentation. However, the accuracy scores have to

be interpreted carefully. Due to the large number of background pixels, the accuracy of a segmentation containing only background labels is 0.83 in DRIVE, which is considered to be a fairly good score in many applications of binary classification. Even in DRIVE – containing the smallest images used for evaluation – a 0.001 increase of accuracy represents the correct classification of approximately 200 additional pixels per image.

The authors of machine learning based approaches usually report another measure: the receiver operating characteristic (ROC) curve plots sensitivity SE as a function of specificity SP, and the corresponding measure is the area under the curve AUC. In fact, the AUC score does not measure the performance of an individual system, although retinal screening systems are based on automated components. Rather, the AUC score is an aggregated measure of classification performance when the parameters of the system are tuned: AUC measures the trade-off of performance when the sensitivity of the method is increased to the detriment of specificity. Although the AUC score is a commonly reported measure, it has to be treated with reservations when vessel segmentation methods are ranked by it. On the one hand, measuring the AUC score supposes that there is a kind of global threshold that can be used to vary SE and SP smoothly. There are algorithms like the proposed one, where neither global probability maps, nor continuous tuning parameters are available to measure various pairs of SE and SP values, plot the ROC curve, and compute the AUC score. On the other hand, there are several techniques used to handle the concavities of ROC curves and measure the area under the curve: one can use simple numerical integration, polynomial fitting, or measure the area under the convex hull of the curve, and each of these techniques leads to different AUC scores. There are also recent criticisms of the AUC score regarding its misleading properties when classifiers are compared by it [133, 134]. Although the proposed method is not suitable for the computation of the AUC score, we have generated some SE and SP pairs by manually varying the thresholds in the weak segmentation operators and the heights of relative intensity surfaces and calculated the AUC score in the same way as described in [118]. Nevertheless, we treat the AUC score as a secondary measure of vessel segmentation performance and consider the accuracy score (ACC) as the primary indicator.

4.4.3 Comparison with existing techniques

The performance of the proposed method is compared to the ones reported and used for comparison in recent papers. The majority of authors calculated the measures in the region covered by the ROI and we consider this method to be the preferred way to measure the performance of vessel segmentation techniques. Nevertheless, some authors used all the pixels of the images to calculate the performance measures [106, 116]. This can be easily verified by solving the equations defining ACC, SE and SP for the unknowns TP, TN, FP and FN using the reported values of the measures and the number of foreground pixels in the test images FG. For the sake of completeness, the values TP, TN, FP and FN are determined by the expressions:

$$TP = SE \cdot FG, \quad FN = FG - TP, \tag{4.13}$$

$$FP = \frac{TP(ACC - 1) + ACC \cdot FN}{\frac{SP}{1-SP}(1 - ACC) - ACC}, \tag{4.14}$$

$$TN = \frac{SP}{1 - SP} FP, \tag{4.15}$$

In both of the mentioned cases, the sum $TP + FN + FP + TN$ equals the number of pixels of the images instead of the number of pixels covered by the ROI. In order to make the proposed method comparable with these techniques, we have also computed the performance measures involving all pixels of the images and these results are referred as *Evaluation without ROI*.

4.4.4 Model selection

The parameters of the proposed method are trained as we have described in Section 4.3. Especially, for the training of template matching operators, we have used the sets Λ, Γ, $\mathcal{Q}$ and $\mathcal{B}$ as given in Section 4.3.3, assuming that the parameter combinations of these sets densely cover the meaningful and reasonable part of the parameter space. An increased resolution of the parameter space could slightly improve the accuracy of segmentation, since more than 90% of the trained weak segmentation operators were discarded when the optimal set of operators was selected by the BFSS method. According to the vast majority of previous techniques, the primary indicator

of performance is the accuracy of segmentation. Thus, the free parameters w^{nb} and t^{dr} are selected to maximize the accuracy of segmentation using the training images of databases.

4.4.5 DRIVE database

We have evaluated the proposed method on database DRIVE. At the end of the training process the set $\mathcal{W}^{red}(\mathcal{S})$ contained 384 weak segmentation operators. The surfaces of relative intensities were extracted (some geodesics can be seen in Figure (4.7) and the free parameters providing the highest accuracy scores on the training images have been determined: $w^{nb} = 3.14$ and $t^{dr} = 10$.

The proposed method was applied to the images of the test set and the averages of the previously introduced performance measures were computed. The mean accuracy, sensitivity and specificity scores are compared to those of previous techniques in the second column of Table 4.1. One can observe, that after the reconstruction of vessel contours the accuracy scores are higher than that of any method published before. One can also observe that even the template matching step alone is able to outperform several techniques in the literature. Considering the gaps between the accuracies of techniques published in the last 10 years, we can assess that the advances of the proposed method are significant. Since the PPV and NPV scores are rarely reported, we cannot compare them to those of previous techniques. Anyway, for the sake of completeness, the average PPV and NPV scores are 0.8397 and 0.9634, respectively. Regarding the AUC score, there is only one technique performing better than the proposed one. For illustration, Figure 4.9 shows the segmentation results of healthy and abnormal images from DRIVE.

4.4.6 STARE database

The proposed method has been evaluated also on database STARE. We have separated the images of the database into two sets: 10 training and 10 test images. The training set contains the first five healthy and first five abnormal images (with identifiers: 0001, 0002, 0003, 0004, 0005, 0162, 0163, 0235, 0236, 0239), and the rest of the images constitute the test set.

At the end of the training process, the set $\mathcal{W}^{red}(\mathcal{S})$ contained 288

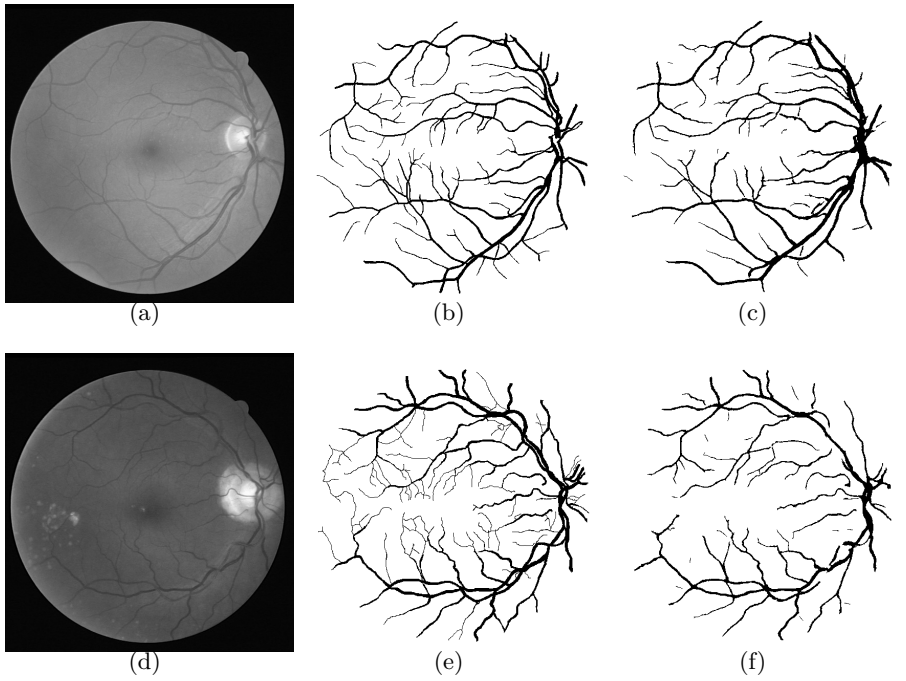


Figure 4.9: Illustration of the proposed method on the database DRIVE. Healthy (a), abnormal (one can observe the presence of bright lesions) (d) images; ground truth annotation (b), (e); segmentations by the proposed method trained on database DRIVE (c), (f)

operators. The relative intensity surfaces were extracted, and the free parameters providing the highest accuracy scores on the training images have been selected: $w^{nb} = 3.12$ and $t^{dr} = 10$. Comparing the optimal values of free parameters to those trained on DRIVE, one can observe, that w^{nb} and t^{dr} are essentially the same, suggesting that these values are related to the nature of the problem of vessel segmentation.

The trained segmentation model was applied to the complete database and the averages of the performance measures were computed for the test images of the database. The results are compared to those of previously published techniques in Table 4.2. One can observe, that the accuracy scores of the proposed technique are much higher than that of any other

Table 4.1: Test results in DRIVE. The highest accuracy scores are indicated by boldface typesetting

Measure	*ACC*	*SE*	*SP*	*AUC*
Method	**Evaluated with ROI**			
Second observer	0.9473	-	-	-
Jiang and Mojon [97]	0.9212	-	-	0.9114
Vlachos and Dermatas [99]	0.9291	0.7472	0.9550	-
Martinez-Perez et al. [98]	0.9181	-	-	-
Espona et al. [135]	0.9352	0.7472	0.9615	-
Nguyen et al. [120]	0.9407	-	-	-
Niemeijer et al. [95]	0.9417	0.6898	0.9696	-
Staal et al. [96]	0.9442	0.7194	0.9773	-
Marín et al. [70]	0.9452	0.7067	0.9801	-
Mendonça and Campilho [132]	0.9452	0.7344	0.9764	-
Soares et al. [100]	0.9466	0.7332	0.9751	0.9614
Lam, Gao, and Liew [122]	0.9472	-	-	0.9614
Fraz et al. [118]	0.9480	0.7406	0.9807	0.9747
The proposed method after template matching	0.9420	0.6497	0.9846	-
The proposed method after contour reconstruction	**0.9491**	0.7344	0.9804	-
	Evaluated without ROI			
Ricci and Perfetti [106]	0.9595	-	-	0.9633
Lupascu, Tegolo, and Trucco [116]	0.9597	0.6728	0.9874	0.9561
The proposed method after template matching	0.9601	0.6495	0.9899	-
The proposed method after contour reconstruction	**0.9652**	0.7446	0.9864	-

technique published before. The results of the template matching are also highly competitive: there is only one method giving better results than the template matching framework. The average PPV and NPV scores are 0.8200 and 0.9768, respectively. The AUC score of the proposed method is more than 0.5% higher than that of the best performing previous technique trained and tested on STARE. Figure 4.10 illustrates the segmentation results of healthy and abnormal images of STARE.

4.5 Discussion and Conclusions

We have proposed a novel approach for the segmentation of vessels in retinal images: since the center lines of vessels have more characteristic visual features than the vessel walls, we have separated the extraction of pixels near the center lines and the reconstruction of vessel contours; and

Table 4.2: Test results in STARE. The highest accuracy scores are indicated by boldface typesetting

Measure	*ACC*	*SE*	*SP*	*AUC*
Method	**Evaluated with ROI**			
Second observer	0.9550	-	-	-
Jiang and Mojon [97]	0.9009	-	-	0.9298
Staal et al. [96]	0.9516	-	-	0.9614
Marín et al. [70]	0.9526	0.6944	0.9819	0.9769
Mendonça and Campilho [132]	0.9440	0.6996	0.9730	-
Soares et al. [100]	0.9479	0.7207	0.9747	0.9671
Lam, Gao, and Liew [122]	0.9567	-	-	0.9739
Fraz et al. [118]	0.9534	0.7548	0.9763	0.9768
The proposed method after template matching	0.9552	0.7126	0.9822	-
The proposed method after contour reconstruction	**0.9610**	0.8012	0.9788	
	Evaluated without ROI			
Ricci and Perfetti [106]	0.9646	-	-	0.9680
The proposed method after template matching	0.9670	0.7122	0.9872	-
The proposed method after contour reconstruction	**0.9714**	0.8030	0.9847	-

we have also taken into consideration that some kinds of vessels may be under-represented in the training databases.

The proposed method uses the trained weak segmentation operators and the extracted relative intensity surfaces to represent the visual features of vessels. Although there are some numerical constants wired in the method, most of them do not affect the segmentation directly, instead, they are related to how various quantities (like the width of vessels or relative intensities) are measured. Changing these constants coherently throughout the method does not affect the quality of the outcome remarkably. Aside from these technical constants and supposing that the initial shape descriptor parameters of weak segmentation operators cover the reasonable part of the parameter space, the proposed method has only two free numerical parameters that have to be trained using the images of a training database.

Considering that the accuracy of the proposed method outperformed all the previous techniques in every test case, we can conclude that the method can become a robust and reliable building block of retinal image processing systems.

Finally, we formulate the third thesis of the dissertation.

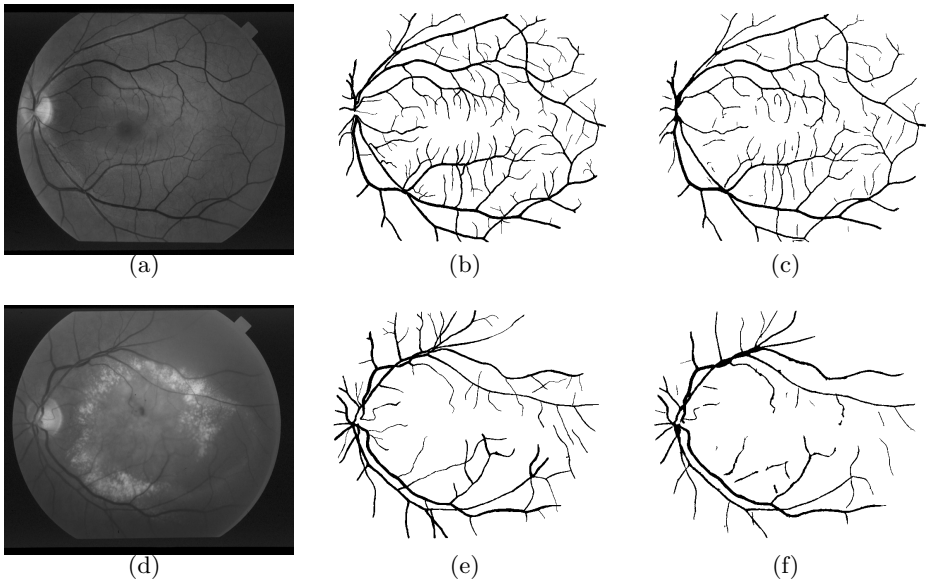


Figure 4.10: Illustration of the proposed method on the STARE database. Healthy (a), abnormal (d) images; ground truth annotation (b), (e); segmentations by the proposed method trained on database STARE (c), (f)

Thesis 3.

1. *I have proposed a novel technique for the segmentation of vessels in retinal images. As a novel contribution to the field, in the development of the method I have taken into consideration that various vessel structures are under-represented in the manually annotated training databases. Correspondingly, the proposed method learns and utilizes the visual features of a large number of predefined vessel profiles individually.*

2. *Based on the test results I can state that the method is highly competitive with the state-of-the-art techniques, and the features the method is based on have high descriptive power for the representation of the vascular system.*

4.6 Outlook

The performance of the proposed method could be further increased by involving weak segmentation operators with templates of other intensity distributions than the ones derived from the generalized Gabor functions. For example, Hermite polynomials can model the intensity profile of arteries (having bright stripes at the center lines due to the high level of oxygen) more accurately than exponential or trigonometric functions [104]. Consequently, the use of weak segmentation operators with templates based on Hermite polynomials could improve the modelling of the vasculature and increase the accuracy of segmentation.

Although the method is developed for the segmentation of retinal vessels, it can be adapted to other problems in which tubular structures are to be segmented, like road networks in aerial images, or vessel segmentation in volumetric images. In the latter case the 3D generalization of template matching operators and that of the contour reconstruction process is needed, which can be done in a straightforward way. The idea of training a large number of classifiers (each of them responsible for the high precision recognition of one specific pattern) can be easily adapted to general classification problems of machine learning.

Chapter 5

Summary

In the dissertation I have discussed some results of several years of research carried out in the field of pattern recognition, especially in image processing and computer vision. According to the title of the dissertation, the results are closely related to the theory and applications of (dis)similarity functions.

In Chapter 2 two novel dissimilarity measures were derived, both of them are approximately invariant to even non-linear monotonically increasing transformations and can be computed for an image in the time of some convolution operations. The proposed measures were tested and evaluated in various scenarios of template matching and compared to other techniques designed to be invariant to monotonic transformations. The results show that the measures are highly competitive, in certain cases giving remarkably better performance than conventional techniques. The measures are independent from the spatial dimensionality of the objects being compared, thus, besides template matching in images (which we use for demonstration and evaluation) both of the proposed measures can be used to compare and measure the dissimilarity of 1D time series or even 3D images.

In Chapter 3 I have examined how the kernel trick can be applied to translation invariant (dis)similarity measures, like the *Pearson correlation coefficient* (PCC) and *centered Euclidean distance* (cEUC). I have given some general results for kernels with finite dimensional feature maps, namely, showed that under some mild conditions there are countably infinite ways for the kernelization of these translation invariant measures. Then, I

have examined the case of the polynomial kernel in details, showed that using the polynomial kernel the cEUC and PCC measures can be kernelized in continuum many different ways and there are at least countably many variants that can be computed efficiently in terms of kernel evaluations. The discrimination power of the kernelized PCC and cEUC measures was evaluated in classification problems using the k-nearest neighbour classifier. The results show that the performance of the kernelized cEUC and PCC measures is highly competitive, and in certain problems better than that of the non-kernelized variants.

Finally, in Chapter 4 a novel technique for the segmentation of the vasculature in retinal images was described. The method is based on template matching operators, and correspondingly, on the use of (dis)similarity measures. The template matching framework we developed for the accurate representation of the vasculature can be generalized to many other problems of pattern recognition. The proposed method was tested on the images of two publicly available datasets and compared to many previously published techniques in terms of the conventional measures of binary classification. It was found that the proposed method outperforms the previous techniques on the *de facto* standard image databases, showing that the template matching based representation of the vasculature is a reasonable technique and the features extracted by the method have high discriminative power.

As we pointed out in the outlook sections, at least as many questions arose as many we have answered, so we can continue the research in each of these interesting topics without loss of momentum.

Chapter 6

Összefoglalás

A disszertációban a digitális képfeldolgozás és alakfelismerés területén végzett több éves kutatómunka néhány eredményét foglaltam össze. A dolgozat címének megfelelően a bemutatott eredmények elsősorban a hasonlósági mérőszámok elméletéhez és gyakorlati alkalmazásaikhoz kapcsolódnak.

Az eredmények első csoportjában (második fejezet) monoton intenzitástranszformációkra közelítőleg invariáns hasonlósági mértékeket vezettem be, melyekre az angol terminológiát használva *Matching by Monotonic Tone Mapping* (MMTM) mértékekként hivatkozok. Elméleti úton beláttam, hogy a javasolt MMTM függvények számos kedvező tulajdonsággal rendelkeznek: abszolút mértékek, azaz a $[0, 1]$ tartományba képeznek és néhány konvolúciónak megfelelő időben számolhatók egy teljes képre. A javasolt módszereket alapos tesztelésnek vetettem alá, és az irodalomban széleskörűen alkalmazott, monoton transzformációkra invariáns hasonlósági mértékekkel hasonlítottam össze. A teszteredmények alapján kijelenthetem, hogy a javasolt módszerek a korábbi (sokszor négyzetes futásidejű) algoritmusokkal összemérhető teljesítményt nyújtanak, és bizonyos problémák esetén jobb eredmények adódnak a javasolt mértékek alkalmazásával. Bár a levezetés során a digitális képfeldolgozás terminológiáját használtam, a módszerek függetlenek az összehasonlított objektumok dimenzionalitásától, így képek mellett akár idősorok vagy volumetrikus adatok összehasonlítására is használhatók.

A bemutatott eredmények második csoportjában (harmadik fejezet) azt vizsgáltam, hogy a kernel trükk alkalmazható-e az eltolásinvariáns Pearson

korrelációs együtthatóra (PCC) és a szintén eltolásinvariáns centralizált euklideszi távolságra (cEUC). Sikerült belátni, hogy véges dimenziós sajátságfüggvénnyel rendelkező kernelek esetén nem túl szigorú feltételek mellett a PCC és cEUC függvények megszámlálhatóan végtelen sok különböző módon kernelizálhatók. A széleskörűen alkalmazott polinom kernel esetére beláttam, hogy a cEUC és PCC hasonlósági mértékek kontinuum sok különböző módon kernelizálhatóak és legalább megszámlálhatóan végtelen sok esetben a kernelizált függvények kiszámíthatóak a sajátságfüggvények ismerete nélkül, kernel kiértékelések segítségével. Így a gyakorlati alkalmazások szempontjából is hasznos, gyorsan számítható kernelizált hasonlósági mértékekhez jutottam. A kernelizált cEUC és PCC függvények diszkriminációs erejét általános osztályozási problémákban vetettem össze a nem kernelizált függvények teljesítményével. A teszteredmények alapján kijelenthetem, hogy általános esetben a cEUC és PCC függvények polinom kernellel történő kernelizálása nem rontja azok teljesítményét, sőt, bizonyos problémákban jobb osztályozási eredményeket kaphatunk a kernelizált mértékek alkalmazásával.

Az eredmények harmadik csoportjában (negyedik fejezet) egy új módszert javasoltam az érhálózat szegmentálására retina képeken. A módszer ún. gyenge szegmentáló operátorokon alapszik, melyek mindegyike csak egy bizonyos típusú érhálózat szegmens felismeréséért felelős. A gyenge szegmentáló operátorokban fontos szerep jut a hasonlósági mértékeknek, ugyanis az operátorok működése mintázat illesztésen alapszik. A javasolt módszert két publikusan elérhető, *de facto* sztenderdnek számító adatbázison hasonlítottam össze korábbi módszerekkel. Az eredmények alapján kijelenthetem, hogy a javasolt módszer pontosabban szegmentálja az érhálózatot mint a korábbi módszerek; a gyenge szegmentáló operátorokon alapuló reprezentáció, illetve a sajátságok, amelyek alapján a szegmentáció történik magas leíró erővel bírnak.

Bár a dolgozatban számos kérdésre választ adtunk, munkánk legalább ugyanennyi további kérdést vetett fel, így számos érdekes irányba folytathatjuk a kutatómunkát ezen érdekes témákban és területeken.

References

[1] Y. Hel-Or, H. Hel-Or, and E. David. "Matching by Tone Mapping: Photometric Invariant Template Matching". In: *IEEE Trans. on Pattern Analysis and Machine Intelligence* 36(2) (2014), pp. 317–330.

[2] M. Sofka and C. V. Stewart. "Retinal Vessel Centerline Extraction Using Multiscale Matched Filters, Confidence and Edge Measures". In: *IEEE Transactions on Medical Imaging* (2006), pp. 1531–1546.

[3] J. V. B. Soares, J. J. G. Leandro, R. M Cesar, and H. F. Jelinek. "Retinal vessel segmentation using the 2-D Gabor wavelet and supervised classification". In: *IEEE Transactions on Medical Imaging* (2006), pp. 1214–1222.

[4] Gyorgy Kovacs and Andras Hajdu. "A self-calibrating approach for the segmentation of retinal vessels by template matching and contour reconstruction". In: *Medical Image Analysis* 29.4 (2016), pp. 24 –46.

[5] Gy. Kovács. "Matching by Monotonic Tone Mapping". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2016). submitted.

[6] S. Theodoridis and K. Koutroumbas. *Pattern Recognition, 4th edn.* Academic Press, New York, 2009, pp. 602–606.

[7] C. Spearman. "The proof and measurement of association between two things". In: *Am. J. Psychol* 15(1) (1904), pp. 72–101.

[8] M. G. Kendall. "A new measure of rank correlation". In: *Biometrika* 30 (1938), pp. 81–93.

[9] G. A. Fredricks and R. B. Nelsen. "On the relationship between Spearman's rho and Kendall's tau for pairs of continuous random variables". In: *J. Stat. Plan. Inference* 137 (2007), pp. 2143–2150.

[10] R. A. Gideon and R. A. Hollister. "A rank correlation coefficient". In: *J. Am. Stat. Assoc.* 82(398) (1987), pp. 656–666.

[11] N. Bhat and S. K. Nayar. "Ordinal measures for image correspondence". In: *IEEE Trans. Pattern Anal. Mach. Intell.* 20(4) (1998), pp. 415–423.

[12] A. Venot and V. Leclerc. “Automated correction of patient motion and gray values prior to subtraction in digitized angiography”. In: *IEEE Trans. Med. Imaging* 3 (1984), pp. 179–185.

[13] A. Venot, J. F. Lebruchec, J. L. Golmard, and J. C. Roucayrol. “An automated method for the normalization of scintigraphic images”. In: *J. Nucl. Med.* 24 (1983), pp. 529–531.

[14] S. Kaneko, I. Murase, and S. Igarashi. “Robust image registration by increment sign correlation”. In: *Pattern Recognition* 35(10) (2002), pp. 2223–2234.

[15] A. Collignon, F. Maes, D. Delacre, D. Vandermeulen, P. Suentens, and A. Marchal. “Automated multi-modality image registration based on information theory”. In: *Proc. Information Processing in Medicine Conf.* 1995, pp. 263–274.

[16] M. P. Wachowiak, R. Smolikova, G. D. Tourassi, and A. S. Elmaghraby. “Similarity metrics based on nonadditive entropies for 2D-3D multimodal biomedical image registration”. In: *Medical Imaging Conf.* Vol. 5032. 2003, pp. 1090–1100.

[17] J. P. W. Pluim, J. B. A: Maintz, and M. A. Viergever. “f-information measures in medical image registration”. In: *IEEE Trans. Med. Imageing* 23(12) (2004), pp. 1506–1518.

[18] M. Mellor and M. Brady. “Phase mutual information as a similarity measure for registration”. In: *Med. Image Anal.* 9 (2005), pp. 330–343.

[19] L. Liu, T. Jiang, J. Yang, and C. Zhu. “Fingerprint registration by maximization of mutual information”. In: *IEEE Trans. Image Process.* 15(5) (2006), pp. 1100–1110.

[20] K. Pearson. “Mathematical contributions to the theory of evolution, XIV, on the general theory of skew correlation and non-linear regression”. In: *Drapers’ Company Research Memoirs, Biometric Series.* Dulau and Co. London, 1905, p. 54.

[21] Y. H. Lau, M. Brain, and B. F. Hutton. “Non-rigid image registration using a median filtered coarse-to-fine displacement field and a symmetric correlation ratio”. In: *Phys. Med. Biol.* 46 (2001), pp. 1297–1319.

[22] R. P. Woods, J. C. Mazziotta, and S. R. Cherry. “MRI-PET registration with automated algorithm”. In: *J. Comput. Assist. Tomogr.* 17(4) (1993), pp. 536–546.

[23] M. K. Hu. “Visual Pattern Recognition by Moment Invariants”. In: *IRE Trans. Info. Theory* IT-8 (1962), pp. 179–187.

[24] J. Flusser and T. Suk. “Rotation Moment Invariants for Recognition of Symmetric Objects”. In: *IEEE Trans. Image Proc.* 15 (2006), pp. 3784–3790.

[25] T. Ojala, M. Pietikainen, and D. Harwood. “A comparative study of texture measures with classification based on feature distributions”. In: *Pattern Recognition* 29 (1996), pp. 51–59.

[26] S. Liao, M. W. K. Law, and A. C: S. Chung. “Dominant Local Binary Patterns for Texture Classification”. In: *IEEE Trans. Image Proc.* 18(5) (2009), pp. 1107–1118.

[27] D. G. Lowe. “Object recognition from local scale-invariant features”. In: *Proceedings of the International Conference on Computer Vision 2*. 1999, pp. 1150–1157.

[28] H. Bay, A. Ess, T. Tuytelaars, and L. V. Gool. “SURF: Speeded Up Robust Features”. In: *Computer Vision and Image Understanding (CVIU)* 110(3) (2008), pp. 346–359.

[29] K. Mikolajczyk and C. Schmid. “A performance evaluation of local descriptors”. In: *IEEE Trans. on Pattern Analysis and Machine Intelligence* 27(10) (2005), pp. 1615–1630.

[30] S. Sarfraz and O. Hellwich. “Head Pose Estimation in Face Recognition across Pose Scenarios”. In: *Proceedings of International Conference on Computer Vision Theory and Applications*. 2008, pp. 235–242.

[31] S. Omachi and M. Omachi. “Fast template matching with polynomials”. In: *IEEE Trans Image Process* 16(8) (2007), pp. 2139–2149.

[32] S. Mattoccia, F. Tombari, and L. Di Stefano. “Fast full-search equivalent template matching by enhanced bounded correlation”. In: *IEEE Trans Image Process* 17(4) (2008), pp. 528–538.

[33] J. Yoo, S. S: Hwang, S. D. Kim, M. S. Ki, and J. Cha. “Scale-invariant template matching using histogram of dominant gradients”. In: *Pattern Recognition* 47(9) (2014), pp. 3006–3018.

[34] Y.-H. Lin and C.-H. Chen. “Template matching using the parametric template vector with translation rotation and scale invariance”. In: *Pattern Recognition* 41(7) (2008), pp. 2413–2421.

[35] E. Elboher and M. Werman. “Asymmetric correlation: a noise robust similarity measure for template matching”. In: *IEEE Trans Image Process* 22(8) (2013), pp. 3062–3073.

[36] K. Fredriksson, V. Makinen, and G. Navarro. “Rotation and lighting invariant template matching”. In: *Invormation and Computation* 205(7) (2007), pp. 1096–1113.

[37] Fast Frequency Template Matching Using Higher Order Statistics. “F. Essannouni and D. Aboutajdine”. In: *IEEE Trans Image Process* 19(3) (2010), pp. 826–830.

[38] H. Duan, C. Xu, S. Liu, and S. Shao. “Template matching using chaotic imperialist competitive algorithm”. In: *Pattern Recognition Letters* 31 (2010), pp. 1868–1875.

[39] R. Brunelli. *Template Matching Techniques in Computer Vision: Theory and Practice*. Wiley, 2009, p. 348.

[40] A. A. Goshtasby. *Image Registration, Principles, Tools and Methods*. Springer-Verlag London, 2012, p. 442.

[41] M. M. Deza and E. Deza. *Encyclopedia of Distances*. Springer, 2009, p. 583.

[42] E. Deza and M. M. Deza. *Dictionary of Distances*. Elsevier, 2006, p. 391.

[43] J. Nocedal and J. S. Wright. *Numerical Optimization (2nd ed.)* Springer Verlag, 2006, p. 449.

[44] M. K. Kozlov, S. P Tarasov, and L. G. Khachlyan. "Polynomial solvability of convex quadratic programming". In: *USSR Computational Mathematics and Mathematical Physics* 20(5) (1980), pp. 223–228.

[45] M. Ayer, H. D. Brunk, G. M. Ewing, W. T. Reid, and T. Silverman. "An ampirical distribution function for sampling with incomplete information". In: *Annals. of Math. Stat.* 5 (1955), pp. 641–647.

[46] A. C. Floudas and P. M. Pardalos (Eds.) *Encyclopedia of Optimization*. Springer US, 2009, p. 4622.

[47] D. Martin, C. Fowlkes, D. Tal, and J. Malik. "A Database of Human Segmented Natural Images and its Application to Evaluating Segmentation Algorithms and Measuring Ecological Statistics". In: *Proc. 8th Int'l Conf. Computer Vision*. Vol. 2. 2001, pp. 416–423.

[48] Federico Tombari, Luigi Di Stefano, Stefano Mattoccia, and Angelo Galanti. "Performance evaluation of robust matching measures". In: *Proc. 3rd International Conference on Computer Vision Theory and Applications (VISAPP)*. 2008.

[49] Gy. Kovács. "On the Optimal Quantization Technique of Photometric Invariant Template Matching and Its Properties in the Presence of Gaussian Noise". In: *Proc. of KÉPAF 2015*. January 27–30. Kecskemét, Hungary, 2015, pp. 657–672.

[50] B. Scholkopf. "The Kernel Trick for Distances". In: *NIPS*. 2000, pp. 301–307.

[51] B. Scholkopf and A. J. Smola. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. Cambridge, MA, USA: MIT Press, 2001, p. 648.

[52] T. Hofmann, B. Scholkopf, and A. J. Smola. "Kernel Methods in Machine Learning". In: *The Annals of Statistics* 31.3 (Jan. 2008), pp. 1171–1220.

[53] C. H. Lampert. "Kernel Methods in Computer Vision". In: *Found. Trends. Comput. Graph. Vis.* 4 (3 2009), pp. 193–285.

[54] V. N. Vapnik. *Statistical Learning Theory*. Wiley-Interscience, Sept. 1998, p. 768.

[55] K. Yu, L. Ji, and X. Zhang. "Kernel Nearest-Neighbor Algorithm". In: *Neural Processing Letters* 15 (2 2002). 10.1023/A:1015244902967, pp. 147–156.

[56] A. Denton and W. Perrizo. "A Kernel-Based Semi-Naive Bayesian Classifier Using P-Trees." In: *SDM'04*. 2004, pp. 427–431.

[57] F. Camastra. "Kernel Methods for Clustering". In: *WIRN/NAIS*. Vol. 3931. Lecture Notes in Computer Science. 2006, pp. 1–9.

[58] B. Schlkopf, A. J. Smola, and K. R. Müller. "Kernel principal component analysis". In: *Advances in kernel methods: support vector learning* (1999), pp. 327–352.

[59] R. Cai, Z. Hao, W. Wen, and H. Huang. “Kernel based gene expression pattern discovery and its application on cancer classification”. In: *Neurocomput.* 73 (13-15 2010), pp. 2562–2570.

[60] J. Zhou, Y. Liu, and Y. Chen. “Face Recognition Using Kernel PCA and Hierarchical RBF Network”. In: *Proceedings of the 6th International Conference on Computer Information Systems and Industrial Management Applications*. Washington, DC, USA: IEEE Computer Society, 2007, pp. 239–244.

[61] N. S. Nehe and R. S. Holambe. “Isolated Word Recognition Using Low Dimensional Features and Kernel Based Classification”. In: *ARTCom.* 2009, pp. 194–198.

[62] Gy. Kovács and A. Hajdu. “Translation Invariance in the Polynomial Kernel Space and Its Applications in kNN Classification”. In: *Neural Processing Letters* 37(2) (2013). (IF=1.237), pp. 207–233.

[63] A. Nachman. “Theory of Reproducing Kernels”. In: *Transactions of the American Mathematical Society* 68.3 (1951), pp. 337–404.

[64] M. E. Blazadonakis and M. Zervakis. “Polynomial and RBF Kernels as Marker Selection Tools-A Breast Cancer Case Study”. In: *Machine Learning and Applications, Fourth International Conference on* 0 (2007), pp. 488–493.

[65] D. Jadhav, J. Kulkarni, and R. Holambe. “Multiresolution Feature Based Fractional Power Polynomial Kernel Fisher Discriminant Model for Face Recognition”. In: *Journal of Multimedia* 3.1 (2008), pp. 47–53.

[66] M. Evans, N. Hastings, and B. Peacock. *Statistical Distributions*. Wiley-Interscience, June 2000.

[67] C. L. Blake and C. J. Merz. *UCI Repository of Machine Learning Databases*. 1998.

[68] M. Niemeijer, B. Van Ginneken, J. Staal, M. S. A. Suttorp-Schulten, and M. D. Abramoff. “Automatic detection of red lesions in digital color fundus photographs.” In: *IEEE Trans. on Medical Imaging* 24.5 (2005), pp. 584–592.

[69] Meindert Niemeijer, Bram van Ginneken, Michael J Cree, Atsushi Mizutani, Gwenola Quellec, Clara I Sanchez, Bob Zhang, Roberto Hornero, Mathieu Lamard, Chisako Muramatsu, Xiangqian Wu, Guy Cazuguel, Jane You, Agustin Mayo, Qin Li, Yuji Hatanaka, Beatrice Cochener, Christian Roux, Fakhri Karray, Maria Garcia, Hiroshi Fujita, and Michael D Abramoff. “Retinopathy online challenge: automatic detection of microaneurysms in digital color fundus photographs.” In: *IEEE Trans Med Imaging* 29.1 (2010), pp. 185–95.

[70] D. Marín, A. Aquino, M. E. Gegúndez-Arias, and J. M. Bravo. “A new supervised method for blood vessel segmentation in retinal images by using gray-level and moment invariants-based features”. In: *IEEE Trans. Med. Imag.* (2011), pp. 146–158.

[71] A. D. Fleming, S. Philip, K. A. Goatman, J. A. Olson, and P. F. Sharp. “Autmoated microaneurysm detection using local contrast normalization and local vessel detection”. In: *IEEE Trans Med Imaging* 25(9) (2006), pp. 1223–1232.

[72] A. Sopharak, B. Uyyanonvara, and S. Barman. “Simple hybrid method for fine microaneurysm detection from non-dilated diabetic retinopathy retinal images”. In: *Computerized Medical Imaging and Graphics* 37(5-6) (2013), pp. 394–402.

[73] A. Bhulyan, R. Kawasaki, M. Sasaki, E. Lamoreux, K. Ramamohanarao, R. Guymer, T. Y. Wong, and K. Yogesan. “Drusen Detection and Quantification for Early Identification of Age Related Macular Degeneration using Color Fundus Images”. In: *J. Clin Exp Ophthalmol* 4(5) (2013), pp. 1–6.

[74] A. Hoover and M. Goldbaum. “Locating the optic nerve in a retinal image using the fuzzy convergence of blood vessels”. In: *IEEE Trans Med Imaging* 22(8) (2003), pp. 951–958.

[75] C-Y Yu, C-C Liu, and S-S Yu. “A Fovea Localization Scheme Using Vessel Origin-Based Parabolic Model”. In: *Algorithms* 7(3) (2014), pp. 456–470.

[76] C. S. Cheung, Z. Butty, N. N. Tehrani, and W. C. Lam. “Computer-assisted image analysis of temporal retinal vessel width and tortuosity in retinopathy of prematurity for the assessment of disease severity and treatment outcome”. In: *J AAPOS* 16(5) (2012), pp. 431–436.

[77] A. Sodi, M. Guarducci, L. Vauthier, A. S. Ioannidis, S. Pitz, G. Abbruzzese, F. Sofi, A. Mecocci, A. Miele, and U. Menchini. “Computer assisted evaluation of retinal vessels tortuosity in Fabry disease”. In: *Acta Ophthalmologica* 91(2) (2013), pp. 113–119.

[78] Q. Li, J. You, W. Jinghua, and A. Wong. “A fully automated system for retinal vessel tortuosity diagnosis using scale dependent vessel tracing and grading”. In: *Proc. of IEEE 23rd International Symposium on Computer-Based Medical Systems (CGMS)*. 2010, pp. 221–225.

[79] N. Chapman, G. Dell’omo, M. S. Sartini, N. Witt, A. Hughes, S. Thom, and R. Pedrinelli. “Peripheral vascular disease is associated with abnormal arteriolar diameter reslationships at bifurcations in the human retina”. In: *Clin Sci (Lond).* 103(2) (2002), pp. 111–116.

[80] B. Al-Diri and A. Hunter. “Automated Measurements of Retinal Bifurcations”. In: *Proc. of IFMBE World Congress on Medical Physics and Biomedical Engineering*. 2009, pp. 205–208.

[81] M. S. Habib, B. Al-Diri, A. Hunter, and D. H. W. Steel. “The association between retinal vascular geometry changes and diabetic retinopathy and their role in prediction of progression – an exploratory study”. In: *BMC Ophthalmology* 14(89) (2014), pp. 1–11.

[82] M. Z. Azemin, D. K. Kumar, T. Y. Wong, J. J. Wang, P. Mitchell, R. Kawasaki, and H. Wu. “Age-related rarefaction in the fractal dimension of retinal vessel”. In: *Neurobiology of Aging* 33(1) (2012), 194.e1–194.e4.

[83] R. Kawasaki, M. Z. Che Azemin, D. K. Kumar, A. G. Tan, G. Liew, T. Y. Wong, P. Mitchell, and J. J. Wang. “Fractal dimension of the retinal vasculature and risk of stroke: a nested case-control study”. In: *Neurology* 76 (2011), pp. 1766–1767.

[84] E. Trucco, A. Ruggeri, T. Karnowski, L. Giancardo, E. Chaum, J. P. Hubschman, B. al Diri, C. Y. Cheung, D. Wong, M. Abrámoff annd G. Lim, D. Kumar, P. Burlina, N. M. Bressler, H. Jelinek, F. Meriaudeau, G. Quellec, T. MacGillivray, and B. Dhillon. “Validating retinal fundus image analysis algorithms: issues and a proposal”. In: *Investigative Ophthalmology and Visual Science* 54(5) (2013), pp. 3546–3559.

[85] Gy. Kovács and A. Hajdu. “A Self-Calibrating Approach for the Segmentation of Retinal Vessels by Template Matching and Contour Reconstruction”. In: *Medical Image Analysis* 29(4) (2016). (IF=4.565), pp. 24–46.

[86] S. Chaudhuri, S. Chatterjee, N. Katz, M. Nelson, and M. Goldbaum. “Detection of blood vessels in retinal images using two-dimensional matched filters”. In: *IEEE Trans. Med. Imag.* 8.3 (1989), pp. 263–269.

[87] I. Liu and Y. Sun. “Recursive tracking of vascular networks in angiograms based on the detection-deletion scheme”. In: *IEEE Trans. Med. Imag.* 12.2 (1993), pp. 334–341.

[88] L. Zhou, M. S. Rzeszotarski, L. J. Singerman, and J. M. Chokreff. “The detection and quantification of retinopathy using digital angiograms”. In: *IEEE Trans. Med. Imag.* 13.4 (1994), pp. 619–626.

[89] G. G. Gardner, D. Keating, T. H. Williamson, and A. T. Elliott. “Automatic detection of diabetic retinopathy using an artificial neural network: A screening tool”. In: *Br. J. Ophthalmol.* 80 (1996), pp. 940–944.

[90] C. Sinthanayothin, J. F. Boyce, H. L. Cook, and T. H. Williamson. “Automated localisation of the optic disc, fovea and retinal blooc vessels from digital colour fundus images”. In: *Br. J. Ophthalmol.* 83 (1999), pp. 902–910.

[91] Y. A. Tolias and S. M. Panas. “A fuzzy vessel tracking algorithm for retinal images based on fuzzy clustering”. In: *IEEE Trans. Med. Imag.* 17.2 (1998), pp. 263–273.

[92] O. Chutatape, L. Zheng, and S. Krishman. “Retinal blood vessel detection and tracking by matched Gaussian and Kalman filters”. In: *IEEE Int. Conf. Eng. Biol. Soc.* Vol. 20. 1998, pp. 3144–3149.

[93] A. Can, H. Shen, J. N. Turner, H. L. Tanenbaum, and B. Roysam. “Rapid automated tracing and feature extraction from retinal fundus images using direct exploratory algorithms”. In: *IEEE Trans. Inform. Technol. Biomed.* 3.2 (1999), pp. 125–138.

[94] A. Hoover, V. Kouznetsova, and M. Goldbaum. “Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response”. In: *IEEE Trans. Med. Imag.* 19.3 (2000), pp. 203–210.

[95] M. Niemeijer, J. Staal, B. v. Ginneken, M. Loog, M. D. Abramoff, J. Fitzpatrick, and M. Sonka. "Comparative study of retinal vessel segmentation methods on a new publicly available database". In: *SPIE Med. Imag.* 5370 (2004), pp. 648–656.

[96] J. Staal, M. D. Abramoff, M. Niemeijer, M. A. Viergever, and B. v. Ginneken. "Ridge based vessel segmentation in color images of the retina". In: *IEEE Trans. Med. Imag.* 23.4 (2004), pp. 501–509.

[97] X. Jiang and D. Mojon. "Adaptive local thresholding by verification based multithreshold probing with application to vessel detection in retinal images". In: *IEEE Trans. Pattern Anal. Mach. Intell.* 25.1 (2003), pp. 131–137.

[98] M. E. Martinez-Perez, A. D. Hughes, S. A. Thom, A. A. Bharath, and K. H. Parker. "Segmentation of blood vessels from red-free and fluorescein retinal images". In: *Med. Imag. Anal.* 11 (2007), pp. 47–61.

[99] M. Vlachos and E. Dermatas. "Multi-scale retinal vessel segmentation using line tracking". In: *Comp. Med. Imag. Graph.* 34.3 (2010), pp. 213–227.

[100] J. V. B. Soares, J. J. G. Leandro, R. M. Cesar Jr, H. F. Jelinek, and M. J. Cree. "Retinal vessel segmentation using the 2D Gabor wavelet and supervised classification". In: *IEEE Trans. Med. Imag.* 25.9 (2004), pp. 1214–1222.

[101] L. Gang, O. Chutatpe, and S. M. Krishnan. "Detection and measurement of retinal vessels in fundus images using amplitude modified second-order Gaussian filter". In: *IEEE Trans. Biomed. Eng.* 49 (2002), pp. 168–172.

[102] M. Al-Rawi and H. Karajeh. "Genetic algorithm matched filter optimization for automated detection of blood vessels from digital retinal images". In: *Comput. Methods Programs Biomed.* 87 (2007), pp. 248–253.

[103] L. Wang, A. Bhalerao, and R. Wilson. "Analysis of Retinal Vasculature using a Multiresolution Hermite-Gaussian Model". In: *IEEE Tran Med Imaging* 26(2) (2007), pp. 137–152.

[104] C. A. Lupascu, D. Tegolo, and E. Trucco. "Accurate estimation of retinal vessel width using bagged decision trees and an extended multiresolution Hermite model". In: *Medical Image Analysis* 17 (2013), pp. 1164–1180.

[105] A. Bhulyan, B. Nath, J. Chua, and R. Kotagirl. "Blood vessel segmentation from color retinal images using unsupervised texture classification". In: *International Conference on Image Processing.* 2007, pp. V521–V524.

[106] E. Ricci and R. Perfetti. "Retinal blood vessel segmentation using line operators and support vector classification". In: *IEEE Trans. Med. Imag.* 26.10 (2006), pp. 1357–1365.

[107] T. McInerney and D. Terzopoulos. "T-snakes: Topology adaptive snakes". In: *Med. Imag. Anal.* 4 (2000), pp. 73–91.

[108] B. Al-Diri, A. Hunter, and D. Steel. "An active contour model for segmenting and measuring retinal vessels". In: *IEEE Trans MEd Imaging* 28(9) (2009), pp. 1488–1497.

[109] B. S. Y. Lam and H. Yan. “A novel vessel segmentation algorithm for pathological retina images based on the divergence of vector fields”. In: *IEEE Trans. Med. Imag.* 27.2 (2008), pp. 237–246.

[110] T. Walter and J. C. Klein. “Segmentation of color fundus images of the human retina: Detection of the optic disc and the vascular tree using morphological techniques”. In: *Med. Data Anal., ser. Lect. Notes in Comput. Sci.* (2001), pp. 282–287.

[111] F. Zana and J. C. Klein. “Segmentation of retinal blood vessels by combining the detection of centerlines and morphological reconstruction”. In: *IEEE Trans. Med. Imag.* 25.9 (2006), pp. 1200–1213.

[112] Z. Xiao, M. adel, and S. Bourennane. “Bayesian Method with Spatial Constraint for Retinal Vessel Segmentation”. In: *Comput. and Math. Methods in Med.* 2013 (2013), p. 9.

[113] Gy. Kovács and A. Hajdu. “Extraction of the vascular system in retina images using averaged one-dependence estimators and orientation estimation in Hidden Markov Random Fields”. In: *8th IEEE Symposium on Biomedical Imaging: From Nano to Macro*. 2011, pp. 693–696.

[114] A. Yureidini, E. Kerrien, and S. Cotin. “Robust RANSAC-based blood vessel segmentation”. In: *SPIE Med. Imag.* Ed. by D. R. Hanor and S. Ourselin. Vol. 8314. SPIE Press, 2012, p. 8314M.

[115] R. C. Wihandika and N. Suciati. “Retinal blood vessel segmentation with optic disc pixels exclusion”. In: *I. J. Imag., Graph. and Signal Proc.* 5.7 (2013), pp. 26–33.

[116] C. A. Lupascu, D. Tegolo, and E. Trucco. “FABC: Retinal Vessel Segmentation Using AdaBoost”. In: *IEEE Trans. Inf. Technol. in Biomed.* 14(5) (2010), pp. 1267–1274.

[117] A. Frangi, W. J. Niessen, K. Vincken, and M. Viergever. “Multiscale vessel enhancement filtering”. In: *in Proc. 1st MICCAI*. 1998, pp. 130–137.

[118] M. M. Fraz, P. Remagnino, A. Hoppe, and B. Uyyanonvara. “An Ensemble Classification-Based Approach Applied to Retinal Blood Vessel Segmentation”. In: *IEEE Trans. Biomed. Eng.* 59(9) (2012), pp. 2538–2548.

[119] E. Moghimirad, S. H. Rezatofighi, and H. Soltanian-Zadeh. “Retinal vessel segmentation using a multi-scale medialness function”. In: *Comput. Biol. Med.* 42(1) (2012), pp. 50–60.

[120] U. T. V. Nguyen, A. Bhuiyan, L. A. F. Park, and K. Ramamohanarao. “An effective retinal blood vessel segmentation method using multi-scale line detection”. In: *Pattern Recognit.* 46 (2013), pp. 703–715.

[121] J. I. Orlando and M. Blaschko. “Learning fully-connected CRFs for blood vessel segmentation in retinal images”. In: *Proc. of Medical Image Computing and Computer Assisted Intervention (MICCAI) 2014*. 2014, pp. 634–641.

[122] B. S. Y. Lam, Y. Gao, and A. W.-C. Liew. “General Retinal Vessel Segmentation Using Regularization-based Multi-concavity Modeling”. In: *IEEE Trans. Med. Imag.* 29(7) (2010), pp. 1369–1381.

[123] M. M. Fraz, P. Remagnino, A. Hoppe, B. Uyyanonvara, A. R. Rudnicka, C. G. Owen, and S. A. Barman. “Blood vessel segmentation methodologies in retinal images – a survey”. In: *Comput. Methods Programs Biomed.* 108.1 (2012), pp. 407–433.

[124] A. A. Youssif, A. Z. Ghalwash, and A. S. Ghoneim. “Comparative study of contrast enhancement and illumination equalization methods for retinal vasculature segmentation”. In: *Cairo International Biomedical Engineering Conference (CIBEC)*. 2006.

[125] W. Li. “Mutual information functions versus correlation functions”. In: *J. Stat. Phys* 60(5-6) (1990), pp. 823–837.

[126] Mehmet Sezgin and Bulent Sankur. “Survey over image thresholding techniques and quantitative performance evaluation”. In: *Journal of Electronic Imaging* 13.1 (2004), pp. 146–165.

[127] J. K. Udupa and P. K. Saha. “Fuzzy connectedness and image segmentation”. In: *Proceedings of the IEEE* 91.10 (2003), pp. 1649–1669.

[128] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi. “Optimization by simulated annealing”. In: *Sci.* 220.4598 (1983), pp. 671–680.

[129] A. M. Deriington and P. Lennie. “Spatial and temporal contrast sensitivities of neurones in lateral geniculate nucleus of macaque”. In: *J. Physiol.* 357.12 (1984), pp. 219–240.

[130] Gy. Kovács and A. Hajdu. “Extraction of vascular system in retina images using averaged one-dependence estimators and orientation estimation in Hidden Markov Random Fields”. In: *8th IEEE International Symposium on Biomedical Imaging (ISBI2011)*. March 30 – April 2. Chicago, USA, 2011, pp. 693–696.

[131] J. De, H. Li, and L. Cheng. “Tracing retinal vessel trees by transductive inference”. In: *BMC Bioinform.* 15.20 (2014).

[132] A. M. Mendonça and A. Campilho. “Segmentation of retinal blood vessels by combining the detection of centerlines and morphological reconstruction”. In: *IEEE Trans. Med. Imag.* 25.9 (2006), pp. 1200–1213.

[133] J. M. Lobo, A. Jlménez-Valverde, and R. Real. “AUC: a misleading measure of the performance of predictive distribution models”. In: *Global Ecology and Biogeography* 17 (2008), pp. 145–151.

[134] B. Hanczar, J. Hua, C. Sima, J. Weinstein, M. Bittner, and E. R. Dougherty. “Small-sample precision of ROC-related estimates”. In: *Bioinformatics* 26(6) (2010), pp. 822–830.

[135] L. Espona, M. J. Carreira, M. Ortega, and M. G. Penedo. “A snake for retinal vessel segmentation”. In: *Pattern Recognit. and Image Anal.* 4478 (2007), pp. 178–185.

Appendix A

Publications used in the dissertation

Journal papers

[1] Gy. Kovács and A. Hajdu. "Translation Invariance in the Polynomial Kernel Space and Its Applications in kNN Classification". In: *Neural Processing Letters* 37(2) (2013). (IF=1.237), pp. 207–233. DOI: 10.1007/s11063-012-9242-0.

[2] Gy. Kovács and A. Hajdu. "A Self-Calibrating Approach for the Segmentation of Retinal Vessels by Template Matching and Contour Reconstruction". In: *Medical Image Analysis* 29(4) (2016). (IF=4.565), pp. 24–46. DOI: 10.1016/j.media.2015.12.003.

[3] Gy. Kovács. "Matching by Monotonic Tone Mapping". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2016). submitted.

Proceedings of international conferences

[1] Gy. Kovács and A. Hajdu. "Extraction of vascular system in retina images using averaged one-dependence estimators and orientation estimation in Hidden Markov Random Fields". In: *8th IEEE International Symposium on Biomedical Imaging (ISBI2011)*. March 30 – April 2. Chicago, USA, 2011, pp. 693–696.

Proceedings of Hungarian conferences

[1] Gy. Kovács. “On the Optimal Quantization Technique of Photometric Invariant Template Matching and Its Properties in the Presence of Gaussian Noise”. In: *Proc. of KÉPAF 2015*. January 27–30. Kecskemét, Hungary, 2015, pp. 657–672.

Appendix B

Full publication list of the author

Books

[1] Gy. Kovács. *Párhuzamos programozási eszközök és összetett alkalmazásaik [Parallel programming and its complex applications]*. Typotex, 2013, p. 321. ISBN: 978-963-279-328-3.

[2] Gy. Kovács. *OpenCL [Hungarian]*. Typotex, 2013, p. 361. ISBN: 978-963-279-332-0.

[3] Gy. Kovács. *OpenCL [English]*. Typotex, 2013, p. 355. ISBN: 978-963-279-334-4.

Lecture notes

[1] Gy. Kovács. *A digitális jelfeldolgozás matematikai alapjai [Introduction to the mathematics of signal processing]*. 2014, p. 270.

Journal papers

[1] Gy. Kovács and A. Hajdu. "Translation Invariance in the Polynomial Kernel Space and Its Applications in kNN Classification". In: *Neural Processing Letters* 37(2) (2013). (IF=1.237), pp. 207–233. DOI: 10.1007/s11063-012-9242-0.

[2] M. Gulácsi, Gy. Kovács, and Z. Gulácsi. "Exact ferromagnetic ground state of pentagon chains". In: *Philosophical Magazine Letters* 94(5) (2014). (IF=1.16), pp. 269–277. DOI: 10.1080/09500839.2014.895873.

[3] M. Gulácsi, Gy. Kovács, and Z. Gulácsi. “Flat band ferromagnetism without connectivity conditions in the flat band”. In: *EPL (Europhysics Letters)* 107(5) (2014). (IF=2.269), p. 57005. DOI: 10.1209/0295-5075/107/57005.

[4] M. Gulácsi, Gy. Kovács, and Z. Gulácsi. “An extension to flat band ferromagnetism”. In: *Mod. Phys. Lett. B* 28 (2014). (IF=0.687), p. 1450220. DOI: 10.1142/S0217984914502200.

[5] Gy. Kovács and Z. Gulácsi. “Pentagon chain in external fields”. In: *Philosophical Magazine* 95(32) (2015). (IF=1.825), pp. 3674–3695. DOI: 10.1080/14786435.2015.1094191.

[6] Gy. Kovács, K. Glukhov, and Z. Gulácsi. “Quadrilateral quantum chain Hamiltonian cast in positive semidefinite form containing non-linear fermionic contributions”. In: *WSEAS Transactions on Applied and Theoretical Mechanics* 10 (2015). (IF=0.32), pp. 187–193.

[7] Gy. Kovács and A. Hajdu. “A Self-Calibrating Approach for the Segmentation of Retinal Vessels by Template Matching and Contour Reconstruction”. In: *Medical Image Analysis* 29(4) (2016). (IF=4.565), pp. 24–46. DOI: 10.1016/j.media.2015.12.003.

[8] S. Keszthelyi, T. Donkó, and G. Kovács. “Computer tomography assisted imaging analysis in damaged maize grain caused by Sitotroga cerealella”. In: *Journal of Plant Diseases and Protection* (2016). (IF=0.477), pp. 89–92. DOI: 10.1007/s41348-016-0009-0.

[9] K. Csomós, E. Kristóf, B. Jakob, I. Csomós, G. Kovács, O. Rotem, J. Hodrea, Z. Bagoly, L. Muszbek, É. Csösz, and L. Fésüs. “Protein cross-linking by chlorinated polyamines and transglutamylation stabilizes neutrophil extracellular traps”. In: *Cell Death Dis.* 7(8) (2016). (IF=5.378), e2332. DOI: 10.1038/cddis.2016.200.

[10] G. Németh, G. Kovács, A. Fazekas, and K. Palágyi. “A Method for Quantitative Comparison of 2D Skeletons”. In: *Acta Polytechnica Hungarica* (2016). (IF=0.544) Accepted.

[11] Gy. Kovács. “Matching by Monotonic Tone Mapping”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2016). submitted.

Proceedings and Abstracts

[1] Gy. Kovács, Zs. Ruttkay, and A. Fazekas. “Virtual Chess Player with Emotions”. In: *Proc. of Fourth Hungarian Conference on Computer Graphics and Geometry*. November 13–14. Budapest, Hungary, 2007, pp. 182–188. ISBN: 9789634209317.

[2] S. A. Kis, L. Trón, G. Opposits, P. Veres, Á. Pányik, G. Kovács, L. Balkay, Á. Szlávecz, J. Molnár, L. Galuska, and M. Emri. “Testing and validating of a parallel image reconstruction software package”. In: *Nucl. Med. Rev.* Vol. 10(1). 2007, p. 59.

[3] A. Fazekas, K. Palágyi, Gy. Kovács, and G. Németh. “Skeletonization Based on Metrical Neighborhood Sequences”. In: *Lecture Notes in Computer Science, Proc. of 6th International Conference on Computer Vision Systems, ICVS2008*. Vol. 5008. May 12–15. Santorini, Greece, 2008, pp. 333–342. ISBN: 9783540795469.

[4] L. Sajó, Gy. Kovács, and A. Fazekas. “An application of multi-modal human-computer interaction”. In: *Proc. of AQTR 2008*. May 22–25. Cluj-Napoca, Romania, 2008, pp. 316–319. ISBN: 978-1-4244-2576-1.

[5] Gy. Kovács and G. Németh. “Skeletonization Based on Neighborhood Sequences”. In: *Proc. of Conference of PhD Students in Computer Science*. July 2–5. Szeged, Hungary, 2008, pp. 36–37.

[6] A. Fazekas, A. Hajdu, L. Sajó, and Gy. Kovács. “A digitális képfeldolgozás területén folyó kutatások a Debreceni Egyetem Informatikai Karán”. In: *Proc. of Informatika a Felsöoktatásban ’08*. Ed. by A. Pethö and M. Herdon. August 27–29. Debrecen, Hungary, 2008, pp. 27–29. ISBN: 978-963-473-129-0.

[7] G. Németh, Gy. Kovács, K. Palágyi, and A. Fazekas. “Generalization of Morphological Skeletons using Neighborhood Sequences”. In: *Proc. of KÉPAF 2009*. January 28–30. Budapest, Hungary, 2009.

[8] Gy. Kovács, G. Opposits, S. A. Kis, L. Balkay, L. Trón, and M. Emri. “Szinogram-korrekciós módszerek fejlesztése és hibaanalízise”. In: *Proc. of Hevesy György MONT XVI. Kongresszusa*. July 2–4. Debrecen, Hungary, 2009.

[9] Gy. Kovács, Béla Kajtár, Gábor Méhes, and Attila Fazekas. “Fast Detection of Chromosome Metaphases in Digitalized Microscopic Slides”. In: *Proc. of the 6th International Symposium on Image and Signal Processing and Analysis (ISPA2009)*. September 16–18. Salzburg, Austria, 2009. ISBN: 9789531841351.

[10] Gy. Kovács, Cs. Makara, and A. Fazekas. “The Multi-modal Rock-Paper-Scissors Game”. In: *Proc. of International Conference on Intelligent Virtual Agents*. Lecture Notes in Computer Science. September 14–16. Amsterdam, The Netherlands, 2009, pp. 564–565.

[11] Gy. Kovács, J. I. Iván, A. Pányik, and A. Fazekas. “The OpenIP open source image processing library”. In: *Proc. of ACM Multimedia 2010*. October 25–29. Florence, Italy, 2010, pp. 1489–1492.

[12] S. A. Kis, G. Opposits, I. Lajtos, T. Spisák, L. Balkay, Gy. Kovács, L. Trón, and M. Emri. “Comparative analysis of image processing algorithms on quadratic and hexagonal grids”. In: *Eur. J. Nucl. Med. Mol. Imaging*. Vol. 38(Suppl. 2). 2011, S275.

[13] Gy. Kovács and A. Fazekas. “Automatikus gamma korrekció”. In: *Proc. of KÉPAF 2011*. January 25–28. Szeged, Hungary, 2011.

[14] Gy. Kovács and A. Hajdu. “Extraction of the vascular system in retina images using averaged one-dependence estimators and orientation estimation in Hidden Markov Random Fields”. In: *8th IEEE International Symposium on Biomedical Imaging (ISBI2011)*. March 30 – April 2. Chicago, USA, 2011, pp. 693–696. ISBN: 978-1-4244-4128-0.

[15] Gy. Kovács, A. Fazekas, and A. Hajdu. “Exponential contrast maximization of intensity images”. In: *Proc. of 7th International Symposium on Image and Signal Processing and Analysis*. September 4–6. Dubrovnik, Croatia, 2011, pp. 139–142. ISBN: 9789531841597.

[16] G. Milisits, T. Donkó, Gy. Kovács, G. Opposits, E. Szentirmai, and Z. Sütö. “Examination of the separate effect of egg weight and egg composition on the hatchability of hen eggs and on the hatching weight of chicks”. In: *Proc. of 24th World’s Poultry Congress 2012*. August 5–9. Salvador, Brasil, 2012.

[17] Gy. Kovács, T. Donkó, G. Milisits, Z. Sütö, E. Szentirmai, and M. Emri. “CT image analysis methods used in Hungary”. In: *Booklet of FAIM I: First Annual Conference on Body and Carcass Evaluation, Meat Quality, Software and Traceability*. September 25–26. Dublin, Ireland, 2012, p. 87.

[18] Gy. Kovács, T. Donkó, G. Milisits, Z. Sütö, E. Szentirmai, and M. Emri. “Computed Tomography Image analysis methods used in Hungary”. In: *Proceedings of Farm Animal Imaging I: First Annual Conference on Body and Carcass Evaluation, Meat Quality, Software and Traceability*. Ed. by C. Maltin, C. Craigie, and L. Bünger. September 25–26. Dublin, Ireland, 2012, pp. 60–63. ISBN: 987-0-9570709-3-6.

[19] K. Csomós, J. Hodrea, O. Rotem, Gy. Kovács, É. Katona, Zs. Bagoly, I. Csomós, Gy. Vámosi, L. Muszbek, and L. Fésüs. “Monoamines compromise Neutrophil Extracellular Trap (NET) structure and function by incorporating into cellular and NET proteins”. In: *21st ECDO Euroconference on Apoptosis "Cell death: a Biomedical paradigm"*. September 25–28. Pasteur Institute, Paris, France, 2013.

[20] Gy. Kovács, T. Donkó, M. Emri, G. Opposits, and I. Repa. “Gabor-filter based automatic removal of troughs from CT images”. In: *Booklet of FAIM II: Second Annual Conference on Body and Carcass Evaluation, Meat Quality, Software and Traceability*. October 29–30. Kaposvár, Hungary, 2013, p. 87.

[21] Gy. Kovács, T. Donkó, M. Emri, G. Opposits, and I. Repa. “Gabor-filter based automatic removal of troughs from CT images”. In: *Proceedings of Farm Animal Imaging II: Second Annual Conference on Body and Carcass Evaluation, Meat Quality, Software and Traceability*. Ed. by C. Maltin, C. Craigie, and L. Bünger. October 29–30. Kaposvár, Hungary, 2013, pp. 80–84. ISBN: 978-0-9570709-9-8.

[22] Gy. Kovács, T. Donkó, Zs. Matics, M. Emri, G. Opposits, and I. Repa. “New segmentation method for CT based selection program in rabbits”. In: *Booklet of FAIM III: Third Annual Conference on Body and Carcass Evaluation, Meat Quality, Software and Traceability*. September 25–26. Taastrup, Denmark, 2014, p. 93.

[23] Gy. Kovács, T. Donkó, Zs. Matics, M. Emri, G. Opposits, and I. Repa. “New segmentation method for CT based selection program in rabbits”. In: *Proceedings of Farm Animal Imaging III: Third Annual Conference on Body and Carcass Evaluation, Meat Quality, Software and Traceability.* Ed. by C. Maltin, C. Craigie, and L. Bünger. September 25–26. Taastrup, Denmark, 2014, pp. 102–104. ISBN: 978-0-9931063-0-9.

[24] E. Szentirmai, G. Milisits, T. Donkó, Gy. Kovács, Z. Budai, J. Ujvári, I. Repa, and Z. Sütö. “Comparison of changes in the bone mineral content and egg shell weight and egg shell ratio in brown and white egg layers during the first egg laying period”. In: *Review on Agriculture and Rural Development, Proceedings of the 13. Wellmann Oszkár international conference.* Vol. 3(1). April 24. Hódmezövásárhely, Hungary, 2014, pp. 198–203.

[25] Gy. Kovács and Z. Gulácsi. “On the exact ground states of quadrilateral chains”. In: *Proc. of International Conference "Smart functional materials shaping our future".* September 18–19. Debrecen, Hungary, 2014, pp. 130–131. ISBN: 978-963-473-724-7.

[26] Gy. Kovács, T. Donkó, Zs. Matics, B. Czakó, R. Garamvölgyi, and G. Bajzik. “On the CT based estimation of muscle weights in living rabbits”. In: *Proc. of KÉPAF 2015.* January 27–30. Kecskemét, Hungary, 2015, pp. 151–166.

[27] Z. Fazekas, A. Hajdu, I. Lázár, Gy. Kovács, B. Csákány, D. M. Calugaru, R. Shah, E. I. Adam, and S. Talu. “Influence of Using Different Segmentation Methods on the Fractal Properties of the Identified Retinal Vascular Networks in Healthy Retinas and in Retinas with Vein Occlusion”. In: *Proc. of KÉPAF 2015.* January 27–30. Kecskemét, Hungary, 2015, pp. 360–373.

[28] Gy. Kovács. “On the Optimal Quantization Technique of Photometric Invariant Template Matching and Its Properties in the Presence of Gaussian Noise”. In: *Proc. of KÉPAF 2015.* January 27–30. Kecskemét, Hungary, 2015, pp. 657–672.

[29] Gy. Kovács, T. Donkó, M. Emri, G. Opposits, R. Garamvölgyi, and G. Bajzik. “Alternatives of PLS regression for the estimation of weight from CT images”. In: *Booklet of FAIM IV: Fourth Annual Conference on Body and Carcass Evaluation, Meat Quality, Software and Traceability.* September 22–24. Edinbourgh, Scotland, 2015.

[30] Gy. Kovács, T. Donkó, M. Emri, G. Opposits, R. Garamvölgyi, and G. Bajzik. “Alternatives of PLS regression for the estimation of weight from CT images”. In: *Proceedings of Farm Animal Imaging IV: Fourth Annual Conference on Body and Carcass Evaluation, Meat Quality, Software and Traceability.* Ed. by C. Maltin, C. Craigie, and L. Bünger. September 22–24. Edinbourgh, Scotland, 2015, pp. 116–119.

[31] B. Czakó, T. Donkó, Gy. Bázf, Gy. Kovács, P. G. Szabó, Zs. Matics, Zs. Gerencsér, I. Radnai, and I. Nagy. “CT-felvételek és vágási tulajdonságok alapján becsült genetikai paraméterek pannon fehér növendék nyulakban [Genetical parameter estimation based on CT-images and slaughter properties of Pannon white rabbits]”. In: *Proc. of 27. Nyúltenyésztési Tudományos Nap*. Ed. by Zs. Matics. May 20. Kaposvár, Hungary, 2015, pp. 69–72. ISBN: 978-963-9821-84-2.

[32] B. Czakó, T. Donkó, Gy. Bázár, Gy. Kovács, Zs. Matics, Zs. Gerencsér, I. Radnai, and I. Nagy. “Genetics parameters of carcass traits in Pannon white rabbits based Computer Tomography and test slaughter”. In: *Proc. of 19. International Symposium on housing and diseases of rabbits, furproviding animals and pet animals*. Ed. by S. Hoy. May 27–28. Celle, Germany, 2015, pp. 128–134. ISBN: 978-3-8359-6344-3.

[33] T. Donkó, B. Czakó, I. Nagy, Gy. Kovács, Ö. Petneházy, P. G. Szabó, R. Kasza, Zs. Szendrö, R. Garamvólgyi, and Zs. Matics. “Nyulak teljes test zsír térfogatának becslése computer tomográffal [Estimation of whole body fat content in rabbits using Computed Tomography]”. In: *Proc. of 28. Nyúltenyésztési Tudományos Nap*. May 25. Kaposvár, Hungary, 2016, pp. 45–49. ISBN: 978-615-5599-30-9.

[34] T. Donkó, B. Czakó, I. Nagy, Gy. Kovács, Ö. Petneházy, R. Kasza, Zs. Szendrö, R. Garamvölgyi, and Zs. Matics. “Total Body Fat Content Determination by Means of Compute Tomography (CT) in Rabbits”. In: *Proceedings of the 11th World Rabbit Congress*. June 15–18. Qingdao, China, 2016, pp. 753–756.

Others

[1] Gy. Kovács, L. Sajó, and A. Fazekas. “Multi-modális gépi sakkozó – Török 2”. In: *Híradástechnika* 5 (2008), pp. 51–54.