

Mathematical Methods for the Quantification of Actin-Filaments in Microscopic Images

Dissertation

zur
Erlangung des akademischen Grades
doctor rerum naturalium (Dr. rer. nat.)
der Mathematisch-Naturwissenschaftlichen Fakultät
der Universität Rostock



vorgelegt von
Harald Birkholz, geb. am 09.06.1983 in Rostock

Erstgutachter:
Prof. Dr. rer. nat. Konrad Engel
Institut of Mathematics
University of Rostock

Zweitgutachter:
Prof. Dr. Bernhard Burgeth
Fachrichtung 6.1 Mathematik
Universität des Saarlandes

Einreichung: Rostock, den 1. November 2011
Verteidigung: Rostock, den 24. Februar 2012

Zusammenfassung

In der Zellbiologie werden konfokale, Laser-Raster-mikroskopische Aufnahmen vom Aktin-Filament humaner Osteoblasten erstellt um den Fortschritt der zellulären Entwicklung zu beurteilen. Das Aktin-Filament kann als geometrisch eingebetteter Graph von Graten im Bild-Relief modelliert werden. In der Literatur ist der Ansatz über morphologische Skelettierung einer kumulativen Niveaumenge gut dokumentiert. Der Gegenstand dieser Arbeit ist ein verfeinerter Ansatz, um genauere quantitative Aussagen über die Menge von Graten im Bild-Relief und damit über das Aktin-Filament in seiner Erscheinung in konfokalen Mikroskopieaufnahmen zu gewinnen. Dazu spielen automatische Vorverarbeitung, Markierung und Quantifizierung zusammen um insgesamt die Fähigkeit des menschlichen Betrachters anzunähern, die Filamente intuitiv richtig zu erkennen. Die erreichte Genauigkeit ist durch numerische Experimente mit mehreren Zufallsmodellen der Daten und mit Anwendungsdaten zu bereits bestätigten Hypothesen belegt.

Abstract

In cell biology confocal laser scanning microscopic images of the actin filament of human osteoblasts are produced to assess the progress of cellular development. The actin filament can be regarded as a geometrically embedded graph of bright ridges in the image relief. The approach using skeletonization of a cumulative level set is well documented in the literature. This thesis aims at an advanced approach for more accurate quantitative measurements about the morphology of the bright-ridge set of microscopic images and thus about the actin filament in its appearance in confocal laser scanning microscopic images. Therefore automatic preprocessing, tagging and quantification interplay to approximate the capabilities of the human observer to intuitively recognize the filaments correctly. The achieved accuracy is proven by numerical experiments with random models of the data and with application data for already confirmed hypotheses.

Contents

1	Introduction	5
1.1	Interdisciplinary Question	5
1.2	Innovative Computer Vision	6
1.3	Notation	7
2	Cell-Morphological Parameters	8
2.1	Input and Output of Feature Detection	9
2.2	Partial Lengths	11
2.2.1	Feature Decomposition	12
2.2.2	Reduction to matching problems	13
2.2.3	The matching algorithm in angle graphs	16
2.3	Length and Orientation	21
2.3.1	Angles on the half circle	23
2.3.2	Length-weighted angles on the half circle	26
3	Straight Ridge Tracking	30
3.1	Review of Literature	30
3.2	Feature Specification	32
3.2.1	Definition of a Vertex-Superset	35
3.2.2	Ridge Criteria	37
3.2.3	Ridge Tracking Algorithm	39
4	Denoising	42
4.1	Image Restoration with the Model of Rudin, Osher and Fatemi	42
4.2	Numerical Realization	44
4.2.1	Degrees of freedom	44
4.2.2	Duality and algorithmic approaches	46
4.2.3	Finite Differences	51
4.3	Higher Order Regularization	53
5	Mathematical Morphology	58
5.1	Basic Operations	58
5.1.1	Erosion and Dilation	59
5.1.2	Opening and Closing	59
5.1.3	Morphological Skeletonization	60
5.2	Specialized Operations	62
5.2.1	Nonlinear Filtering	62
5.2.2	Homotopic Skeletons	63
5.2.3	Ridge Detection by Skeletonization	65

6	Experimental Validation	69
6.1	Qualitative Consistency	69
6.2	Random Model	71
6.2.1	EDGE	72
6.2.2	TRAIL	74
6.2.3	ELLIPSE without and with noise	76
7	Summary and Conclusions	80

1 Introduction

Interdisciplinary work needs a common language. Proverbial a picture is worth a thousand words. Image processing can provide a common platform for collaborative efforts of distinct scientific disciplines. The situation is as follows. Cell biologists are interested in facilitating the survey of cultured cells. Mathematicians are in the need of examples for employable and aesthetic theory. Hence, there is mutual benefit in improving the foundation of biological research by finding automatic techniques for the evaluation of microscopic image data.

The data is a multitude of digital images. They expose prominent fibers forming a filament. This filament consists of actin, an important structural protein in eucaryotic cells, abbreviated as F-actin, which is selectively fluorescence stained before image acquisition. For automatic analysis the filament is modelled as a geometrically embedded graph, which is subsequently evaluated for biologically meaningful parameters. The greater challenge is to identify and implement suitable image analysis tools which define the feature as accurate as necessary to be able to derive biological parameters properly. Furthermore, these tools must be flexible and applicable for different modalities of image acquisition like contrast, noise and blur.

Cyto-skeletal actin is relevant for formation, migration and signal propagation in most eucaryotic cells. It can be observed thanks to selective staining and microscopic techniques. This allows image acquisition with strong contrast and adjustable focus. Actin often forms a well defined filament, which is as a network of interconnected and interwoven fibers within the cell body. The morphology of the filament is a strong indicator for the cellular adherence progress and thus of particular interest in biomaterial research. Providing appropriate techniques for automatic quantification of morphological parameters describing mainly the actin filament of cells observed with confocal laser scanning microscopy is the main goal of this thesis.

1.1 Interdisciplinary Question

Confocal laser scanning microscopy is an image acquisition technique with the appealing capability to collect data from a thin focal plane. This renders it especially suitable for studying cells on biomaterial surfaces. Osteoblastic cells are monitored for studies about the deterministic part of their behavior in relation to well defined environmental situations, like in a biomaterial surface of certain microstructure. The microscopic technique is based on laser-light focussed on single spots in the specimen which is fluorescence stained to highlight the cytoskeleton. The final state of the cytoskeleton is a textural network, which is inaccessible for a filament-wise quantification. The focus of this thesis is the developmental phase in which the filament is yet existent but not yet as dense as a texture.

The emphasis of this work is the accurate and efficient determination of parameters which describe expensive data. The cost of a measurement is assumed to be great enough, that nonlinear image processing is appropriate. This is the case for any research with an interdisciplinary effort for sample generation, as e.g. cell culture on structured biomaterials with appropriate cell-lines which develop F-actin. There are different staining techniques and there will be probably novel techniques in the future. A different stain or a different medium possibly changes the modalities of the inspected samples. The feature might be displayed with different perturbations or different morphology. So a technique for automatic quantification must be flexible. This is the case if it is abstract. The feature is thus only a geometrically embedded graph, which renders the solution not only applicable to different types of data, but also to different modifications of the samples by novel interdisciplinary technique. How can biological data be quantified with confocal laser scanning microscopy? Ridge detection in such images is already a challenging endeavor.

This thesis is based on cell images and a qualitative description of the wanted quantification. There is no ideal result model. This is replaced by a random model of the provided cell images. This random model is claimed to display the governing sources of errors to highlight the difficulties of recovering the ground truth by automatic quantification.

1.2 Innovative Computer Vision

In [14] there is a thorough treatment of bright ridges for a wide range of applications. Why another treatment for this application? The reason is accuracy. Image analysis is applied to digital images, but studied in terms of a deeper model enriched with a-priori knowledge. This is what happens here. Ridges are not the goal. The goal is the morphology of the set of ridge pixels. An ad hoc application of the ridge model for twice continuously differentiable functions to the data at hand does not reach sufficient accuracy for derivation of high level descriptive parameters. It is a Procrustes-bed for the real world data. This is illustrated by the analysis of connectivity with the partition into almost straight trails problem. The point of failure is the assumption on data smoothness. This assumption is avoided in the preprocesses but must be introduced for feature detection. There it is reduced to the essential smoothness perpendicular to the ridge. Consequently the data is unchanged along the ridge so the feature is more accurately detected. There are other pragmatic solutions with regard to accuracy for a limited scope of parameters [2, 26]. This work is replicated and tested against the new model. It turns out, that the requirements to image quality are stronger in the former models. Especially the problem of noise is relevant for microscopic data as a result of the magnification, see [33].

The concepts of the novel model were developed in several stages, see [5, 8,

9]. The central results concern a discussion of a split approach compared with the literature, which has innovations in every part. The treatment of quantification in Section 2 explains a heuristic which paves the way to new parameters efficiently. The feature detection in Section 3 resolves the basic problem for quantification with a discrete model under assumptions which are enforced with preprocessing methods described in Sections 4 and 5. Hence there is progress in working knowledge as well as in classical theory of mathematical image processing.

The results are presented mainly in the Sections 2, 4 and 6. Section 2 reveals combinatorial and statistical models, which advance from the state of the art in a natural way, well motivated by the biological application background. The extensive Section 4 shows new results about the efficient minimization of total variation of second order. In Section 6 a selection of numerical experiments shows the achievement of accuracy constraints of all parts of the computer vision process proposed here.

1.3 Notation

The subsequent notion of a digital image is the following. For a positive integer k let $[k] = \{1, \dots, k\}$ and, for two integers k, ℓ with $k \leq \ell$, let $[k, \ell] = \{k, k+1, \dots, \ell\}$ and $[k, \ell]_{\mathbb{R}} = \{x \in \mathbb{R} : k \leq x \leq \ell\}$. Let $H = [0, 255]$. Let an *digital image* $f \in H^{\nu \times \nu}$ be given. Although f is a matrix it is not denoted with indices but as a function as a reference to an underlying continuous image model. A digital image f is interpreted as the sampled discrete values of a function $\hat{f} : \mathbb{R}^2 \rightarrow \mathbb{R}$. The value $\lfloor t \rfloor$ is the greatest integer not greater than $t \in \mathbb{R}$.

Furthermore, let V be a finite set of points in the plane and let E be a finite set of (straight) segments in the plane with endpoints from V . Thus $G = (V, E)$ can be interpreted as an (undirected) graph that is embedded in the plane. As usual, the elements of V are called *vertices* and the elements of E *edges*.

Vectors are noted as bold symbols, e.g. \mathbf{v} and the Euclidean norm of a vector is denoted by $\|\cdot\|$. Canonically the discrete ball with radius $r > 0$ around a point $\mathbf{v} \in \mathbb{Z}^2$ of an infinite grid is noted as

$$B_r(\mathbf{v}) := \{\mathbf{w} \in \mathbb{Z}^2 : \|\mathbf{v} - \mathbf{w}\| \leq r\}.$$

For the radius $r = \sqrt{2}$ this ball degenerates to a 3×3 square. The pixels $B_{\sqrt{2}}(\mathbf{v}) \setminus \mathbf{v}$ are called the eight-neighbors of the pixel \mathbf{v} . In contrast, the neighbors $N_G(\mathbf{v})$ of a vertex $\mathbf{v} \in V$ of the graph $G = (V, E)$ is the set vertices $\mathbf{w} \in V$ with an edge $\mathbf{v}\mathbf{w} \in E$. Hence there are two distinct notions of neighbors if the vertex set of a graph consists of pixels. The correct notion becomes clear from the context.

2 Cell-Morphological Parameters

Quantification of cell-morphological parameters has a purpose in cell biology for revealing biologically relevant effects in complex chemical experiments. Accuracy on the one hand and flexibility on the other hand motivate the choice of ridge-tracking with a predictor-corrector approach and preprocessing with mathematical morphology and variational methods. This enables but also constrains the method to be controlled by input parameters. The constraint can be lifted by experience with favorable values for a stack of images with similar modalities. A good start is the experiment with a random model that simulates the image modalities in a determined setting. To this end feature detection can be thought of a method to separate image modalities from image content. Then this content is the set of output parameters with a cell-morphological meaning.

A geometrically embedded graph is both feasible and reasonable for a specific quantification of a ridge set morphology in microscopic images. For the greater accuracy the neighborhood of feature elements must be considered. A computer vision model must suit the purpose of processing a certain type of image. For example in an optical character recognition setting, there is a difference between images acquired from postal address fields of letters or of vehicle registration plates of passing cars. In this case the purpose is classification. There is also extensive knowledge about the feature to be detected, a known finite alphabet with coupled specificity. This can be seen in contrast to quantitative evaluation of biological parameters from microscopic images. In this latter setting there is greater control of image acquisition but much less is determined about the feature.

A *geometrically embedded graph* is a triple $G = (V, E, b)$ of a vertex set V , an edge set

$$E \subset \binom{V}{2}$$

and a geometrical embedding $b : V \rightarrow \Omega$, into the computational domain Ω . Equivalently the embedding is implicitly involved into the vertex set, by writing the same graph as $G' = (V', E')$ with

$$V' := \{b(v) : v \in V\}$$

and

$$E' := \{b(v)b(w) : vw \in E\}.$$

For the sake of brevity the edges without orientation are denoted without set-symbols. Hence $vw = wv \in E$.

Close at hand the geometrically embedded graph can be thought of a line process with connectivity. This line process can be analyzed for preferred alignment to a certain orientation. Furthermore the whole feature graph with its connectivity can be used to derive partial length information. In the

setting of actin filaments it is cell-morphologically meaningful to separate single fibers from each other. They are superposed and branched, which cannot be determined from the images. Still the resolution of both cases for a probable one allows for a decomposition of the feature graph into almost straight trails. These are a representation for single fibers, which are assumed to be of minimal curvature given the context.

2.1 Input and Output of Feature Detection

Processing the images acquired by microscopic techniques starts with a pre-process. Most of what can be achieved by that could also be achieved with modified image acquisition but mostly at a disadvantageous ratio of cost and effect. To this end preprocessing is theoretically optional. In practice it is not. As the feature detection is explicitly formulated, without any influence of randomness or hidden assumptions, it bases on an image model, that has to be approximately determined for every given digital image.

The first assumption is, that brightness in the image represents the filament of a biological cell. So this structure has to be stained properly, i.e. exclusively and uniformly. For actin in osteoblastic cells observed by confocal laser scanning microscopy there is the effect of background illumination of cells. Within the cell body there is unspecific illumination which might originate from filaments beyond the focal plane or below the scale of a pixel. Thus a preprocess which normalizes the background illumination is advisable. This is achieved by the top-hat transform presented in Section 5.2.1. The framework of mathematical morphology would allow for the introduction of greater a-priori knowledge. But given only that the illumination is a cloud of sub-pixel brightness or a blurry signal from outside the focal plane, the knowledge is limited. A disc-structuring element

$$g : B_r(\mathbf{0}) \rightarrow \mathbb{R}$$

which is zero everywhere is a compromise close to both disturbances.

The top-hat transform filters the loss of details by morphological opening of the digital image. These details are bright extensions into dark background which do not contain the domain of this structuring element. Filaments are assumed to be elongated and equally thin. The thickness is at the same scale as the noise. The latter is addressed by the subsequent denoising method. The input parameter for the top-hat transform is simply the radius $r > 0$ of the structuring element domain. It has to exceed the width of the widest filament but should be as small as possible to be effective.

Denoising is no greater effort. A fidelity parameter $\lambda > 0$ for the weight of a total variation of second order noise penalizer has to be chosen, see Section 4.3. As a rule of thumb this value has to be greater, the more intense the noise is. In confocal laser scanning microscopic images a dominant noise phenomenon is shot-noise, which originates from the error in counting a small

number of photons emitted by the laser excited stain binding the filament. Theoretically this noise is rather modelled by the Poisson distribution than by the normal distribution. Still variational methods based on total variation, of which total variation of second order is a modification, appears to be suitable for Poisson noise as well, see [33].

On the image thus prepared for ridge tracking the assumption of an image with exclusively brightness from stained filament holds. The remaining modality is the unequal contrast of filaments along their course. Therefore, the ridge-tracking algorithm allows for setting criteria for the contrast of vertices and edges of a graph, where a relaxed edge criterion lets edges overcome gaps of straight filaments with improper contrast. The whole method presented in Section 3.2 has several input parameters.

The blobness criterion for vertices and the ridgeness criterion for graph edges are defined by threshold parameters $\ell_0 < 0$ and $r_0 > 0$ respectively. The edge length must be restricted to an interval between $[\rho_0, \rho_1]_{\mathbb{R}}$ for computational efficiency. For the desired loop-free graph $\rho_0 > 0$. The greatest possible graph edge-length ρ_1 is a system variable to define the greatest gap of stain luminance along a single filament to be bridged. The shortest possible length ρ_0 determines which curvature of filaments is still captured, a greater curvature needs a smaller value. A lower bound for ρ_0 should be the last parameter $\alpha > 0$ which is the thickness of the filament model used for the determination of a ridgeness-measurement. So far it is a fixed input parameter. Nonetheless, it could be automatically determined by testing the ridgeness for a finite set of distinct α and choosing the greatest resulting ridgeness as the appropriate model for the graph edge in question. This option is not discussed in this thesis as the direct use destabilizes the tracking of filaments.

Given these 7 values for preprocessing and ridge tracking, the techniques described in Sections 3 to 5 produce a geometrically embedded graph. This feature model is evaluated for output parameters.

The cell-morphological parameters as output parameters of the quantification method should serve as a source for statistically resilient figures which represent a good approximation of what observers could detect manually. The most obvious observations concern parallel alignment and length of filaments. Mathematically this can be captured by the analysis of the empirical distribution of orientation weighted with length and by a decomposition of the edge set of the feature graph into curves representing the single filaments. The latter would then allow for an analysis of the empirical distribution of filament lengths, although the greatest length that occurs is already a meaningful parameter.

All parameters involve meaningful angles of feature measurements. In the determination of length, the decision, which part of the graph belongs to one filament depends on the angles between edges of the graph. Furthermore the global analysis of orientation behavior of the inner phenotype of the cells

requires accurate measurements of orientation. The data, however, is given in terms of brightness information on a regular lattice. In this lattice there are preferred orientation inherent, which superpose with any information given by the data. So the feature detection must be invariant to this bias. The feature must be defined independent from the grid, though applicable to grid-bound data.

2.2 Partial Lengths

In the feature graph, all detected filaments are joined. A vertex of degree greater than 2 can denote a junction or a crossing of filaments in the focal plane. A vertex of degree 1, however, denotes the end of a filament. From the biochemical behavior of F-actin, the molecule forming the filament can be modelled as a stiff almost straight fiber. In the following model for F-actin all vertices of degree greater than 2 are counted as crossings. For the graph this means, that all paths of neighboring edges with an almost straight course are distinct filaments. Vertices of degree 2 must be distinguished by their angle. A threshold criterion allows for the decision whether a path is possibly formed by a single filament, or it is the contact point of the ends of two distinct filaments. For a distinction of such paths one can define the **PAST**-problem for geometrically embedded graphs.

Let

$$\beta(\mathbf{u}, \mathbf{v}, \mathbf{w}) := \arccos \left(\frac{(\mathbf{u} - \mathbf{v})^T (\mathbf{w} - \mathbf{v})}{l_{\{\mathbf{uv}\}} l_{\{\mathbf{vw}\}}} \right)$$

be the angle enclosed by two neighboring edges $\mathbf{uv}, \mathbf{vw} \in E$. Let γ be a parameter derived from β by a threshold criterion

$$\gamma = \begin{cases} \beta & \text{if } \beta \geq \beta_0, \\ 0 & \text{otherwise.} \end{cases}$$

In these terms, the **PAST**-problem is to find a partition $E = \bigcup_{j=1}^k E_j$ such that

$$\sum_{j=1}^k \sum_{\{\mathbf{uv}\}, \{\mathbf{vw}\} \in E_j} \gamma(\mathbf{u}, \mathbf{v}, \mathbf{w}) \rightarrow \max$$

and for all $j = 1, \dots, k$ the j -th trail

$$E_j = \{\mathbf{v}_1 \mathbf{v}_2, \mathbf{v}_2 \mathbf{v}_3, \dots, \mathbf{v}_{k_j-1} \mathbf{v}_{k_j}\}$$

visits distinct vertices $\mathbf{v}_1, \dots, \mathbf{v}_{k_j} \in V$. There is a polynomial deterministic algorithm for finding the solution. It can be provided by a distinction of cases for the graphs with bounded maximal degree or by a matching approach for graphs with arbitrary degrees.

The information content of the decomposition allows for a quantification of the longest filament. Furthermore there is a filament number $k \in \mathbb{N}$ and a process $(t_j)_{j=1}^k$ with estimated filament lengths

$$t_j := \sum_{uv \in E_j} l_{uv}.$$

Experimentally this process produces empirical distribution which seem to be almost Pareto distributed, i.e. for the random variable X the probability

$$P(X > x) = \begin{cases} \left(\frac{x_0}{x}\right)^\kappa & \text{if } x \geq x_0, \\ 1 & \text{otherwise} \end{cases}$$

with parameters $\kappa > 0, x_0 > 0$. The lower bound x_0 is technically provided by the system variable ρ_0 , defined in Section 3.2.3. Furthermore a dominance of short filaments is in line with the biological process of permanent formation and degradation of F-actin, that proposes a pool of fragments to nurture a few longer filaments.

The next Sections 2.2.1 to 2.2.3 are devoted to a slightly more generalized version of this problem in a precise graph-theoretical language, developed in collaboration with Konrad Engel, see [7].

2.2.1 Feature Decomposition

A *trail* is a sequence $T = (e_1, \dots, e_k)$ of pairwise distinct edges such that there are vertices v_0, \dots, v_k with $e_i = v_{i-1}v_i$ for $i \in [k]$. The edge set of T is denoted by $E(T)$ and the set of *inner points* of T , i.e. $\{v_1, \dots, v_{k-1}\}$, by $\widehat{V}(T)$.

Let e, e' be two adjacent edges, i.e. e, e' have a common endpoint v . Then $\angle(e, e')$ denotes the angle that is spanned by the segments e, e' in the plane. This angle is considered as non-oriented and taken from the interval $[0, \pi]$. Let an increasing function $z : [0, \pi] \rightarrow \mathbb{R}_+$ be given. In the following three types of such functions are close at hand, depending on a threshold $\beta_0 \in [0, \pi]$:

$$z_1(\beta) := \begin{cases} \beta & \text{if } \beta \geq \beta_0, \\ 0 & \text{otherwise,} \end{cases} \quad z_2(\beta) := \begin{cases} \sin(\beta/2) & \text{if } \beta \geq \beta_0, \\ 0 & \text{otherwise,} \end{cases}$$

and

$$z_3(\beta) := \begin{cases} \max\{0, -\cos(\beta)\} & \text{if } \beta \geq \beta_0, \\ 0 & \text{otherwise.} \end{cases}$$

Note that the choice in Section 2.2 is $\gamma = z_1$. An angle β_0 helps to forbid that two segments e, e' with a common endpoint v are two parts of one larger segment if $\angle(e, e') < \beta_0$. Therefore, β_0 is chosen near to π in concrete

applications. Moreover, in the case $\angle(e, e') \geq \beta_0$, the value $z(\angle(e, e'))$ may be interpreted as a quantitative measure for the combination of e and e' to a larger segment. If the two other endpoints p, p' of e and e' lie on a unit circle around v , then $z_1(\angle(e, e'))$ gives the distance between p and p' on the circle, $z_2(\angle(e, e'))$ the half of the Euclidean distance between p and p' and $z_3(\angle(e, e'))$ the length of the projection of e' onto the line containing e if $\beta_0 \geq \pi/2$. Of course, other examples for z are possible. A function z which leads to the switching property introduced in Section 2.2.3 is sufficient.

With each trail $T = (e_1, \dots, e_k)$ a weight w is associated as follows:

$$w(T) = \sum_{i=2}^k z(\angle(e_{i-1}, e_i)).$$

This weight can be interpreted as a measure of “straightness” of the trail. The *Partition into Almost Straight Trails problem*, abbreviated as **PAST**-problem, is the following:

Find a partition of G into trails T_1, \dots, T_c , more precisely $E = E(T_1) \dot{\cup} \dots \dot{\cup} E(T_c)$, such that $\sum_{j=1}^c w(T_j)$ is maximal.

Note that the trail $T = (e_1, \dots, e_k)$ can be partitioned in the case $\angle(e_{i-1}, e_i) < \beta_0$ into two subtrails $T_1 = (e_1, \dots, e_{i-1})$ and $T_2 = (e_i, \dots, e_k)$ with $w(T) = w(T_1) + w(T_2)$. Hence w.l.o.g. it can be assumed that all trails $T = (e_1, \dots, e_k)$ from the partition have the property $\angle(e_{i-1}, e_i) \geq \beta_0$ for $i \in [2, k]$.

For illustration two examples with $z = z_1$ are studied. First consider a regular n -gone, see Figure 1. Then each angle has value $\frac{n-2}{n}\pi$ and hence the n -gone can be decomposed into only one trail, the whole n -gone, if $\frac{n-2}{n}\pi \geq \beta_0$, i.e. if $n \geq \frac{2\pi}{\pi - \beta_0}$, and has to be decomposed into n trails (edges), otherwise.

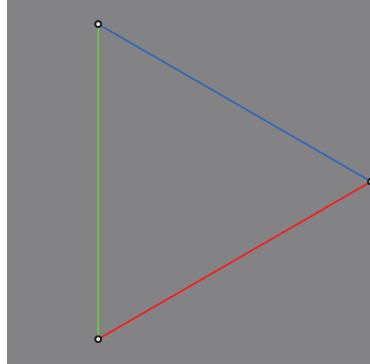
In the first case the value of the objective function is $\frac{(n-1)(n-2)}{n}\pi$ and in the second case it is 0. Now consider a regular n -star, see Figure 2, where n is odd. The largest angle is $\frac{n-1}{n}\pi$. So the star can be decomposed into $\frac{n-1}{2}$ trails of two edges and an additional trail of one edge if $\frac{n-1}{n}\pi \geq \beta_0$, i.e. if $n \geq \frac{\pi}{\pi - \beta_0}$, and has to be decomposed into n trails (edges), otherwise. In the first case the value of the objective function is $\frac{(n-1)^2}{2n}\pi$ and in the second case it is 0.

2.2.2 Reduction to matching problems

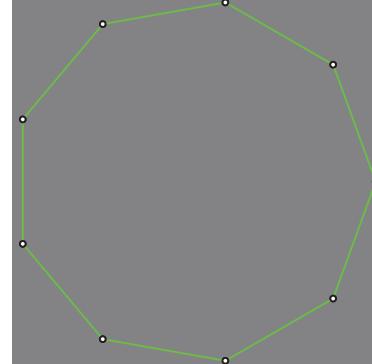
Let E_v be the set of edges having v as endpoint and let K_v be the complete graph with vertex set E_v . Let e, e' be any two vertices from K_v . The edge $\{e, e'\}$, in brief ee' , is weighted in K_v with

$$f(ee') = z(\angle(e, e')).$$

Thus a weight function f is given on the edge set of K_v .

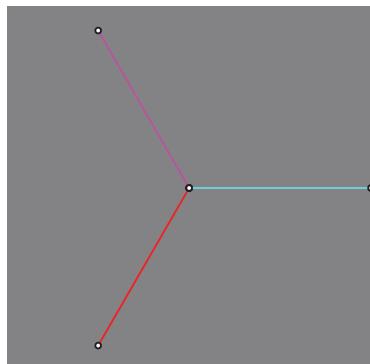


(a)

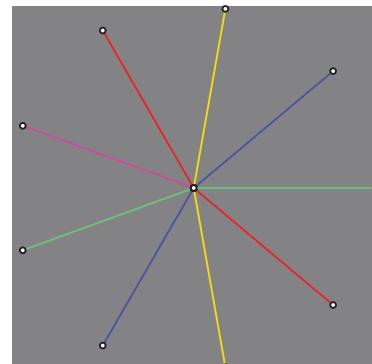


(b)

Figure 1: A **PAST** of a regular n -gone for $n = 3$ and $n = 9$ and $\beta_0 = \frac{4\pi}{6}$.



(a)



(b)

Figure 2: A **PAST** of a regular n -star for $n = 3$ and $n = 9$ and $\beta_0 = \frac{5\pi}{6}$.

Recall that a *matching* in a graph is a set of pairwise non-adjacent edges. Let for a matching M in K_v

$$f(M) = \sum_{ee' \in M} f(ee').$$

Given a partition of G into trails T_1, \dots, T_c , a matching M_v in K_v is obtained for each $v \in V$ as follows:

$$M_v = \{ee' : e, e' \in E_v \text{ and } \exists j \in [c] \text{ such that } e, e' \text{ are consecutive edges in } T_j\}.$$

Conversely, given matchings M_v in K_v for all $v \in V$, a partition of G into trails is obtained as follows: The auxiliary graph H is considered on the vertex set E where e, e' are joined by an edge if ee' belongs to M_v for some $v \in V$. Then the vertex set of each connected component of H is the edge set of a trail in G . Thus a bijection between partitions of G into trails T_1, \dots, T_c is given together with families of matchings M_v in K_v , $v \in V$, so it is possible to construct the trails algorithmically if the matchings are given.

Lemma 1. *The following formula holds:*

$$\sum_{j=1}^c w(T_j) = \sum_{v \in V} f(M_v).$$

Proof. For an inner point v of a trail T let $e_{v,T}$ be the set of the two edges e and e' of T having v as an endpoint. Note that $e_{v,T}$ is an edge in K_v . With this notation it follows

$$M_v = \{ee' : \exists j \in [c] \text{ such that } ee' = e_{v,T_j}\}.$$

Note that there cannot be two different j_1 and j_2 such that $ee' = e_{v,T_{j_1}} = e_{v,T_{j_2}}$. Moreover, if $ee' = e_{v,T_j}$, then $v \in \widehat{V}(T_j)$. Changing the order of summation gives the asserted identity:

$$\begin{aligned} \sum_{j=1}^c w(T_j) &= \sum_{j=1}^c \sum_{v \in \widehat{V}(T_j)} f(e_{v,T_j}) \\ &= \sum_{v \in V} \sum_{j \in [c]: v \in \widehat{V}(T_j)} f(e_{v,T_j}) \\ &= \sum_{v \in V} \sum_{j \in [c]: v \in \widehat{V}(T_j)} \sum_{e, e' \in E_v: ee' = e_{v,T_j}} f(ee') \\ &= \sum_{v \in V} \sum_{e, e' \in E_v} \sum_{j \in [c]: ee' = e_{v,T_j}} f(ee') \\ &= \sum_{v \in V} f(M_v). \end{aligned}$$

□

Obviously, the sum of the LHS of the identity in Lemma 1 is maximal if and only if $f(M_v)$ is maximal for each $v \in V$. Thus the **PAST**-problem can be solved by $|V|$ independent matching problems and the subsequent construction of the trails as described before. Though polynomial, matching algorithms in arbitrary graphs are sophisticated. Note that the deletion of edges of weight 0 from K_v may still result in a graph that has arbitrarily long odd holes, i.e. cycles without chords. But for the concrete geometric situation there is an essentially easier and faster matching algorithm which will be presented now.

2.2.3 The matching algorithm in angle graphs

For a single K_v , a new geometric embedding is considered. A Cartesian coordinate system is fixed in the plane and an orientation on the unit circle C around the origin o . The original graph G in the plane may be translated such that v becomes the origin. Then each edge vv' from E_v uniquely defines a point p' on C , namely the intersection of C with the ray whose origin is $v = o$ and whose direction is given by the vector $\overrightarrow{vv'}$. Note that for two edges $e' = vv'$ and $e'' = vv''$ and the corresponding points p', p'' on C , it follows $\angle(e', e'') = \angle(op', op'')$. Thus each edge $e'e''$ from K_v can be represented by $p'p''$ and the weight can be carried over:

$$f(p'p'') = f(e'e'') = z(\angle(op', op'')).$$

For the matching problem edges of weight 0 and isolated points may be deleted.

Consequently, the following problem must be solved: Let $P = \{p_1, \dots, p_n\}$ be a set of points on the unit circle C which are w.l.o.g. consecutive with respect to the given orientation. Let

$$Q = \{p_i p_j \in P : \angle(op_i, op_j) \geq \beta_0\}$$

and let $A = (P, Q)$ be the corresponding graph, where it is assumed that it does not contain isolated vertices. Such a graph is called an *angle graph*. Let $f : Q \rightarrow \mathbb{R}_+$ be the weight given by

$$f(p_i p_j) = z(\angle(op_i, op_j)) \quad \text{for } p_i p_j \in Q.$$

A maximal weighted matching in the angle graph A should be determined. By elementary geometric reasons, the graph A has the following *density property*: If p_i, p_j, p_k are three different points from P such that $p_i p_j, p_i p_k \in Q$ and if p_ℓ is a point from P lying on the circle segment between p_j and p_k that does not contain p_i , then also $p_i p_\ell \in Q$, see Figure 3. Further, the *angle graph* A has the following *switching property*: If p_i, p_j, p_k, p_ℓ are four different

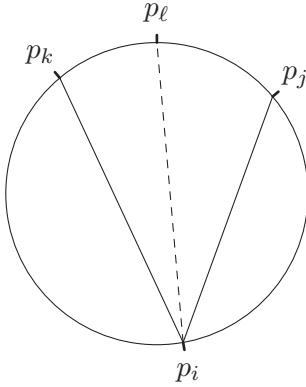


Figure 3: density property of the angle graph

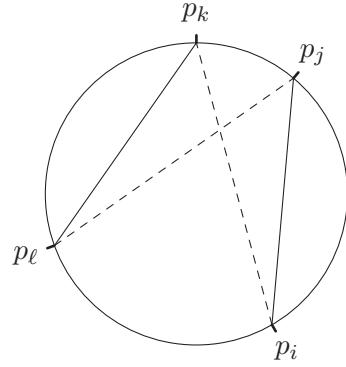


Figure 4: switching property of the angle graph

points from P such that the straight-line segments p_ip_j and p_kp_ℓ as well as p_ip_ℓ and p_jp_k do not intersect and if $p_ip_j, p_kp_\ell \in Q$ then also $p_ip_k, p_jp_\ell \in Q$, see Figure 4. If, moreover,

$$f(p_ip_j) + f(p_kp_\ell) \leq f(p_ip_k) + f(p_jp_\ell),$$

then it is said that *the weight function f has the switching property*. Recall from Section 2.2 that for $z = z_1$ the weight $f(p_ip_j)$ is the length of the shorter arc between p_i and p_j on C and for $z = z_2$ it is equal to the half of the Euclidean distance between p_i and p_j . An elementary geometric analysis shows that in these two cases f has the switching property. Also, if $z = z_3$ and $\beta_0 \geq \pi/2$, f has the switching property which can be verified either geometrically or by using the fact that the function $\cos(x + \beta) - \cos(x)$ is increasing for $\pi/2 \leq x \leq x + \beta \leq 3\pi/2$.

Theorem 1. *If f has the switching property then there is a maximal weighted matching M in the angle graph A such that any two members of M intersect as straight-line segments in the circle.*

Proof. Let M^* be a matching from the set of all maximal weighted matchings, for which the number of non-intersecting unordered pairs $\{q, q'\}$ of members of M^* is minimal. It is sufficient to show that this number is 0. Assume the contrary and let $p_i p_j$ and $p_k p_\ell$ be two non-intersecting members of M^* . W.l.o.g. also the straight-line segments $p_i p_\ell$ and $p_j p_k$ do not intersect. Let

$$M' = (M^* \setminus \{p_i p_j, p_k p_\ell\}) \cup \{p_i p_k, p_j p_\ell\}.$$

By the switching property of A and f , M' is a maximal weighted matching in A , too. Moreover, it is easy to check that the number of non-intersecting unordered pairs of members decreases after this switching by at least 1, a contradiction to the choice of M^* . \square

A matching M in A is said to be *intersecting* if any two members of M intersect. By Theorem 1 it is sufficient to determine a maximal weighted intersecting matching. First a more *specific problem* should be considered: For a fixed edge $q \in Q$, find a maximal weighted intersecting matching containing the edge q . Using a suitable rotation it can be assumed w.l.o.g. that $q = p_1 p_h$ with $2 \leq h \leq n$. Further let

$$I = [2, h - 1] \quad \text{and} \quad J = [h + 1, n].$$

A matching M in A is called an $I - J -$ matching if all members have one endpoint in I and the other endpoint in J . A maximal weighted $I - J -$ matching can be found e.g. by the Hungarian algorithm applied to the corresponding bipartite graph and this gives the solution to the specific problem: only the edge q has to be added. This matching is perhaps not intersecting, but it is known that there must be an intersecting matching of the same maximal weight by Theorem 1. The property of being intersecting is only a tool for the algorithmic treatment, it is not really needed for the original problem.

The determination of a maximal weighted $I - J -$ matching can still be accelerated using the concrete geometric situation. As before, it is sufficient to consider only intersecting $I - J -$ matchings. For $i \in I$ let $N_i = \{j \in J : p_i p_j \in Q\}$. By the density property, each N_i , $i \in I$, is an interval $N_i = [\ell_i, r_i]$ where $h + 1 \leq \ell_i \leq r_i \leq n$. Moreover, the switching property of the angle graph A implies

$$\ell_i \leq \ell_{i+1}, \quad r_i \leq r_{i+1}, \quad \ell_{i+1} \leq r_i + 1, \quad \text{for all } i \in I \setminus \{h - 1\}.$$

A dynamic programming approach is presented in Algorithm 1. The idea is to iteratively determine a maximal weighted intersecting $I - J -$ matching

having only endpoints in $[2, i]$ and $[h + 1, j]$, where $j \leq r_i$. Note that edges having one endpoint in $[2, i]$ have the other endpoint in $[h + 1, r_i]$.

Let the edges $p_i p_j$ with $i \in I$ and $j \in J$ be lexicographically ordered by

$$p_2 p_{\ell_2}, \dots, p_2 p_{r_2}, p_3 p_{\ell_3}, \dots, p_3 p_{r_3}, \dots, p_{h-1} p_{\ell_{h-1}}, \dots, p_{h-1} p_{r_{h-1}}$$

and denote them by q_1, q_2, \dots, q_m where $m = \sum_{i=2}^{h-1} r_i - \ell_i + 1$. Iteratively the algorithm stores for each i, j the weight ν_j of the actual matching, the last added edge q_k in form of a label $e_j = k$ and its predecessor $\text{pred}(e_j)$, i.e. the label of the last added edge that is not adjacent to q_k . Further the old numbers ν_j, e_j are stored which were obtained for $i - 1$ in form of numbers ν'_j, e'_j .

Algorithm 1 The matching algorithm for bipartite angle graphs

```

for all  $k = 1, \dots, m$  do
     $\text{pred}(k) := -1$ ,
end for
for all  $j = h, \dots, n$  do
     $\nu_j := 0, e_j := -1$ ,
end for
 $k := 1$ ,
for all  $i = 2, \dots, h - 1$  do
    for all  $j = \ell_i - 1, \dots, r_i$  do
         $\nu'_j := \nu_j, e'_j := e_j$ ,
    end for
    for all  $j = \ell_i, \dots, r_i$  do
        if  $\nu_j < \nu_{j-1}$  then
             $\nu_j := \nu_{j-1}, e_j := e_{j-1}$ ,
        end if
        if  $\nu_j < f(p_i p_j) + \nu'_{j-1}$  then
             $\nu_j := f(p_i p_j) + \nu'_{j-1}$ ,
             $e_j := k, \text{pred}(k) := e'_{j-1}$ ,
        end if
         $k := k + 1$ ,
    end for
end for
 $e := e_{r_{h-1}}$ ,
 $M := \emptyset$ ,
while  $e \neq -1$  do
     $M := M \cup \{q_e\}$ ,
     $e := \text{pred}(e)$ ,
end while.

```

Theorem 2. Algorithm 1 determines a maximal weighted intersecting $I-J$ -matching in time $O(|Q|)$ and $\nu_{r_{h-1}}$ equals this maximal weight.

Proof. For clarity, upper indices are added. So let ν_j^i be the corresponding values which are given after the step which is given by j in the inner loop and by i in the outer loop. For brevity, the convention $\nu_j^i = 0$ is used if $j < \ell_2$.

The statement on the time complexity is trivial. For the correctness proof subgraphs are considered which contain only the first k edges from the ordering of the edges, $k = 0, \dots, \sum_{i=2}^{h-1} r_i - \ell_i + 1$. Let $Q_k = \{p_2 p_{\ell_2}, \dots, p_i p_j\}$ be the set of these first k edges. Let $I_k = [2, i]$, $J_k = [\ell_2, j]$ and $A_k = (I_k \cup J_k, Q_k)$. By induction on k it is proven that the application of Algorithm 1 to A_k provides the maximal weight ν_j^i of an intersecting $I_k - J_k -$ matching in A_k and a matching M_j^i of weight ν_j^i . Taking in the end $e = e_j$ instead of $e = e_{r_{h-1}}$. The cases $k = 0, 1$ are trivial. Now the induction step from numbers smaller than k to k is considered. Let μ_j^i and $\bar{\mu}_j^i$ be the maximal weight of an intersecting $I_k - J_k -$ matching in A_k containing $p_i p_j$ and not containing $p_i p_j$, respectively. Clearly, $\max\{\mu_j^i, \bar{\mu}_j^i\}$ is the maximal weight of an intersecting $I_k - J_k -$ matching in A_k . By the induction hypothesis,

$$\begin{aligned}\bar{\mu}_j^i &= \max\{\nu_j^{i-1}, \nu_{j-1}^i\}, \\ \mu_j^i &= f(p_i p_j) + \nu_{j-1}^{i-1}.\end{aligned}$$

The algorithm indeed determines ν_j^i as

$$\nu_j^i = \max\{\nu_j^{i-1}, \nu_{j-1}^i, f(p_i p_j) + \nu_{j-1}^{i-1}\},$$

i.e. the actual maximal weight. If the maximum is attained for ν_j^{i-1} or ν_{j-1}^i the last added edge was not changed and hence, by the induction hypothesis M_j^i has weight ν_j^i . If the maximum is attained for $f(p_i p_j) + \nu_{j-1}^{i-1}$ it follows that $M_j^i = M_{j-1}^{i-1} \cup \{p_i p_j\}$ and, again by the induction hypothesis, M_j^i has weight $\nu_j^i = f(p_i p_j) + \nu_{j-1}^{i-1}$. \square

Now a maximal weighted matching in the angle graph can be determined as follows. Choose an edge q for which the sum of degrees of its vertices is minimal and consider the set Q' of all edges that are adjacent or equal to q . Obviously, a maximal weighted matching contains at least one edge from Q' . So Algorithm 1 is applied to all edges $q' \in Q'$. This gives $|Q'|$ maximal weighted intersecting matchings $M_{q'}$ containing the corresponding edges q' . By Theorem 1, the matching M_{q^*} with

$$f(M_{q^*}) = \max\{f(q') : q' \in Q'\}$$

is the desired matching. Let δ_L be the minimum degree of the line graph of A , i.e. the size of $Q' \setminus \{q\}$.

Corollary 1. *A maximal weighted matching in the angle graph $A = (P, Q)$ can be determined in time $O((\delta_L + 1)|Q|)$.*

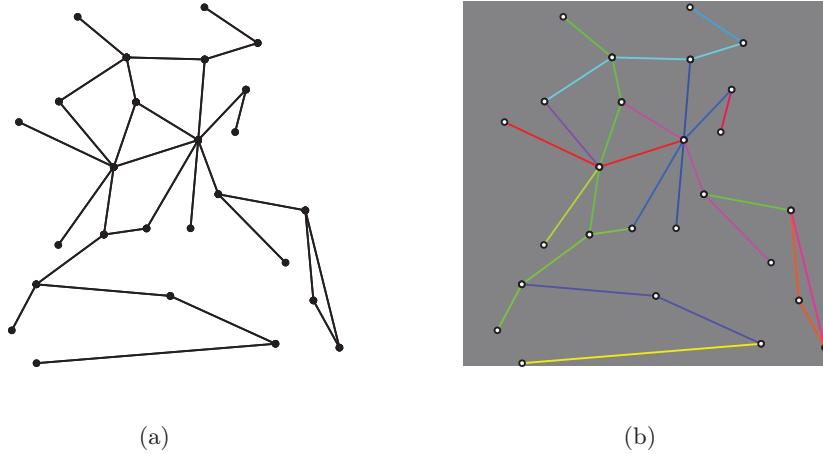


Figure 5: **PAST** of a random planar graph with angle threshold $3\pi/4$.

Note that the time complexity $O(|P||Q| + |P|^2 \log |P|)$ could be achieved with the fastest matching algorithms for arbitrary graphs, see [25]. But these algorithms are very involved and very difficult to implement.

An artificial example illustrates the **PAST**-coloring of a random graph and a graph from application in image analysis. The colors indicate different trails. The subgraph of a Delaunay-triangulation of a point-process in the unit-square in Figure 5 shows the behavior of the decomposition at vertices of maximal degree 7. Figure 6 shows the coloring for a microscopic image.

2.3 Length and Orientation

Statistics on the circle are not sufficiently studied to have a thorough treatment of the following model [16, 28]. However it can be found as an heuristic generalization and restriction of statistics on the circle.

Concerning length and orientation parameters the connectivity within the feature graph plays a minor role. Only when considering a decomposition selecting the filament of a single cell in an image showing several cells this would be relevant. Under the assumption that the feature graph represents the filament of a single cell, its alignment can be studied at the corresponding line-process. The orientation is measured by means of the angle of a vector with the horizontal axis of the image. Therefore, the arcus tangens function

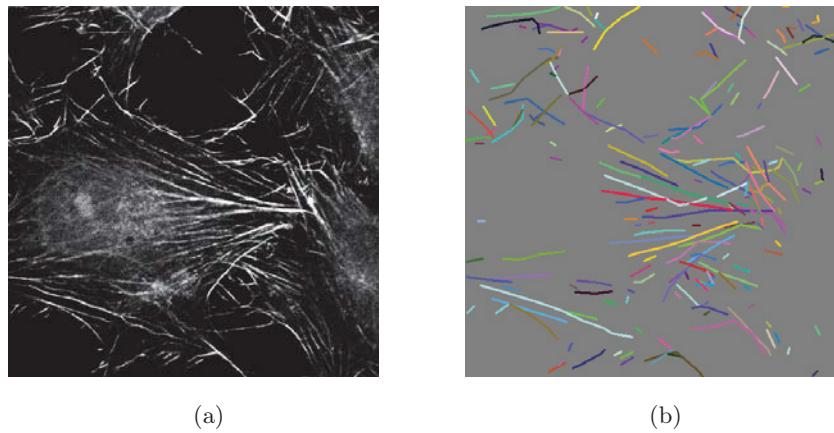


Figure 6: **PAST** of a planar graph detected in a microscopic image of the actin-filament of a human MG-63 osteoblast provided by B. Nebe (AB Zellbiologie, Rostock) with angle threshold $\arccos(-0.9)$.

on the pointed plane with two arguments

$$\arctan_2(x, y) := \begin{cases} \arctan\left(\frac{y}{x}\right) & \text{if } x > 0, \\ \pi + \arctan\left(\frac{y}{x}\right) & \text{if } x < 0, y \geq 0, \\ -\pi + \arctan\left(\frac{y}{x}\right) & \text{if } x < 0, y < 0, \\ \frac{\pi}{2} & \text{if } x = 0, y > 0, \\ -\frac{\pi}{2} & \text{if } x = 0, y < 0, \end{cases}$$

is used. Let $G = (V, E)$ be the feature graph with $V \subset \Omega \subset \mathbb{R}^2$ vertices in the plane. Then the corresponding line-process is

$$L := \{(l_{uv}, \varphi_{uv}, x_{uv}, y_{uv}) : uv \in E\},$$

i.e. the set of all lines represented by the edge set embedded with the vertices, denoted by length

$$l_{uv} := \|u - v\|,$$

orientation derived from the double angle representation adopted from [18]

$$\varphi_{uv} := \frac{1}{2} \arctan_2 \left((u_1 - v_1)^2 - (u_2 - v_2)^2, 2(u_1 - v_1)(u_2 - v_2) \right),$$

and horizontal and vertical center coordinate

$$(x_{uv}, y_{uv}) := \frac{1}{2}(u + v).$$

The parameter $T > 0$ for the total length of feature lines is then

$$T := \sum_{uv \in E} l_{uv}.$$

It facilitates to quantify the overall expression of F-actin, i.e. the cytoskeletal feature to quantify the overall progress of cellular settlement by means of the images. For quantification of the assumption of a preferred alignment a fitting of a unimodal distribution on the half-circle is suitable.

2.3.1 Angles on the half circle

The mere orientation vector $\varphi = (\varphi_{uv})_{uv \in E}$ can be tested for preferred orientations by the maximum-likelihood estimators κ, μ of a von Mises distribution. Let the functions $I_\nu(x)$ be the modified Bessel functions of first kind. The density property of the von Mises distribution uses the following fact from [42, p.181]

$$I_\alpha(x) = \frac{1}{\pi} \int_0^\pi \exp(x \cos \theta) \cos(\alpha \theta) d\theta - \frac{\sin(\alpha \pi)}{\pi} \int_0^\infty \exp(-x \cosh t - \alpha t) dt \quad (1)$$

Definition 1. Let $\Phi : [-\pi/2, \pi/2] \rightarrow \mathbb{R}$ be a random variable. The random variable Φ is said to be von Mises distributed, if

$$P(\Phi \leq \varphi_0) = \int_{-\frac{\pi}{2}}^{\varphi_0} \frac{1}{\pi I_0(\kappa)} \exp(\kappa \cos(2\varphi - 2\mu)) d\varphi.$$

The parameter $\kappa > 0$ is called concentration and $\mu \in [-\pi/2, \pi/2]$ is called mean of this distribution. The value

$$\delta = \frac{1}{2} \sqrt{1 - \frac{I_1(\kappa)}{I_0(\kappa)}}$$

is called the angular deviation of this distribution.

The angular deviation is proposed in [45, p.617] and [28]. In Definition 1 the value ranges from 0 to $\frac{1}{2}$ rad. Note that the range differs by a constant factor in the literature. This presentation adheres to [28].

Lemma 2. The function $f(\varphi|\kappa, \mu) = \frac{1}{\pi I_0(\kappa)} \exp(\kappa \cos(2\varphi - 2\mu))$ is a probability density function.

Proof. This immediately follows from equation (1) and $I_0(x) > 0$ if $x > 0$. \square

Lemma 3. The equations

$$\cos(2\mu) = \frac{1}{n} \sum_{i=1}^n \cos(2\varphi_i) \tag{2}$$

$$\sin(2\mu) = \frac{1}{n} \sum_{i=1}^n \sin(2\varphi_i) \tag{3}$$

define a maximum-likelihood estimator for μ , given realizations φ_i of n i.i.d. random variables Φ_i with von Mises distribution.

Proof. The log-likelihood function is

$$l(\mu) = \log \prod_{i=1}^n f(\varphi_i|\kappa, \mu) = \kappa \sum_{i=1}^n \cos(2\varphi_i - 2\mu) + C$$

with a well defined constant C in μ . With the estimator equations it follows that

$$l'(\mu) = -2\kappa \sum_{i=1}^n \sin(2\varphi_i - 2\mu) = -2\kappa \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n \sin(2\varphi_i - 2\varphi_j) = 0.$$

\square

An addition theorem for the sine of the sum of angles is used in the second equation. Note also that the summand is zero for $i = j$ and of alternative sign for (i, j) and (j, i) . If not both right-hand sides vanish, the system of equations (2) and (3) can be analytically solved by

$$\mu = \frac{1}{2} \arctan_2 \left(\frac{1}{n} \sum_{i=1}^n \cos(2\varphi_i), \frac{1}{n} \sum_{i=1}^n \sin(2\varphi_i) \right). \quad (4)$$

Lemma 4. *The equation*

$$\frac{I_1(\kappa)}{I_0(\kappa)} = \frac{1}{n} \sum_{i=1}^n \cos(2\varphi_i - 2\mu) \quad (5)$$

defines a maximum-likelihood estimator for κ , given μ by equation (4).

Proof. The log-likelihood function is

$$l(\kappa) = \log \prod_{i=1}^n f(\varphi_i | \kappa, \mu) = \kappa \sum_{i=1}^n (\cos(2\varphi_i - 2\mu) - \log(I_0(\kappa))) + C.$$

With the estimator equations it follows that

$$l'(\kappa) = \sum_{i=1}^n \left(\cos(2\varphi_i - 2\mu) - \frac{I'_0(\kappa)}{I_0(\kappa)} \right) = \sum_{i=1}^n (\cos(2\varphi_i - 2\mu)) - n \frac{I_1(\kappa)}{I_0(\kappa)} = 0.$$

□

Equation (5) can be solved approximately by a piecewise rational function with limited error

$$\kappa = h \left(\frac{1}{n} \sum_{i=1}^n \cos(2\varphi_i - 2\mu) \right)$$

with

$$h(s) := \begin{cases} 2s + s^3 + 5s^5/6 & \text{if } 0 < s < 0.53, \\ -0.4 + 1.39s + 0.43/(1-s) & \text{if } 0.53 \leq s < 0.85, \\ 1/(s^3 - 4s^2 + 3s) & \text{if } 0.85 \leq s < 1 \end{cases}$$

as proposed in [16]. Note that the angular deviation

$$\delta = \frac{1}{2} \sqrt{1 - \frac{I_1(\kappa)}{I_0(\kappa)}} = \frac{1}{2} \sqrt{1 - \frac{1}{n} \sum_{i=1}^n \cos(2\varphi_i - 2\mu)}$$

does not depend on an explicit solution of equation (5).

2.3.2 Length-weighted angles on the half circle

A generalization of Definition 1, which is similar to a model discussed in [23], is the following.

Definition 2. Let $(\Phi, L) : [-\pi/2, \pi/2] \times [0, 1] \rightarrow \mathbb{R}$ be a random variable. It is called length-weighted von Mises distributed if

$$P(\Phi \leq \varphi_0, L \leq \ell_0) = \int_0^{\ell_0} \int_{-\frac{\pi}{2}}^{\varphi_0} \frac{1}{\pi I_0(l\kappa)} \exp(l\kappa \cos(2\varphi - 2\mu)) d\varphi dl.$$

Given realizations $(\varphi_i, \ell_i)_{i=1}^n$ of n i.i.d. random vectors (Φ_i, L_i) , the parameters κ and μ can be estimated similar to the non-weighted case in Section 2.3.1. A line-process with arbitrary lengths fits into this model by normalization of all lengths by division by an upper bound for the length of a line. Figure 7 shows the graph of the density.

Lemma 5. The equations

$$\cos(2\mu) = \sum_{i=1}^n \ell_i \cos(2\varphi_i) \quad (6)$$

$$\sin(2\mu) = \sum_{i=1}^n \ell_i \sin(2\varphi_i) \quad (7)$$

define a maximum-likelihood estimator for μ .

Proof. The log-likelihood function is

$$l(\mu) = \log \prod_{i=1}^n f(\varphi_i, \ell_i | \kappa, \mu) = \kappa \sum_{i=1}^n \ell_i \cos(2\varphi_i - 2\mu) + C.$$

With the estimator equations it follows that

$$l'(\mu) = -2\kappa \sum_{i=1}^n \ell_i \sin(2\varphi_i - 2\mu) = -2\kappa \sum_{i=1}^n \sum_{j=1}^n \ell_i \ell_j \sin(2\varphi_i - 2\varphi_j) = 0.$$

□

If not both right-hand sides vanish, the system of equations (6) and (7) can be solved analytically by

$$\mu = \frac{1}{2} \arctan_2 \left(\sum_{i=1}^n \ell_i \cos(2\varphi_i), \sum_{i=1}^n \ell_i \sin(2\varphi_i) \right) \quad (8)$$

Lemma 6. The equation

$$\sum_{i=1}^n \ell_i \frac{I_1(\ell_i \kappa)}{I_0(\ell_i \kappa)} = \sum_{i=1}^n \ell_i \cos(2\varphi_i - 2\mu) \quad (9)$$

defines an estimator for κ , given μ by equation (8).

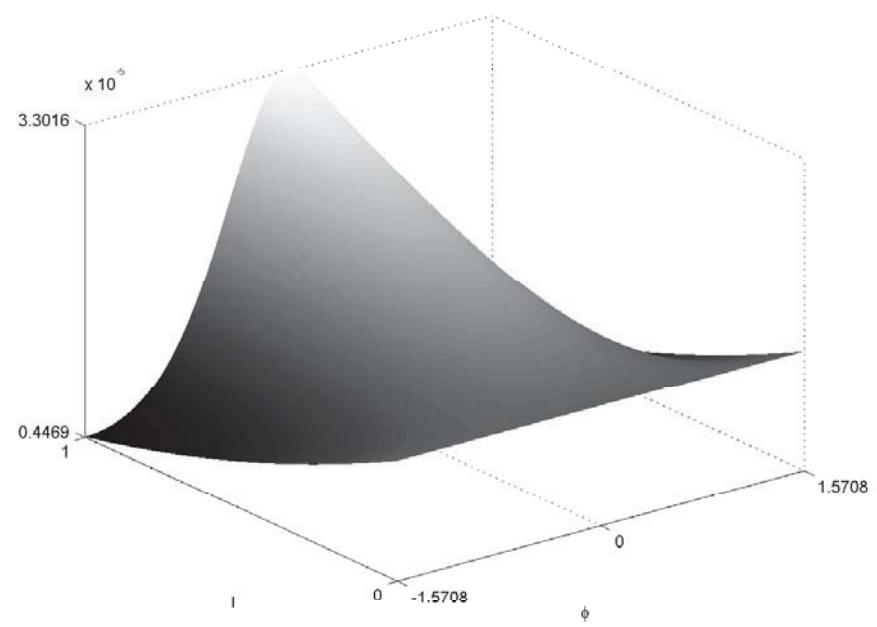


Figure 7: Density of the length-weighted von Mises distribution for $\mu = 0$ and $\kappa = 1$.

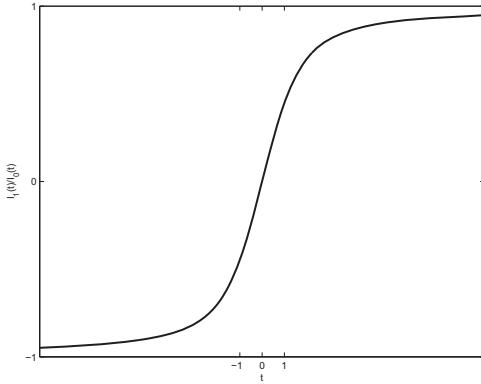


Figure 8: Graph of the function $t \mapsto \frac{I_1(t)}{I_0(t)}$.

Proof. The log-likelihood function is

$$l(\kappa) = \log \prod_{i=1}^n f(\varphi_i, \ell_i | \kappa, \mu) = \kappa \sum_{i=1}^n (\ell_i \cos(2\varphi_i - 2\mu) - \log(I_0(\ell_i \kappa))) + C.$$

With the estimator equations it follows

$$l'(\kappa) = \sum_{i=1}^n \ell_i \cos(2\varphi_i - 2\mu) - \ell_i \frac{I'_0(\ell_i \kappa)}{I_0(\ell_i \kappa)} = \sum_{i=1}^n \ell_i \cos(2\varphi_i - 2\mu) - \ell_i \frac{I_1(\ell_i \kappa)}{I_0(\ell_i \kappa)} = 0.$$

□

The existence and uniqueness of κ follows from the observation that

$$t \mapsto \frac{I_1(t)}{I_0(t)} : \mathbb{R} \leftrightarrow (-1, 1)$$

is a continuous, sigmoid, monotone increasing bijection, see Figure 8. However the solution of equation (9) is only determined numerically in the scope of this presentation. However, a length-weighted angular deviation can be defined accordingly such that κ is not needed explicitly

$$\delta := \frac{1}{2} \sqrt{\sum_{i=1}^n \ell_i \left(1 - \frac{I_1(\ell_i \kappa)}{I_0(\ell_i \kappa)} \right)} = \frac{1}{2} \sqrt{\sum_{i=1}^n \ell_i (1 - \cos(2\varphi_i - 2\mu))}.$$

This is found to be a generalization of the von Mises distribution with the choice of a fixed value $1/n$ for all n realizations of L . As an example for the weighted von-Mises distribution applied to features of microscopic images see Figure 9.

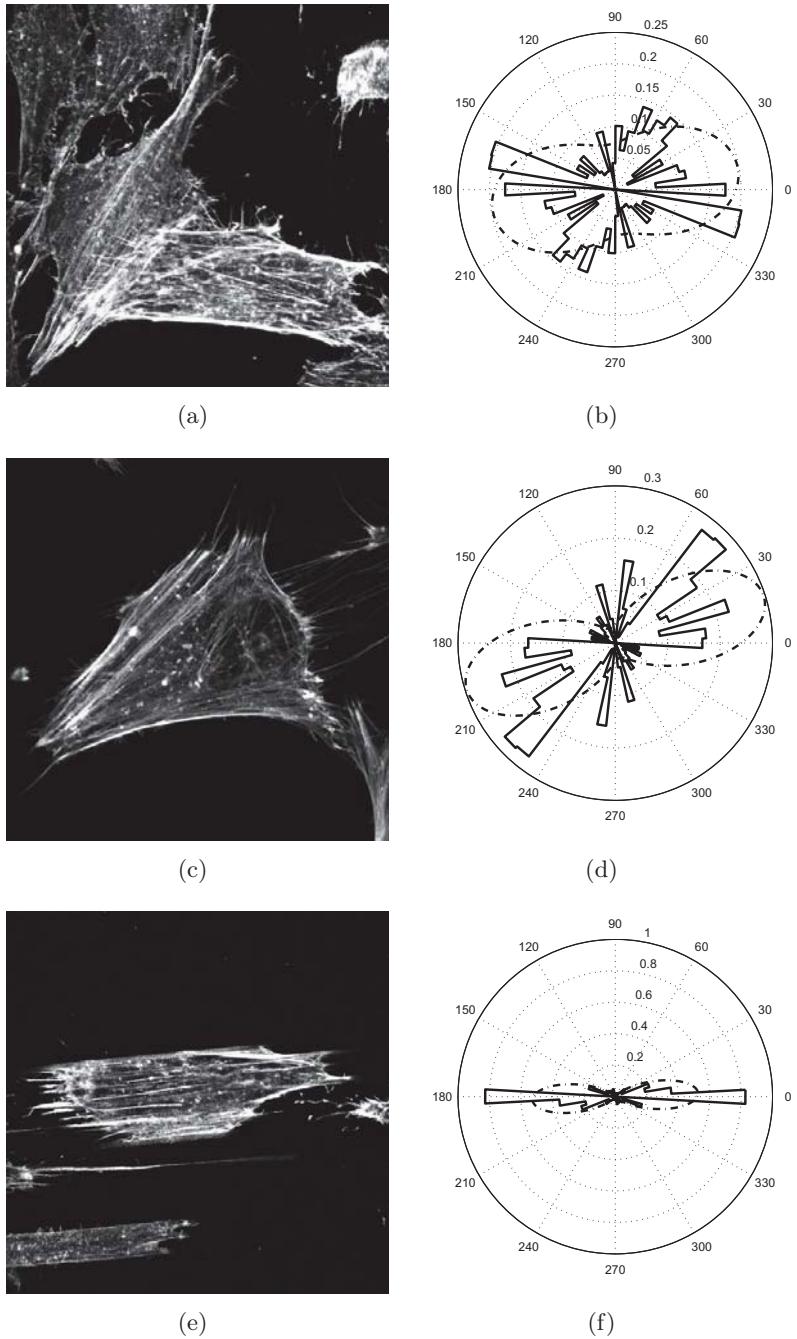


Figure 9: Illustration of parameter estimation for filament alignment in confocal laser scanning microscopic images provided by C. Matschegewski (AB Zellbiologie Rostock). The angular deviations δ are 25.7° for part (a), 22° for part (c) and 14.2° for part (e). The stepped line at the polar plots (b), (d) and (f) indicates the length-weighted histogram of edge orientation. The dash-dotted line represents the density function with the estimated parameters.

3 Straight Ridge Tracking

Ridge tracking is finding ridges under the assumption that the data only consist of ridges. This can be provided approximately by the preprocessing. A more selective preprocess comes at cost of contrast. Still the quantification by means of the **PAST** is very sensitive for errors in the connectivity. Thus the following proposals lead to a contrast stable and connectivity maximizing method. The stability is possible under the assumption of small curvature within single filaments, which also plays a role for the **PAST**.

3.1 Review of Literature

Ridges are object to ongoing research since their first explicit treatment as a sole feature in [19]. A milestone in this research is the extensive treatment of ridge localization by local differential geometric properties of an interpolated image, see [14]. It is introduced with a phenomenological note: “... I[David Eberly] have hiked many times in the Rocky Mountains of Colorado. Certainly I could point out to you the peak of a mountain from a distance. Certainly I could point out to you the ridge of a mountain from a distance. But could I recognize a ridge if I were standing on one? I would like to think so, but the answer to this question is profoundly philosophical!”

The need to believe in the argument for a positive answer to this question is avoided. In all evidence of the differential geometric rationale data with clearly observable ridges still slip from the hands of a strict definition. One reason might be the minimalist treatment of the surrounding of ridge points. This loss of context is reduced in the scale space approaches of [27] for ridge and also for image edge detection. In this the notion of scale space is still limited to linear filtering. Closer to application and with less support from analytical preliminaries this introduction of context is replaced by nonlinear filters here. The motivation builds a closed line to the former work. So far this can be thought of as suitable predictors in the sense of the predictor-corrector approach. The corrector must be more accurate and it must benefit from the predictor. An example well suitable for straight ridges will be the major innovation of the following algorithm.

A less phenomenological philosophy is behind this. Tagging automatically what the human observer can tag with his implicit knowledge of all information lost in the imaging process can be an efficient purpose of computer vision. This thesis is guided by an argument illustrated in [22, p.138]: “Wahrnehmungssysteme haben sich entwickelt, damit sich ein Lebewesen in Bezug auf die gegebenen Umweltbedingungen angemessen verhalten kann. Mit Verhalten sind dabei im Wesentlichen motorische Aktivitäten, wie Fortbewegungen unterschiedlichster Art, Greifen oder anderweitiger Glieder- und Körperbewegungen gemeint.” This is to say, perceptual systems had evolved such that organisms could behave appropriately in relation to their environ-

ment. Behavior, in this, consisted mainly of motor activities like movement of different kinds, grasping or other gestures of the body. So tagging something individually is an activity that the computer vision system should realize automatically. For hikers ridges are an efficient path to move from peak to peak.

The article [27] takes a very general point of view on ridge and also image edge detection. Scale space techniques are greatly emphasized. They are useful to overcome weaknesses of former approaches in application to macroscopic scene *classification*. In this case the feature model is not refined beyond a pixel set.

Closer to application to microscopic images, the monograph [14] forestalls the scale space idea in that a spline interpolation of arbitrary order is proposed. Furthermore the feature model is refined to a graph on adjacent inter-pixel vertices with a discrimination of cases for the introduction of the edge set. The link to medial axis features is set, with medialness measurements. This can be refined if the medialness is equipped with prior knowledge from the surrounding given by reasonable rules behind the real world object and a weak estimation in advance. This decomposition into two steps allows for a combination of different medialness-measurements with declined impact of noise in the final result.

The article [2] is very close to the application of ridge detectors for *quantification* in microscopic images. There the image is acquired via scanning electron microscopy, which allows for quite accurate segmentation by thresholding without great sorting out of data. This coarse feature detector is refined by morphological operations which result in a thin-line pixel set. By adjacency criteria and geometric post-processing this is transferred into a geometrically embedded graph representing the homotopy of the ridge-network.

Recent articles like [32] and [44] illustrate the topicality of automatic quantification of filaments in microscopic images. The work of [44] is based on the local ridge model in [14], where the false detection is improved by an extensive, adaptive post-process. This is limited to reasonable criteria for meaningful results of feature detection. The bias by a post process can well be analyzed in relation to the result of feature detection. However, the approach is reported to limit the application to a low density of filaments. The density tolerance of filament quantification for cell biology is an important criterion for an automatic approach. Actin filaments change in density and the range of observation by automatic quantification is limited by the failure in the final high density state, where the filament becomes a texture with the visible parts of the filaments as textons, see [24]. The work [32] is very close to the approach of the feature detection part of this thesis. The filament is tracked by local cross-correlation of an ideal cross-section model with the image. But notably the filter for cross correlation is sensitive to low contrast, as it only takes the absolute brightness and not the extremal nature of ridge into account. In conclusion there are drawbacks which are

claimed to be less significant for the following approach, at the expense of higher cost for combinatorial look-up and with system variables which must be chosen carefully.

3.2 Feature Specification

Let the unknown of the feature detection problem given an image be a geometrically embedded graph with pixels as nodes. The meaning of an edge is then a straight part of the bright ridge set of the image. Adjacent edges form paths which track isolated ridges or lead to bifurcations with other paths. The embedding could possibly range over the continuous image domain. However, the restriction to the pixel lattice greatly simplifies the feature detection process for the benefit of knowledge to be introduced from the context independent from a fixed interpolation scheme. Still the edge set of the graph has a geometrical meaning by the embedding. The correspondence between graph edges and pixels must thus be specified.

Figure 10 illustrates the phenomenon of ridges in images from application. Sampled over lines perpendicular to their course they are local maxima of the image relief.

Specifying the feature in the image implies the following argumentation about its detection. The global point of view is that it is the medial axis of a given set, e.g. a cumulative level-set of the image, see [2]. This is pixel accurate and strongly depends on the quality of the segmentation. The local point of view is that a *bright* ridge-point is a point in the image domain, with a *negative smallest* eigenvalue of the image Hessian and a gradient perpendicular on the corresponding eigenvector, see [14]. This is sub-pixel accurate if applied to an interpolated version of the digital image. The global point of view operates on the range with the segmentation threshold and on the domain with a fix-point criterion for a thinning operator. The local point of view operates on the function in a neighborhood.

A compromise is the following. A straight-ridge is defined as a line where the mean of the smoothed perpendicular derivative is significantly negative, represented by a rectangle with this line as its principal axis. This releases the local approach from the noise-prone Hessian-eigenvector. The global approach is equipped with a medialness criterion. The precondition is that the line is known by its end-points. These can be preselected by the merely local approach. This is advantageous, because the straight ridge less independent on contrast loss at the image acquisition process. Furthermore it is stable against artifacts persistent in nonlinear preprocesses.

The following definition models ridges from a local point of view.

Definition 3. Let $f \in C^2(\mathbb{R}^2, \mathbb{R})$. Let $\lambda_1 \leq \lambda_2$ be the eigenvalues of the Hessian $\nabla^2 f(\mathbf{x})$ of f in \mathbf{x} . Let $\mathbf{v}_1, \mathbf{v}_2 \in \mathbb{R}^2$ be the corresponding eigenvectors. The point \mathbf{x} is a ridge point if $\mathbf{v}_1^T \nabla f(\mathbf{x}) = 0$ and $\lambda_1 < 0$. The point \mathbf{x} is a

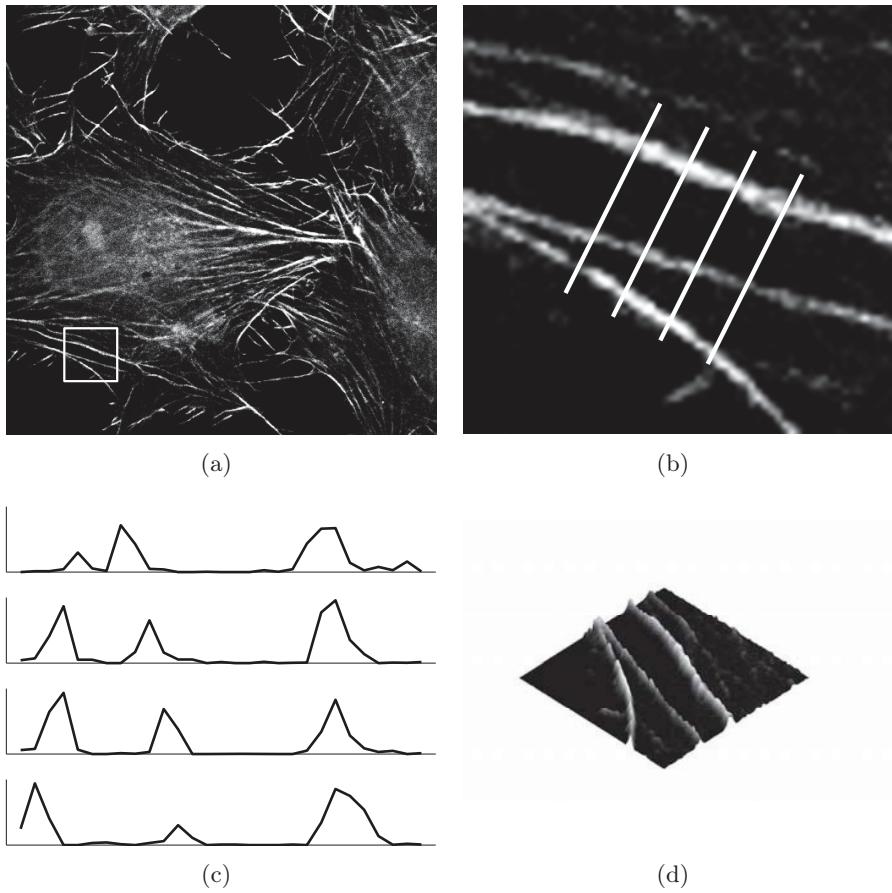


Figure 10: Selecting orthogonal profiles (c) of three ridges in a microscopic image (a) reveals the extremal nature of ridges. The one-dimensional profile over orthogonal lines through the ridges show a local maximum for each one. For a smooth image with proper contrast one could assume that this is a feasible definition of the feature. In [14] this leads to a strict definition based only on the first two derivatives of a twice differentiable interpolated image relief, see part (d).

valley point if $\mathbf{v}_2^T \nabla f(\mathbf{x}) = 0$ and $\lambda_2 > 0$.

The definition with successive generalizations and the following example with a nontrivial ridge-set for the case given here can be found in [14]. See also Figure 11.

Example: Let

$$f = x_1 x_2^2$$

noted without the spatial argument for the sake of brevity. Consequently

$$\nabla f = \begin{pmatrix} x_2^2 \\ 2x_1 x_2 \end{pmatrix}$$

and

$$\nabla^2 f = \begin{pmatrix} 0 & 2x_2 \\ 2x_2 & 2x_1 \end{pmatrix}.$$

The corresponding eigensystem is

$$\lambda_{1,2} = x_1 \mp \sqrt{x_1^2 + 4x_2^2}, \quad \mathbf{v}_1^T = (\lambda_2, -2x_2), \quad \mathbf{v}_2^T = (\lambda_1, -2x_2).$$

So the set of ridge points is the set of all (x_1, x_2) with

$$\mathbf{v}_1^T \nabla f = \lambda_1 x_2^2 - 4x_1 x_2^2 = 0 \Leftrightarrow (x_1 \geq 0 \wedge x_2 = \pm\sqrt{2}x_1) \vee x_2 = 0.$$

and

$$\lambda_1 = x_1 - \sqrt{x_1^2 + 4x_2^2} < 0 \Leftrightarrow x_1 < 0 \vee x_2 \neq 0.$$

The set of valley points given by the two conditions

$$\mathbf{v}_2^T \nabla f = \lambda_2 x_2^2 - 4x_1 x_2^2 = 0 \Leftrightarrow (x_1 \leq 0 \wedge x_2 = \pm\sqrt{2}x_1) \vee x_2 = 0$$

and

$$\lambda_2 = x_1 + \sqrt{x_1^2 + 4x_2^2} > 0 \Leftrightarrow x_1 > 0 \vee x_2 \neq 0.$$

This leads to a feature that is in practice not properly conveyed into junctions. Furthermore it strongly depends on a preprocess that eliminates false positives from greater scales or it fails to select the apparent feature. This is why this definition is at most a good predictor for accurate quantitative ridge feature analysis.

The twice continuously differentiable function f in Definition 3 has to be an approximation to apply for digital images. In practice B-splines are proposed, which is close to isotropic diffusion for small time. The B-spline approximation allows for evaluation of derivatives directly from the polynomial coefficients. An alternative are smoothed derivatives evaluated with linear filtering.

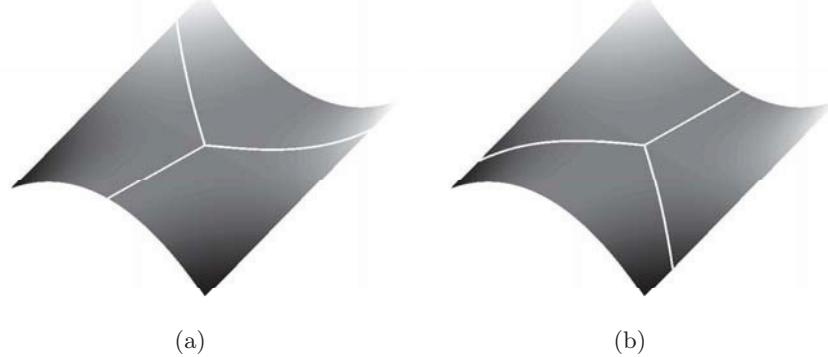


Figure 11: Ridges in (a) and valleys in (b) of the function $f(x_1, x_2) = x_1x_2^2$ near the origin as white lines on the surface, see Example 11.

With uncertainty assumed for the microscopic data, it is inaccurate to evaluate the Hessian both with B-spline interpolation and with smoothed derivatives. Small perturbations of the sampled image intensities can greatly influence the eigenspaces which are essential in Definition 3. This can lead to a loss of ridge points which would be critical for quantification of connectivity-based parameters. To stabilize connectivity, skeletonization approaches are equipped with criteria of homotopy preservation. This is only useful if the homotopy is properly represented in the segmentation in the first place. This does not have to be assumed here. The feature specification thus must be sensitive for context information. A scale-space approach proposed in [27] introduces context from all directions equally.

With a coarse segmentation of the feature, with possible loss of connectivity, the course of straight ridges is not totally random. It must be close to lines between points from the segmentation. Given the course of a predicted straight ridge in advance greatly eases the decision, whether it really is one. All along the line there must be a strong negative second directional derivative of the image function perpendicular to the line. So the burden of principal direction estimation is lifted and an important part of the context is considered. This is also easily extendable to an optimization of the thickness in feature specification.

3.2.1 Definition of a Vertex-Superset

The problem of finding suitable end points for straight ridges and selecting the most characteristic ones does not really benefit from the local ridge-finders. There are more simple and thus more stable criteria to select a set, that comprises the end-points of straight ridges. The most simple predictor would be absolute brightness. But this has the disadvantage that it strongly depends on local contrast. So an advanced choice is a blob-detector based

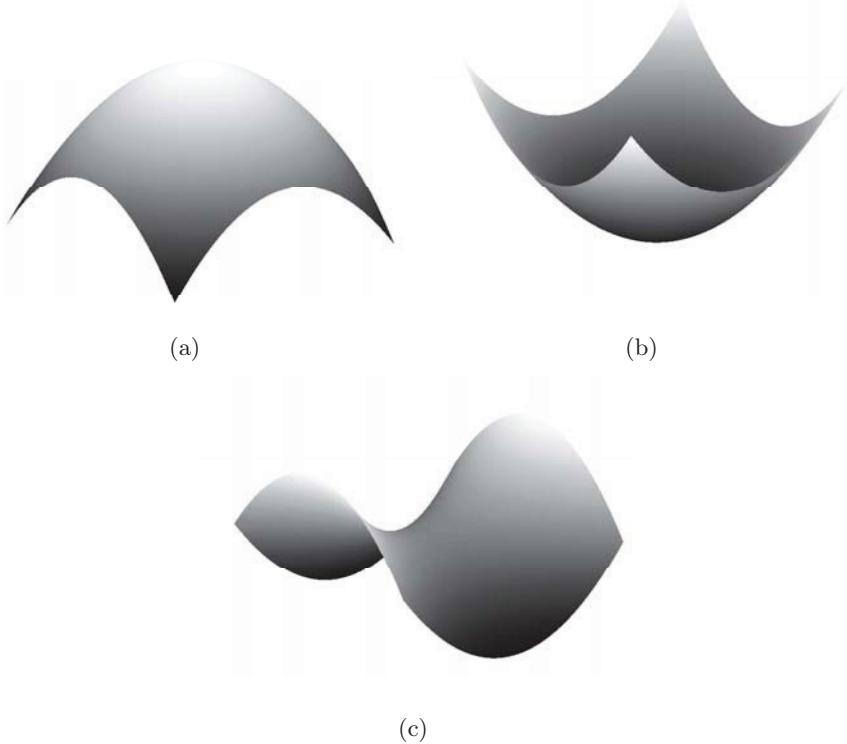


Figure 12: Second order term of Taylor-series of an sufficiently smooth image function. The cases are a negative definite Hessian (a), a positive definite Hessian (b) and an indefinite Hessian (c).

on the image Laplacian after preprocessing similar to the Marr-Hildreth-Operator. This blobness shows a strong response to the inherent shot-noise of confocal laser scanning microscopic images, so the argument against blob detectors in [27] applies. This however can be avoided by a proper preprocess. The *digital laplacian* $Lf \in \mathbb{Z}^{\nu \times \nu}$ of f is defined by

$$(Lf)(\mathbf{x}) = f(x_1 + 1, x_2) + f(x_1, x_2 + 1) \\ + f(x_1 - 1, x_2) + f(x_1, x_2 - 1) - 4f(\mathbf{x})$$

for $\mathbf{x} \in [\nu] \times [\nu]$ with the conventions

$$f(0, x_2) = f(1, x_2),$$

$$f(\nu + 1, x_2) = f(\nu, x_2)$$

$$f(x_1, 0) = f(x_1, 1)$$

and

$$f(x_1, \nu + 1) = f(x_1, \nu)$$

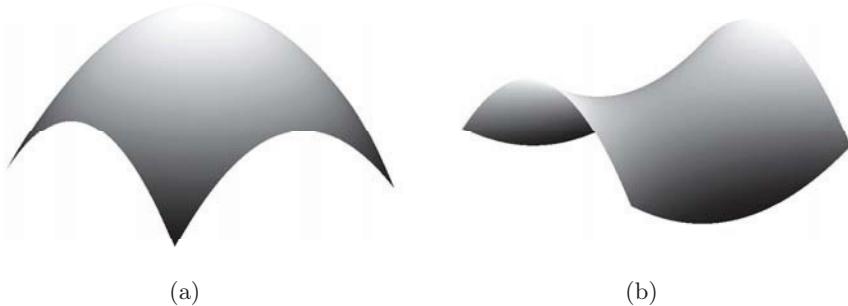


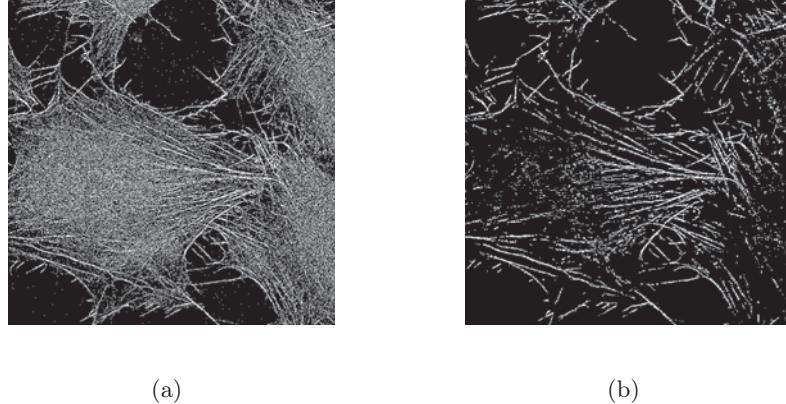
Figure 13: Second order term of Taylor-series at points selected by Laplacian thresholding. The cases are a negative definite Hessian (a) and an indefinite Hessian with dominant negative eigenvalue (b).

for off-boundary values. Thresholding with a negative value selects pixels with local curvature like in Figure 13 from all possible cases, see Figure 12. In [27] there is a more accurate notion of a predicted ridge point, which is less prone to select blobs. The selection of blobs is treated in the following Sections 4 and 5 as a nonlinear denoising problem. This is possible, because greater blobs are ignored after top-hat transform. The whole procedure of preprocessing is nonlinear bandpass filtering which accentuates ridges. An alternative would be coherence enhancing diffusion proposed in [43] and applied in [26]. The drawbacks are the difficult choice of parameters and the loss of connectivity at junctions of ridges.

3.2.2 Ridge Criteria

There are several ways to define ridgeness models in the setting, where the end-points are provided. A relaxation from Definition 3 is the use of the second derivative and the replacement of the first derivative with the prior knowledge. The direction perpendicular to the ridge can be directly assessed for one-dimensional concavity. Hence, the second directional derivative must be negative. This opens the degree of freedom to assign this criterion for every point on the segment \mathbf{xy} to the whole. It must be considered that this semi-discrete definition must make sense in a fully discrete setting. Thus an mean concavity of equidistant points along the segment seems suitable. The median experimentally does not seem appropriate. It has a tendency to overestimate the ridgeness of a suspected ridge, which crosses background after an ending with full contrast.

Let *ridgeness* of edges be measured with a change of coordinates. It requires potential end-points and represents the mean concavity of the image along the line. For given end-points \mathbf{x}, \mathbf{y} of a line in the computational domain let



(a)

(b)

Figure 14: Avoidance of blob sensitivity of Laplacian thresholding by a non-linear preprocess of successive white top-hat and total variation of second order minimization, without preprocessing (a), with preprocessing (b). The data is Figure 6(a).

$T_{\mathbf{x}, \mathbf{y}}$ be the unique affine transformation with positive determinant, which maps \mathbf{x} to the origin and \mathbf{y} to $(0, \lfloor \|\mathbf{y} - \mathbf{x}\| \rfloor)$. Brightness outside of the image domain is padded with zeros. Furthermore let

$$g_\alpha(s) := \exp\left(-\frac{s^2}{2\alpha^2}\right)$$

be a Gaussian function with inflection points at $\pm\alpha$. Its second derivative

$$h_\alpha(s) := g_\alpha''(s)$$

serves as a filter to approximate the second derivative of the smoothed image profile perpendicular to the ridge orientation, which is a measurement for concavity in case of negative values. Thus the ridgeness measurement $r_{\mathbf{x}, \mathbf{y}}$ for ridges of effective width 2α in the image f along a line between pixels \mathbf{x} and \mathbf{y} is given by

$$r_{\mathbf{x}, \mathbf{y}} = -\frac{\sum_{\xi_2=0}^{\lfloor \|\mathbf{y} - \mathbf{x}\| \rfloor} \left(\sum_{\xi_1 \in \mathbb{Z}} \hat{f}(T_{\mathbf{x}, \mathbf{y}}^{-1}(\xi_1, \xi_2)) h_{\sqrt{2}\alpha}(\xi_1) \right)}{\lfloor \|\mathbf{y} - \mathbf{x}\| \rfloor \sum_{\xi_1 \in \mathbb{Z}} |h_{\sqrt{2}\alpha}(\xi_1)|}$$

with \hat{f} the bilinear interpolated image f , see [35]. The change of sign lets a great ridgeness indicate bright ridges with strong contrast at scale α . Note that $t = \sqrt{2}\alpha$ maximizes $\int_{\mathbb{R}} h_t(s) g_\alpha(s) ds$ for $t > 0$. So the filter is chosen to fit the ridge cross-Section model g_α in the sense of optimal response. A normalization with respect to the absolute sum of the one-dimensional filter $h_{\sqrt{2}\alpha}$ renders the value less dependent on the absolute brightness.

3.2.3 Ridge Tracking Algorithm

Let $\ell_0, r_0, \alpha > 0$ be tuning parameters. The *ridge supergraph* is a weighted graph $G' = (V', E')$ where V' is the set of pixels \mathbf{x} with $-Lf(\mathbf{x}) \geq \ell_0$ and E' is the set of pixel pairs \mathbf{xy} from V' with length between $\rho_0 > 0$ and $\rho_1 > \rho_0$ weighted with their ridgeness $r_{\mathbf{xy}}$, see Section 3.2.2. For $\mathbf{x} \in V'$ let $N(\mathbf{x})$ be the neighborhood in the supergraph. For three points $\xi, \mathbf{x}, \mathbf{y} \in [n] \times [n]$ let $pr(\xi; \mathbf{x}, \mathbf{y})$ be the orthogonal projection of ξ onto the line $\overline{\mathbf{xy}}$. Moreover, let the distance ρ of ξ to $\overline{\mathbf{xy}}$ be defined as

$$\rho(\xi; \mathbf{x}, \mathbf{y}) = \begin{cases} \|\xi - pr(\xi; \mathbf{x}, \mathbf{y})\| & \text{if } pr(\xi; \mathbf{x}, \mathbf{y}) \in \overline{\mathbf{xy}}, \\ \infty & \text{otherwise,} \end{cases}$$

see Figure 15. The *representing strip* $S_{\mathbf{x}, \mathbf{y}}$ of the line $\overline{\mathbf{xy}}$ is defined as the pixel set

$$S_{\mathbf{x}, \mathbf{y}} := \{\mathbf{z} \in [n] \times [n] : \rho(\mathbf{z}; \mathbf{x}, \mathbf{y}) < \rho_2\}.$$

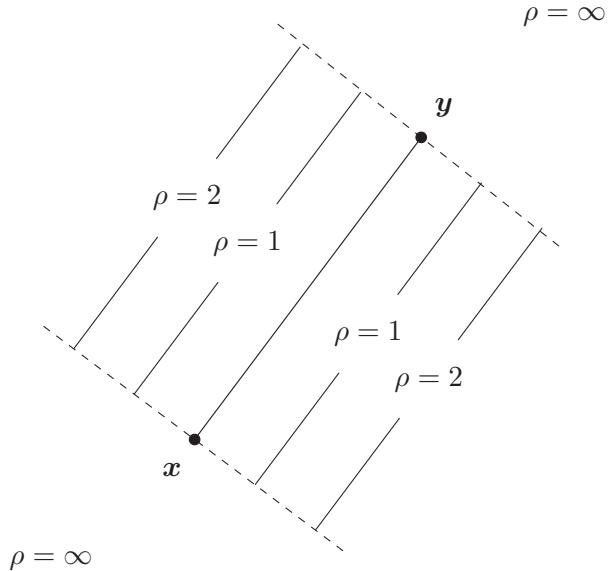


Figure 15: Rectangular distance from a segment

The following Algorithm 2 determines a subgraph (V, E) of G' representing the ridges of f as trails. The basic concept is a sequential search for an extension of a component by brightest vertices and longest edges first. Redundancy is avoided by bookkeeping of the pixels already represented by edges collected in the forbidden set Q . A pixel is assigned as represented by an edge \mathbf{xy} if it is contained in the representing strip $S_{\mathbf{x}, \mathbf{y}}$. Edges \mathbf{xy} and their end-points are added if the ridgeness $r_{\mathbf{xy}}$ exceeds r_0 . As an exception, edges with sufficient ridgeness and a representing strip that touches

the current vertex set are not added for symmetry reasons. For the sake of simplicity, the width ρ_2 of the representing strip and the minimal length of edges ρ_0 are chosen dependent on the width of the cross-section model as $2\sqrt{2}\alpha$.

Algorithm 2 Ridge tracking

```

 $Q := V := E := \emptyset.$ 
while  $V' \setminus Q \neq \emptyset$  do
    Choose  $\xi \in V' \setminus Q$  as the lexicographically first pixel with maximal
    image matrix entry.
     $L := \{\xi\}.$ 
    while  $L \neq \emptyset$  do
        Choose  $x \in L$  as the lexicographically first pixel with maximal image
        matrix entry.
         $Q := Q \cup \{x\}, L := L \setminus \{x\}.$ 
        while  $N(x) \cap (V' \setminus Q) \neq \emptyset$  do
            Choose  $y \in N(x) \cap (V' \setminus Q)$  as the lexicographically first pixel with
            maximal distance from  $x$ .
            if  $w(xy) \geq r_0, V \cap S_{x,y} = \emptyset$  then
                 $L := L \cup \{y\}.$ 
                 $V := V \cup \{x, y\}.$ 
                 $E := E \cup \{xy\}.$ 
                for all  $z \in S_{x,y} \cap (V' \setminus Q)$  do
                     $Q := Q \cup \{z\}.$ 
                    if  $z \in L$  then
                         $L := L \setminus \{z\}.$ 
                    end if
                end for
            end if
        end while
    end while
end while

```

The tracking algorithm starts a new component with the selection of the first brightest vertex. This component is then extended sequentially. In each step a starting point of potential ridges is selected from the end-points of the component and possibly connected to either an open end of the component or a point from the segmentation that is not represented by a ridge of the current graph. The end-points are chosen by distance from the starting point, the greatest distance first so that the greatest possible context is considered. The condition that the representation of a new edge is far from all current vertices ensures that ridges are not repeatedly represented. For example a bright point in the middle of a straight ridge could otherwise be chosen as a starting point for edges to both ends of the ridge. Then the component

would be completed with a graph edge that connects these end points, so that the ridge would be represented twice. This is thus avoided. Assembling the loops, however, is necessary for capturing a graph with circles properly. At the choice of pixels ξ , x and y the lexicographical order is a resort in case the ordering by brightness or distance does not select uniquely. This renders the method not rotationally invariant. The complexity of the algorithm is proportional to $|V|(\rho_1^2 - \rho_0^2)$. In the worst case all vertices are visited once and all vertices in the ring of inner radius ρ_0 and outer radius ρ_1 are tested as end-points with the time-critical evaluation of ridgeness. Hence, the adaption of the length constraints allows for tuning the performance at cost of context consideration at the ridgeness criterion.

4 Denoising

Straight ridge tracking will only be effective if the image is of sufficient quality. Mostly it is possible to acquire images with a good signal to noise ratio. However, an automatic method benefits from stabilization against perturbations. Methods for restoring the images to allow a processing independent of details of the acquisition procedure are useful and help to keep the further procedure simple.

At this point of presentation the mathematical terminology has to overcome an ambiguity. The Euclidean norm is used for discrete representations of the data, local continuous features and distances within the computational domain. In the following, the notion of convexity and saddle point is abstract and no longer of geometrical meaning for the image relief. Nevertheless the abstraction is not equipped with a new set of symbols here, as the meaning is clear from the context.

4.1 Image Restoration with the Model of Rudin, Osher and Fatemi

For image restoration linear filtering is the method both fast to implement and with a wide range of applications. This, however, comes at the expense of quality in the resulting image. Linear filtering can be introduced in different ways, e.g. as solution of a diffusion equation, as a variational problem or as spline interpolation problem. After all it is a global filter that introduces context information for each pixel at cost of inter-pixel contrast. This loss of information is hard to avoid totally. The model proposed in [34] gives an answer for edge-preserving restoration of images degraded with Gaussian white noise. This has inspired a continuing discussion about total variation penalizers in variational methods for image restoration, see e.g. [4, 11, 29, 40, 46].

Definition 4. Let $f \in L^1([0, n]_{\mathbb{R}})$. If

$$TV[f] := \sup_{\substack{0=x_0 < x_1 < \dots < x_m = n, \\ m \in \mathbb{N}}} \sum_{i=0}^{m-1} |f(x_{i+1}) - f(x_i)|$$

is finite, this value is called total variation of f . The function f is said to be of bounded variation.

Definition 4 defines a functional on the space of absolutely integrable functions. It catches the overall fluctuation of monotonicity and does not need to be finite. This leads to a generalization in a different formulation to arbitrary dimensional arguments. For image processing the two-dimensional case is most important. The geometrical meaning of first order total variation is expressed in the coarea formula as the product of the height of jump

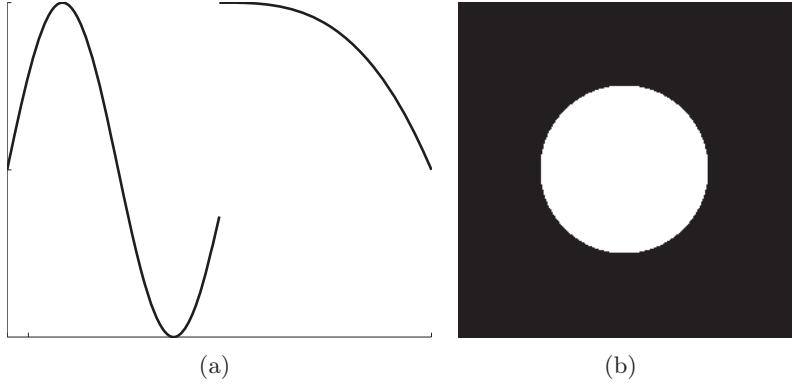


Figure 16: Image function sketches for a univariate function with maximum 1 and minimum -1 in Figure 16(a) with total variation of 6 and for a bivariate function on a square of side-length 1 in Figure 16(b) with total variation of $\frac{\pi}{2}$, which is a consequence of the coarea formula for total variation, see [12, p. 54]

with the length of the discontinuity, see Figure 16. A similar result for the second order case would be appealing.

Rudin, Osher and Fatemi proposed to penalize total variation of a function degraded by Gaussian white noise in order to reconstruct the original signal, see [34]. A continuous formulation as an unconstrained optimization problem is

$$\frac{1}{2} \int_{\Omega} (u(x) - f(x))^2 dx + \lambda \text{TV}[u] \rightarrow \min_u$$

for some fixed $\lambda > 0$. This can be extended to higher dimensional data and to higher order total-variation type penalizers. The purpose remains, the reconstruction of a function degraded by noise, while preserving prominent features such as image edges, ridges and valleys. A general discrete formulation is to find some vector $\mathbf{u} \in \mathbb{R}^n$ as a solution of the unconstrained minimization problem

$$\min \left\{ \frac{1}{2} \|\mathbf{u} - \mathbf{d}\|^2 + \sum_{i=1}^p \|P_i A \mathbf{u}\| \right\},$$

with a problem matrix $A \in \mathbb{R}^{m \times n}$ and orthogonal projectors $P_i \in \mathbb{R}^{m \times m}$, $i = 1, \dots, p$ which form a decomposition of unity $\sum_{i=1}^p P_i = I_m$ with $P_i P_j = 0$ ($i \neq j$). Note that this implies $P_i^2 = P_i$ and P_i symmetric for all i . This problem is subsequently referred to as GEN.

4.2 Numerical Realization

The non-smooth problem GEN cannot be treated directly by any standard optimization method. For finding a minimizer numerically, parabolic equations and duality approaches are popular. Duality approaches can be very efficient and accurate, because they do not need a regularization of penalizers. In the following two algorithms derived via duality observations are briefly presented.

4.2.1 Degrees of freedom

The projectors P_i must be chosen with respect to the discretization matrix A . All components of $A\mathbf{u}$ can be selectively associated by the projectors P_i . The Euclidean norms of associated components build terms of the penalizer. The important examples so far are diagonal matrices P_i , where every diagonal entry is equal to 1 at exactly one P_i and vanishes in all others. That means the penalizer is the sum of Euclidean norms of a decomposition into subsets of components of $A\mathbf{u}$. Obviously the case of total variation is that the Euclidean norm of the vertical and the horizontal finite difference is associated by projection. The Frobenius-norm of the Hessian can be modelled by the Euclidean norm of second finite differences, see [40]. For the introduction of anisotropy the 1-norm of a subset of components is modelled by projectors on only one component, such that their absolute value is penalized.

Fortunately the following efficient numerical treatment in Section 4.2.2 allows for more. An example shows that the models in use are extended nontrivially. Let $n = 1$. For $m = 1$ the only non-zero projector is $(1) \in \mathbb{R}^{1 \times 1}$. There is already a non-trivial example for $m = 2$. First for $p = 1$ the only projector is the identity matrix I_2 . For $p = 2$ it is known that $P_i = P_i^T$ and $P_1 + P_2 = I_2$, so

$$P_1 = \begin{pmatrix} a & b \\ b & c \end{pmatrix}, \quad P_2 = \begin{pmatrix} 1-a & -b \\ -b & 1-c \end{pmatrix}.$$

Solving the system $P_1 P_2 = 0$ leads to

$$P_1 = \begin{pmatrix} \frac{1}{2} + \sqrt{\frac{1}{4} - b^2} & b \\ b & \frac{1}{2} - \sqrt{\frac{1}{4} - b^2} \end{pmatrix}$$

and

$$P_2 = \begin{pmatrix} \frac{1}{2} - \sqrt{\frac{1}{4} - b^2} & -b \\ -b & \frac{1}{2} + \sqrt{\frac{1}{4} - b^2} \end{pmatrix}$$

with $0 \leq b \leq \frac{1}{2}$. The extremal cases are

$$P_1 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad P_2 = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$$

and

$$P_1 = \frac{1}{2} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, \quad P_2 = \frac{1}{2} \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}$$

Higher dimensional projectors can be found directly by solving an associated non-linear system of equations or by means of the following Structure Lemma.

Lemma 7. *Let P_1, \dots, P_p and Q_1, \dots, Q_p be projectors for GEN with $m = m_1$ and $m = m_2$ respectively. Note that zero-matrices are also admissible. Let Π be a permutation matrix of dimension m_1 . Then the row-column permuted matrices*

$$\{\Pi^T P_i \Pi : i = 1, \dots, p\}$$

are projectors for GEN with $m = m_1$, the block-diagonal matrices

$$\{\text{diag}(P_i, Q_i) : i = 1, \dots, p\}$$

are projectors for GEN with $m = m_1 + m_2$ and the Kronecker-product matrices

$$\{P_i \otimes Q_j : i, j = 1, \dots, p\}$$

are projectors for GEN with $m = m_1 m_2$.

Proof. As a sketch of the proof, the first claim follows with

$$P_i P_j = 0 \Leftrightarrow \Pi^T P_i \Pi \Pi^T P_j \Pi = \Pi^T P_i P_j \Pi = \Pi^T 0 \Pi = 0.$$

and

$$\sum_{i=1}^p \Pi^T P_i \Pi = \Pi^T \left(\sum_{i=1}^p P_i \right) \Pi = \Pi^T I \Pi = I.$$

The second claim follows by the conclusion from the blocks to the block-diagonal matrix. The third claim uses the distributivity of the Kronecker-product. \square

A meaningful model with $m = 3, p = 2$ exemplifies the consequences. Let

$$H = \begin{pmatrix} a_i & b_i \\ b_i & c_i \end{pmatrix}$$

be a symmetric matrix feature for the pixel i , like a consistently discretized Hessian, generated by A as

$$(A\mathbf{u})_{i,i+n,i+2n} = \begin{pmatrix} a_i/\sqrt{2} \\ b_i \\ c_i/\sqrt{2} \end{pmatrix}.$$

Then the penalizer-term of GEN $\|P_i A\mathbf{u}\| + \|Q_i A\mathbf{u}\|$ with

$$P_i = \frac{1}{2} \begin{pmatrix} 1 & 0 & -1 \\ 0 & 2 & 0 \\ -1 & 0 & 1 \end{pmatrix} \otimes \mathbf{e}_i \mathbf{e}_i^T$$

and

$$Q_i = \frac{1}{2} \begin{pmatrix} 1 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 1 \end{pmatrix} \otimes \mathbf{e}_i \mathbf{e}_i^T.$$

is the spectral radius of H where \mathbf{e}_i is the canonical unit vector in \mathbb{R}^n , see also Definition 6. This penalizer can be used to solve variational methods involving the total variation of second order.

4.2.2 Duality and algorithmic approaches

In the setting established above there is a general duality theorem. It is a generalization of the argument in [11] and a special case of the argument in [40]. The proof is inspired by [46, 47].

Theorem 3. Let $\mathbf{d} \in \mathbb{R}^n$. Let $A \in \mathbb{R}^{m \times n}$ be a nonzero matrix. Let $P_i \in \mathbb{R}^{m \times m}$ be orthogonal projectors onto linear subspaces which decompose the identity on \mathbb{R}^m , i.e.

$$I = \sum_{i=1}^p P_i$$

with $P_i P_j = 0$ ($i \neq j$).

Then for each optimal solution $\mathbf{x}^T \in \mathbb{R}^m$ of

$$(Q) \quad \min \frac{1}{2} \|A^T \mathbf{x} - \mathbf{d}\|^2, \quad \text{s.t.} \quad \|P_i \mathbf{x}\| \leq 1, i = 1, \dots, p,$$

the vector $\mathbf{u}^T = \mathbf{d} - A^T \mathbf{x}^T$ is an optimal solution of

$$(P) \quad \min \left\{ \frac{1}{2} \|\mathbf{u} - \mathbf{d}\|^2 + \sum_{i=1}^p \|P_i A \mathbf{u}\| \right\}.$$

Proof. The non-differentiable part of the objective function of (P) is a so-called support function

$$\sum_{i=1}^p \|P_i \mathbf{y}\| = \max_{\substack{\|P_i \mathbf{x}\| \leq 1, \\ i=1, \dots, p}} \mathbf{x}^T \mathbf{y},$$

as it is an upper bound of $\mathbf{x}^T \mathbf{y}$ such that $\|P_i \mathbf{x}\| \leq 1$ for all i because

$$\mathbf{x}^T \mathbf{y} = \sum_{i=1}^p \mathbf{x}^T P_i \mathbf{y} \leq \sum_{i=1}^p \|P_i \mathbf{x}\| \|P_i \mathbf{y}\| \leq \sum_{i=1}^p \|P_i \mathbf{y}\|,$$

and

$$\mathbf{x} := \sum_{\substack{\|P_j \mathbf{y}\| > 0, \\ j=1, \dots, p}} \frac{P_j \mathbf{y}}{\|P_j \mathbf{y}\|}$$

is a corresponding maximizer. Indeed, on the one hand

$$\|P_i \mathbf{x}\| = \left\| P_i \sum_{\substack{\|P_j \mathbf{y}\| > 0, \\ j=1, \dots, p}} \frac{P_j \mathbf{y}}{\|P_j \mathbf{y}\|} \right\| = \begin{cases} 1 & \text{if } \|P_i \mathbf{y}\| > 0, \\ 0 & \text{otherwise} \end{cases}$$

and thus

$$\|P_i \mathbf{x}\| \leq 1$$

for all $i = 1, \dots, p$, and on the other hand

$$\mathbf{x}^T \mathbf{y} = \mathbf{y}^T \sum_{\substack{\|P_j \mathbf{y}\| > 0, \\ j=1, \dots, p}} \frac{P_j \mathbf{y}}{\|P_j \mathbf{y}\|} = \sum_{\|P_j \mathbf{y}\| > 0} \|P_j \mathbf{y}\| = \sum_{j=1}^p \|P_j \mathbf{y}\|.$$

Consequently Problem (P) can be reformulated as a minimax problem

$$\begin{aligned} & \min_{\mathbf{u} \in \mathbb{R}^n} \left\{ \frac{1}{2} \|\mathbf{u} - \mathbf{d}\|^2 + \sum_{i=1}^p \|P_i A \mathbf{u}\| \right\} \\ &= \min_{\mathbf{u} \in \mathbb{R}^n} \max_{\substack{\|P_i \mathbf{x}\| \leq 1, \\ i=1, \dots, p}} \underbrace{\left\{ \frac{1}{2} \|\mathbf{u} - \mathbf{d}\|^2 + \mathbf{x}^T A \mathbf{u} \right\}}_{=: l(\mathbf{u}, \mathbf{x})}. \end{aligned}$$

Theorem 4.3.1 in Chapter VII of [20] provides the existence of saddle-points of the objective function l . This implies that the order of minimization and maximization does not matter. Four conditions are sufficient to apply the Theorem.

H1: \mathbb{R}^n is a closed, convex set.

H2: $l(\mathbf{u}, \mathbf{x})$ is convex in \mathbf{u} and concave in \mathbf{x} .

H3: For $\mathbf{x} = \mathbf{0}$, which is admissible, the objective function $l(\mathbf{u}, \mathbf{x}) \rightarrow \infty$ as $\|\mathbf{u}\| \rightarrow \infty$.

H4: The set of all admissible \mathbf{x} is bounded.

Since H1-H4 are fulfilled, there exists a non-empty set of saddle-points of l . Problem (P) can be reformulated as

$$\max_{\substack{\|P_i \mathbf{x}\| \leq 1, \\ i=1, \dots, p}} \min_{\mathbf{u} \in \mathbb{R}^n} \left\{ \frac{1}{2} \|\mathbf{u} - \mathbf{d}\|^2 + \mathbf{x}^T A \mathbf{u} \right\}.$$

There is an explicit solution to the inner problem. The objective function is strictly convex in \mathbf{u} and thus a vanishing gradient is necessary and sufficient for a minimizer, i.e.

$$\mathbf{u} - \mathbf{d} + A^T \mathbf{x} = \mathbf{0}.$$

Hence the variable \mathbf{u} can be replaced by $\mathbf{d} - A^T \mathbf{x}$. As a result, the objective simplifies to

$$\begin{aligned} & \max_{\substack{\|P_i \mathbf{x}\| \leq 1, \\ i=1, \dots, p}} \left\{ \frac{1}{2} \mathbf{x}^T A A^T \mathbf{x} + \mathbf{x}^T A (\mathbf{d} - A^T \mathbf{x}) \right\} \\ = & - \min_{\substack{\|P_i \mathbf{x}\| \leq 1, \\ i=1, \dots, p}} \left\{ \frac{1}{2} \mathbf{x}^T A A^T \mathbf{x} - \mathbf{x}^T A \mathbf{d} \right\} \\ = & - \min_{\substack{\|P_i \mathbf{x}\| \leq 1, \\ i=1, \dots, p}} \left\{ \frac{1}{2} \|A^T \mathbf{x} - \mathbf{d}\|^2 - \frac{1}{2} \|\mathbf{d}\|^2 \right\}. \end{aligned}$$

The constant $\frac{1}{2} \|\mathbf{d}\|^2$ and the sign of the objective value both do not interfere with the minimizer. Thus a solution \mathbf{x}^T of (Q) can be completed to a saddle-point $(\mathbf{d} - A^T \mathbf{x}^T, \mathbf{x}^T)$ of l , of which the first component $\mathbf{u}^T = \mathbf{d} - A^T \mathbf{x}^T$ is a minimizer of (P). \square

With the Duality Theorem in mind, a semi-implicit steepest descend algorithm can be derived, referred to as SISD. Suppose that $A = \Lambda B$ with a parameter matrix $\Lambda = \sum_{i=1}^p \lambda_i P_i$ and $\lambda_i > 0$ for all i . Note that $\Lambda^{-1} = \sum_{i=1}^p \lambda_i^{-1} P_i$. The factorization of A is helpful to improve the convergence of SISD, because the maximal step-size of the iteration must be not greater than one over the squared spectral-norm of B , see [11] and [40]. Problem (Q) can be rewritten as

$$(R) \quad \min \|B^T \mathbf{y} - \mathbf{d}\|^2, \quad \text{s.t.} \quad \|P_i \mathbf{y}\| \leq \lambda_i, i = 1, \dots, p$$

of which the optimal \mathbf{x} is recovered as $\Lambda^{-1} \mathbf{y}$. The Lagrangian of (R) with squared constraint inequalities is

$$L(\mathbf{y}, \boldsymbol{\alpha}) = \|B^T \mathbf{y} - \mathbf{d}\|^2 + \sum_{i=1}^p \alpha_i (\|P_i \mathbf{y}\|^2 - \lambda_i^2).$$

The weak Slater assumption is fulfilled because $0 = \|P_i \mathbf{0}\| < \lambda_i$ for all i . Thus the Karush-Kuhn-Tucker conditions are necessary and sufficient for optimality. The existence of some $\alpha_i \geq 0$ such that

$$\alpha_i (\|P_i \mathbf{y}\|^2 - \lambda_i^2) = 0 \quad (i = 1, \dots, p) \tag{10}$$

is provided and, at the optimal \mathbf{y} , the gradient vanishes:

$$\nabla_{\mathbf{y}} L(\mathbf{y}, \boldsymbol{\alpha}) = 2B(B^T \mathbf{y} - \mathbf{d}) + \sum_{i=1}^p 2\alpha_i P_i \mathbf{y} = \mathbf{0}. \tag{11}$$

A distinction of two cases shows that

$$\alpha_i = \frac{\|P_i \mathbf{z}\|}{\lambda_i} \tag{12}$$

with $\mathbf{z} = B(B^T \mathbf{y} - \mathbf{d})$: Indeed, if $\alpha_i > 0$ then $\|P_i \mathbf{y}\| = \lambda_i$ can be concluded from (10). Furthermore after multiplication with P_i equation (11) results in

$$P_i \mathbf{z} = -\alpha_i P_i \mathbf{y}$$

and thus

$$\|P_i \mathbf{z}\| = \alpha_i \|P_i \mathbf{y}\| = \alpha_i \lambda_i.$$

If, alternatively, $\alpha_i = 0$ the result is $P_i \mathbf{z} = \mathbf{0}$ from (11). Thus in this case

$$\alpha_i = 0 = \frac{\|P_i \mathbf{z}\|}{\lambda_i},$$

too. Hence the semi-implicit steepest descent Algorithm 3 (SISD) in line with [11] is proposed.

Algorithm 3 Semi-implicit steepest descent algorithm for GEN (SISD)

Input: Data vector $\mathbf{d} \in \mathbb{R}^n$, problem matrix $A = \Lambda B$, parameter-matrix

$\Lambda = \sum_{i=1}^p \lambda_i P_i$, fidelity parameters $\lambda_i > 0$, discretization matrix $B \in \mathbb{R}^{m \times n}$, step-size $\tau > 0$, step-count $N \in \mathbb{N}$.

$\mathbf{y}^0 := \mathbf{0} \in \mathbb{R}^m$.

for $k = 0, \dots, N - 1$ **do**

$$\begin{aligned} \mathbf{z}^k &:= B(B^T \mathbf{y}^k - \mathbf{d}). \\ \mathbf{y}^{k+1} &:= \left(I_m + \tau \sum_{i=1}^p \frac{\|P_i \mathbf{z}^k\|}{\lambda_i} P_i \right)^{-1} (\mathbf{y}^k - \tau \mathbf{z}^k). \end{aligned}$$

end for

Output: simplified data vector $\mathbf{u} := \mathbf{d} - B^T \mathbf{y}^N \in \mathbb{R}^n$.

The semi-implicitness is in the descent direction $-\nabla_{\mathbf{y}} L(\mathbf{y}, \boldsymbol{\alpha}(\mathbf{y}))$, which in the k -th step partly depends on the $(k + 1)$ -th step. This results in the matrix-inversion at the \mathbf{y} update. All Lagrange-multipliers α_i are chosen as a function of \mathbf{y} using (12).

The algorithm SISD can be proven to converge, see [11], but its performance is inferior to current algorithms for total variation denoising. Thus the model GEN is transferred into a more efficient algorithm (PDHG), which however lacks a convergence result in this formulation. Following the lines of [46] the algorithm is applied to the minimax-problem

$$\min_{\mathbf{u} \in \mathbb{R}^n} \max_{\substack{\|\mathbf{x}\| \leq 1, \\ i=1, \dots, p}} l(\mathbf{u}, \mathbf{x})$$

for

$$l(\mathbf{u}, \mathbf{x}) = \frac{1}{2} \|\mathbf{u} - \mathbf{d}\|^2 + \mathbf{x}^T A \mathbf{u}.$$

The objective l appears in the Duality Theorem 3 as an intermediate step, which leads to a rewritten problem with an analytical solution to the maximization. Similar to the proposal in [46] for a special case of GEN this step is replaced by an approximation now, which experimentally benefits the convergence, see [6]. A dual step for the inner maximization is switched with a primal step for the outer minimization. In every dual step the constraints are enforced by projection of the iterate onto the feasible set. The steps themselves are explicit gradient-ascent for the primal and gradient-descent for the dual optimization. The primal step can also be regarded as a weighted average of the current primal iterate \mathbf{u}^k and the primal counterpart of the current dual iterate \mathbf{x}^k . Briefly this can be formulated as Algorithm 4.

Algorithm 4 Primal-Dual Hybrid Gradient Algorithm for GEN (PDHG)

Input: Data vector $\mathbf{d} \in \mathbb{R}^n$, problem matrix A , step-count $N \in \mathbb{N}$, series $(\tau^k)_{k=1}^N, (\theta_k)_{k=1}^N$ of dual and primal step-sizes.
 $\mathbf{x}^0 := \mathbf{0} \in \mathbb{R}^m$.
 $\mathbf{u}^0 := \mathbf{d}$.

for $k = 0, \dots, N - 1$ **do**

$$\begin{aligned}\mathbf{y}^{k+1} &:= \mathbf{x}^k + \tau^k A \mathbf{u}^k \\ \mathbf{x}^{k+1} &:= \sum_{i=1}^p \frac{1}{\max\{1, \|P_i \mathbf{y}^{k+1}\|\}} P_i \mathbf{y}^{k+1} \\ \mathbf{u}^{k+1} &:= (1 - \theta^k) \mathbf{u}^k + \theta^k (\mathbf{d} - A^T \mathbf{x}^{k+1})\end{aligned}$$

end for

Output: simplified data vector \mathbf{u}^N .

A key ingredient for this to work is the series of step-sizes. A fixed choice of this series allows for sufficient experimental results. In [46] the proposal is

$$\tau^k = \frac{1}{5} + \frac{2}{25} k$$

and

$$\theta^k = \frac{1}{\tau^k} \left(\frac{1}{2} - \frac{5}{15 + k} \right).$$

In [38] a convergence claim is formulated without proof. The condition $\tau\theta \leq 1/\|A\|^2$ was sufficient for convergence. As $\|A_{\text{TV}_1}\| = 2\lambda$ and $\|A_{\text{TV}_2}\| = 4\lambda$ these step-widths allow for $\lambda \in [0, 1/2]$ for total variation of first order and $\lambda \in [0, 1/8]$ for total variation of second order. However, convergence is observed already for $\tau\theta < 1/\|A\|$.

4.2.3 Finite Differences

Finite differences allow for discretization of differential operators, which are consistent for dense sampling. The error is quantified by the step-width order of the leading error term in the polynomial approximation at a sampling point. For a univariate function $f \in C^3(\mathbb{R})$ the first derivative $f'(x)$ can be approximated by the following rationale. The Taylor-expansion of f in $x \pm h$ for positive h is

$$f(x \pm h) = f(x) + hf'(x) + \frac{h^2}{2}f''(x) + O(h^3)$$

for $h \rightarrow 0$. By rearrangement of these equations one can conclude that

$$f'(x) = \frac{f(x+h) - f(x)}{h} + O(h)$$

and

$$f'(x) = \frac{f(x) - f(x-h)}{h} + O(h)$$

are first order and

$$f'(x) = \frac{f(x+h) - f(x-h)}{2h} + O(h^2)$$

is a second order finite difference approximation to the first derivative of f . For image processing usually the grid is fixed and thus the step-width h can be defined as 1. In this setting the approximation can be more easily denoted as a matrix-vector product. Therefore, the sampling of the function f on the interval $[0, n]_{\mathbb{R}}$ given by the data is modelled as a vector $\mathbf{f} = (f(ih))_{i=0}^{n-1}$, with a natural number n . Notably this means $f_{i+1} = f(ih)$. So the differences $D^+ f_i := f_{i+1} - f_i$ approximate the first derivative of f in ih for some $i = 0, \dots, n-2$. For differences across the boundary one must consider the continuous problem to be discretized. For variational methods in image processing Neumann boundary conditions are a natural way to model that the image is thought to be smooth across the boundary, so the differences vanish. Thus a natural choice would be e.g. $D^+ f_{n-1} := 0$.

Hence let

$$F := \begin{pmatrix} -1 & 1 & & & \\ & -1 & 1 & & \\ & & \ddots & \ddots & \\ & & & -1 & 1 \\ & & & & 0 \end{pmatrix} \in \mathbb{R}^{n \times n}$$

be a finite differences discretization matrix. Obviously

$$F^T F = \begin{pmatrix} 1 & -1 & & & \\ -1 & 2 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 2 & -1 \\ & & & -1 & 1 \end{pmatrix}.$$

This leads to a definition of discrete total variation of first and second order for one-dimensional data.

Definition 5. Let $\mathbf{f} \in \mathbb{R}^n$. The discrete one-dimensional total variation for step-width $h = 1$ of \mathbf{f} is

$$\text{TV}_1(\mathbf{f}) = \sum_{i=1}^n |(F\mathbf{f})_i| = \sum_{i=1}^{n-1} |f_{i+1} - f_i|.$$

The discrete one-dimensional total variation of second order for step-width $h = 1$ of \mathbf{f} is

$$\text{TV}_2(\mathbf{f}) = \sum_{i=1}^n |(F^\mathsf{T} F \mathbf{f})_i| = |f_1 - f_2| + \sum_{i=2}^{n-1} |f_{i-1} - 2f_i + f_{i+1}| + |f_{n-1} - f_n|.$$

A general treatment of total variation type penalizers for discrete signals, can be found in [41]. As an interpretation, one can think of the sum of discretized absolute second derivatives.

In the two dimensional setting, higher order partial derivatives must be discretized with care. Mostly a choice of all operators with the same order avoids the introduction of artifacts when solving a problem with finite differences. The basic equation for verifying a discretization of first and second order derivatives is the Taylor-expansion

$$f(\mathbf{x} + \mathbf{h}) = f(\mathbf{x}) + \mathbf{h}^\mathsf{T} \nabla f + \frac{1}{2} \mathbf{h}^\mathsf{T} \nabla^2 f(\mathbf{x}) \mathbf{h} + O(\|\mathbf{h}\|^3)$$

of a function $f \in C^4(\mathbb{R}^2)$ at a point dislocated by \mathbf{h} from the center of development \mathbf{x} . In a similar fashion one can derive second order consistent discretizations of the entries of the Hessian. The second derivative in horizontal direction is approximated by

$$\frac{\partial^2 f}{\partial x_1^2}(\mathbf{x}) = \frac{f(x_1 - h, x_2) - 2f(x_1, x_2) + f(x_1 + h, x_2)}{h^2} + O(h^2).$$

Similarly the second derivative in vertical direction is discretized as

$$\frac{\partial^2 f}{\partial x_2^2}(\mathbf{x}) = \frac{f(x_1, x_2 - h) - 2f(x_1, x_2) + f(x_1, x_2 + h)}{h^2} + O(h^2).$$

Finally the mixed derivative must be chosen with care. With the choice

$$\frac{\partial^2 f}{\partial x_1 \partial x_2}(\mathbf{x}) = \frac{\sum_{|h_1|=|h_2|=h>0} \text{sgn}(h_1 h_2) f(\mathbf{x} + \mathbf{h})}{4h^2} + O(h^2)$$

the whole eight-neighbors of a pixel \mathbf{x} are involved for a discrete symmetric Hessian and all differences are of the same order of consistency. In order to apply this formalism to a square digital image $\mathbf{f} \in \mathbb{R}^{\nu \times \nu}$ let it be thought of as a column vector $\mathbf{f} \in \mathbb{R}^n$ which is formed by the vertical concatenation of the columns of the image matrix, so $n = \nu^2$. Then the spectral norm of the Hessian serves for the definition of total variation of second order.

Definition 6. Let $\mathbf{f} \in \mathbb{R}^n$. Let

$$C := \frac{1}{2} \begin{pmatrix} -1 & 1 & & & \\ -1 & 0 & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 0 & 1 \\ & & & -1 & 1 \end{pmatrix} \in \mathbb{R}^{\nu \times \nu}.$$

The two-dimensional discrete total variation of second order for step-width $h = 1$ of \mathbf{f} is

$$\text{TV}_2(\mathbf{f}) = \sum_{i=1}^n \left| \frac{a_i - c_i}{2} \right| + \frac{1}{2} \sqrt{(a_i - c_i)^2 + 4b_i^2}$$

with discretized Hessian-matrix entries

$$a_i := ((I_n \otimes F^\mathbf{T} F) \mathbf{f})_i,$$

$$b_i := ((C^\mathbf{T} \otimes C) \mathbf{f})_i$$

and

$$c_i := ((F^\mathbf{T} F \otimes I_n) \mathbf{f})_i.$$

Remarkably, every summand of the total variation of second order sum can be interpreted as the spectral radius of the Hessian matrix at the corresponding pixel. This definition of a regularizer of higher order differs from the literature, see [36, 40]. Still the use of the spectral norm, which is identical to the spectral radius in the case of the symmetric Hessian, occurs to be the natural choice.

4.3 Higher Order Regularization

In [27] the link between ridge-feature and the following preprocess is motivated. A ridgeness-measurement is defined by the maximal absolute eigenvalue of the Hessian. This is by definition the spectral radius. So a preservation of this local image feature, a great spectral radius of the Hessian, is worth being preserved while reducing noise that interferes with the predictor part of feature detection. Figure 17 illustrates the effect which was initially studied in [41]. This figure shows the minimizers of a one-dimensional total variation regularization problem

$$\frac{1}{2} \|\mathbf{u} - \mathbf{f}\|^2 + \lambda \text{TV}_1(\mathbf{u}) \rightarrow \min$$

and of a one-dimensional total variation of second order regularization problem

$$\frac{1}{2} \|\mathbf{u} - \mathbf{f}\|^2 + \lambda \text{TV}_2(\mathbf{u}) \rightarrow \min$$

for data $\mathbf{f} \in \mathbb{R}^{256}$, which is either a step signal or a hat signal with additive Gaussian white noise and fixed parameter $\lambda > 0$. These optimization problems are special cases of GEN. Remarkably the location of the peak of the hat-signal, a zero-dimensional ridge, is well restored with total variation of second order penalizer. Note the similarity of Figure 17(b) with the features of the profiles in Figure 10(c).

It seems that this also applies for two-dimensional signals and thus for the restoration of ridges or valleys where the maximal principal curvature is significant. A heuristic argument for this is that the total variation penalized model restores image edges, where the discrete gradient has great length. The total variation of second order penalizes ridges and valleys with a significant spectral radius of the Hessian, so they are preserved at noise reduction. In [27] one proposal for a local ridgeness measurement, which is in fact also a valleyness measurement, is the maximal absolute eigenvalue of the Hessian. Hence total variation of second order minimization seems to be suitable for problems of denoising signals with prominent ridges and valleys. This proposes it as an appropriate technique for preprocessing microscopic images for the detection of well defined fibers of unknown spatial distribution. As shown in Section 4.2.2 the implementation can be done efficiently. So the preprocess proposed for denoising of two dimensional signals is

$$\frac{1}{2}\|\mathbf{u} - \mathbf{f}\|^2 + \lambda \text{TV}_2(\mathbf{u}) \rightarrow \min,$$

where \mathbf{f} is a column-wise reshaped square digital image with TV_2 formulated in Definition 6.

To illustrate the two-dimensional case, a numerical example is presented, see Figure 18 for the result. Let f be a digital gray-value image degraded by noise. Then a minimizer $u \in \mathbb{R}^{\nu \times \nu}$ of the discretized problem

$$\sum_{x_1, x_2=1}^{\nu} (u(\mathbf{x}) - f(\mathbf{x}))^2 + \lambda \sum_{x_1, x_2=1}^{\nu} (Ru)(\mathbf{x}) \rightarrow \min \quad (13)$$

for $\lambda > 0$ is to be found. Hereby, the operator $(Ru)(\mathbf{x})$ evaluates the spectral radius of the Hessian by finite differences with homogeneous Neumann boundary conditions and is given by

$$(Ru)(\mathbf{x}) = \frac{\frac{1}{2}|(D^{xx}u)(\mathbf{x}) + (D^{yy}u)(\mathbf{x})|}{+\frac{1}{2}\sqrt{((D^{xx}u)(\mathbf{x}) - (D^{yy}u)(\mathbf{x}))^2 + 4(D^{xy}u)(\mathbf{x})^2}},$$

where the linear operators D^{xx}, D^{xy}, D^{yy} provide approximations of the entries of the Hessian. The discrete differential operators can be chosen as second order consistent, see e.g. [21] for details. For inner points and grid-

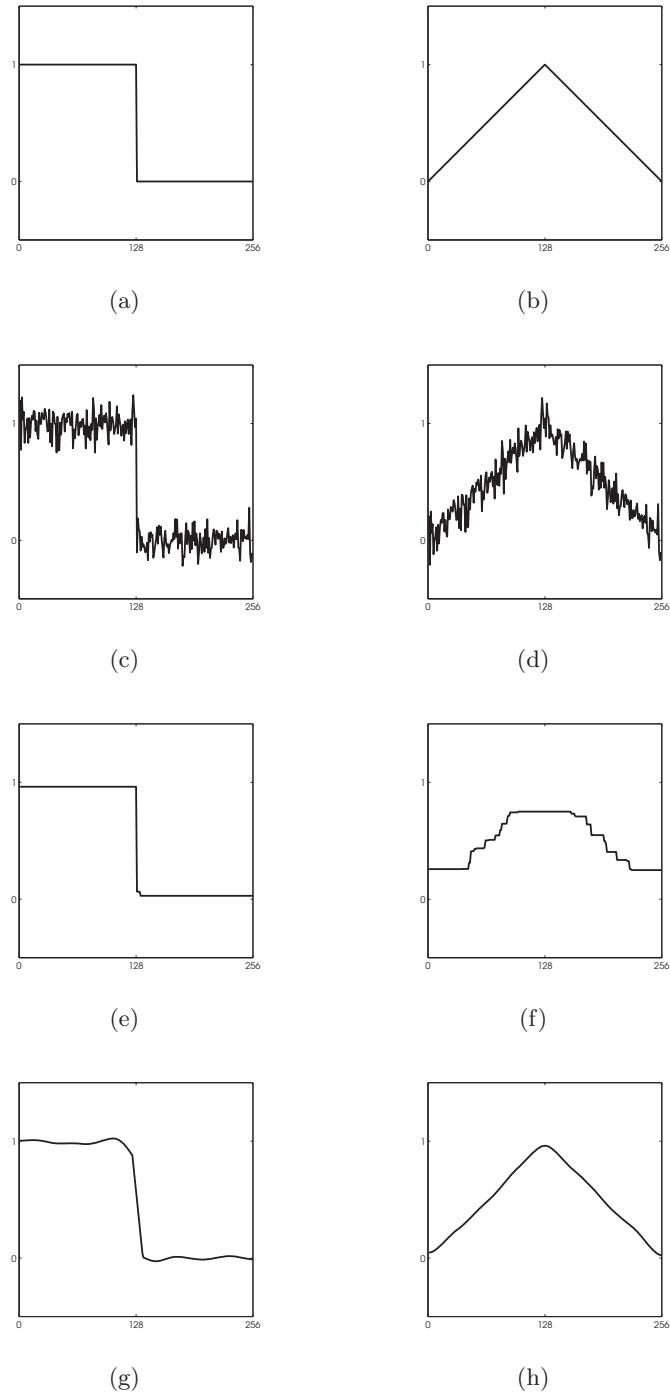


Figure 17: One-dimensional denoising of a piecewise constant and a piecewise linear signal. Parts (a) and (b) are the originals, parts (c) and (d) are respective degraded versions of the originals, parts (e) and (f) are restored by one-dimensional minimization of total variation and parts (g) and (h) are restored by minimization of total variation of second order.

width $h = 1$ this is realized with

$$\begin{aligned}
(Ru)(\mathbf{x}) = & \frac{1}{2} |4u(\mathbf{x}) - u(x_1 - 1, x_2) - u(x_1 + 1, x_2) \\
& - u(x_1, x_2 - 1) - u(x_1, x_2 + 1)| \\
& + \frac{1}{4} \left(4(u(x_1 - 1, x_2) + u(x_1 + 1, x_2) \right. \\
& - u(x_1, x_2 - 1) - u(x_1, x_2 + 1))^2 \\
& - (u(x_1 + 1, x_2 + 1) + u(x_1 - 1, x_2 - 1) \\
& \left. - u(x_1 + 1, x_2 - 1) - u(x_1 - 1, x_2 + 1))^2 \right)^{\frac{1}{2}}.
\end{aligned}$$

To calculate a solution of (13), Algorithm 4 is used.

Note that it is natural to use the spectral norm instead of the Frobenius norm of the Hessian, because the spectral norm is the matrix norm induced by the Euclidean norm for the first order case.

This method leads directly to another theoretical question related to current research about scale space. The notion of ridges is extended with regard to scale in [30]. An automatic scale selection allows for accurate results with lower requirements on image quality and prior knowledge. So far only Gaussian scale space is addressed. The scale space formed by total variation of second order might be well suited for a new practical notion of ridges.

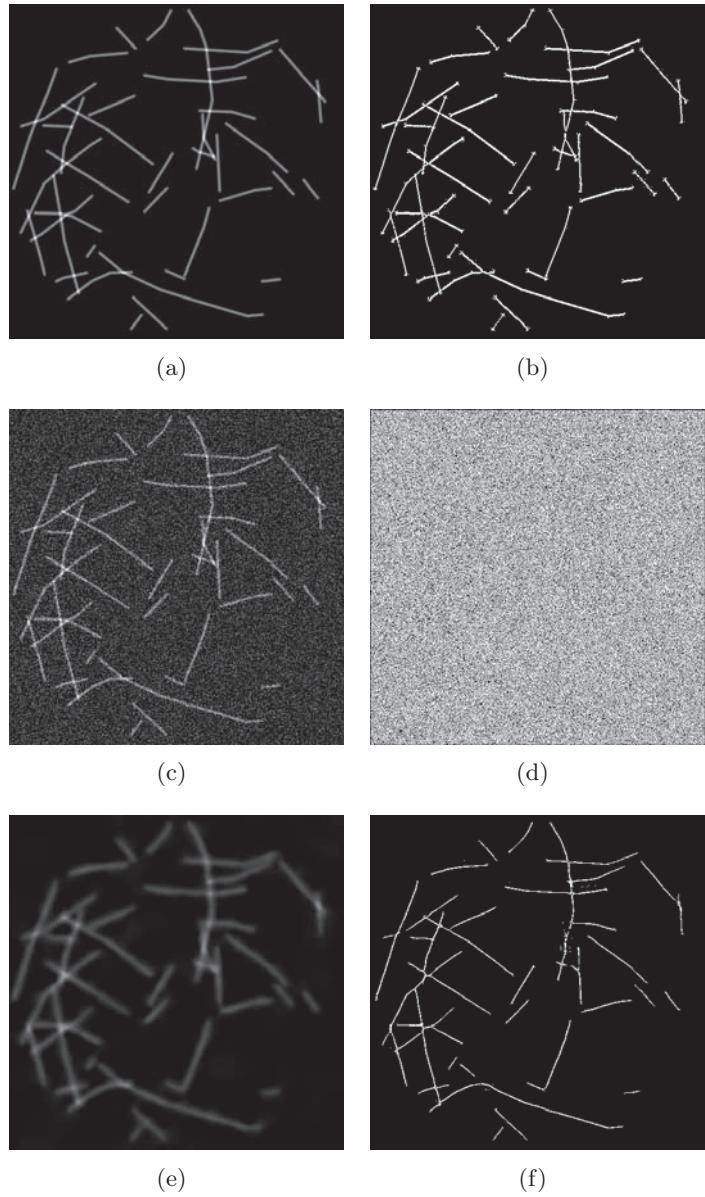


Figure 18: Illustration of ridge preserving image restoration with a total variation of second order penalizer. The artificial image (a) with intensity range $[0.0, 1.0]$ has a digital ridge-set shown in (b) by application of Definition 3 with a finite difference approximation of derivatives. The same procedure leads to a meaningless result (d) for a version (c) of (a), which is degraded with pseudo-random Gaussian white noise of standard deviation 0.25. Part (e) shows the minimizer of problem (13) with (c) as f and $\lambda = 1.0$. The corresponding ridge set (f) is then similar to the original, which might be unknown in applications.

5 Mathematical Morphology

Mathematical morphology is a fundamental area of visual computing devoted to the analysis of spatial structures. The basic phenomena modelled are shape and form. That explains the term morphology, which sets a link to the study of shape and form in biology for example. The greater strength of the theory is that it is founded on few mathematical axioms, which do not loose their diversity for long sequences of concatenation. For this, a basic morphological operator and the data it is applied to have exchangeable roles. A recent treatment of a broad range of applications can be found in [39].

Two leafs from the tree of morphological operations are of special interest for ridge detection. There is the medial axis transform, which provides a notion of a centerline of a given subset of the Euclidean plane. This is important when thickness of elongated structures in networks is negligible. The medial axis is the state of the art for quantification based on ridge detection, see e.g. [2]. This has the disadvantage of spurious ridges that do not reflect substantial information contained in the image. There are approaches to prune discrete medial axes or skeletons, but this involves additional parameter-tuning to this end. Skeletons are meant to present the state of the art and its short-comings here. A similar approach has been used for classification in [26].

Furthermore there is the top-hat transform which allows for the extraction of details from images in disposing brightness variation on a coarse scale relative to the structuring element of the operator involved. The top-hat transform is a method of gray-scale morphology. At the occurrence of uneven background illumination this nonlinear operation is suitable for a reduction of this effect, see [17]. This allows for a preprocessing step that stabilizes the subsequent feature detection.

5.1 Basic Operations

Mathematical morphology describes morphological operations for binary, scalar and even matrix-valued images. The goal is the extraction of relevant structures by exploring the data-set with structuring elements. These allow for the introduction of a-priori knowledge about the feature in their degrees of freedom. The notion of a structuring element is best introduced for binary images. These are the result of any kind of prior segmentation of image objects leading to bounded subsets of the image domain, e.g. the segmentation of a cumulative level-set by a brightness criterion in a digital image. Equivalently, a binary image can be described by either a digital image $f : \Omega \rightarrow \{0, 1\}$ or its one-set $X = \{\mathbf{x} \in \Omega : f(\mathbf{x}) = 1\}$, where $\Omega \subset \mathbb{Z}^2$ is the computational domain. In the set representation a structuring element can be described in the same manner, as a bounded subset $Y \subset \mathbb{Z}^2$ of the infinite grid. In this setting the basic operations of Erosion and Dilation can

be expressed in terms of set-theory, thus forming the whole bandwidth of morphological operations.

5.1.1 Erosion and Dilation

The application of a structuring element Y is based on its shift by \mathbf{x} into the image-domain $(\mathbf{x} + Y) \cap \Omega$. For the sake of simplicity the intersection with Ω is not included in the subsequent formulas. An Erosion \ominus then tests for each pixel \mathbf{x} of the image-domain, whether the structuring element Y shifted there fits into the set-representation X . In terms of set-theory this can be expressed by:

$$X \ominus Y := \{\mathbf{x} \in \Omega : \mathbf{x} + Y \subset X\}.$$

Dilation \oplus is then a test, whether the structuring element hits the set-representation of the binary image:

$$X \oplus Y := \{\mathbf{x} \in \Omega : (\mathbf{x} + Y) \cap X \neq \emptyset\}.$$

Erosion and Dilation are dual in the sense of the set-identity

$$(X \ominus Y)^c = X^c \oplus (-Y).$$

The effect of both operations can be studied by considering $Y = B_r(\mathbf{0})$ a ball of radius $r > 0$ around the origin. Then the erosion $X \ominus B_r(\mathbf{0})$ shifts the boundary of X orthogonally to its interior. Dilation is an orthogonal shift of the boundary to the outside of the set, see Figure 19.

5.1.2 Opening and Closing

Concatenation of erosion and dilation defines further basic operations. Erosion not only removes details from a binary image. Also coarse structures are reduced. Morphological opening \circ is successive erosion and dilation with the same structuring element:

$$X \circ Y := (X \ominus Y) \oplus Y.$$

This reverses the effect of the structuring element on the coarse structures, which are still represented in the eroded binary image. Details of the object boundary extending into the background which do not fit the structuring element are not represented in the eroded binary image and are thus removed by opening. To achieve the same effect for the background of the binary image, for closing \bullet the order of dilation and erosion is exchanged:

$$X \bullet Y := (X \oplus Y) \ominus Y.$$

The notion is that thin links and gaps of the set X are disconnected and linked respectively by applying a structuring element that does not fit in

between. Furthermore small blobs are filtered out. The duality of erosion and dilation is conveyed into opening and closing as:

$$(X \circ Y)^c = X^c \bullet (-Y).$$

5.1.3 Morphological Skeletonization

Based on mathematical morphology one can define discrete realizations of the medial axis. There are several medial axis definitions, which lead to similar features in the continuous setting. In contrast, there are several different notions of skeletons in the discrete setting implementing the main properties of the medial axis. A popular illustration of the concept has been introduced in [10]. A subset of the plane is thought of as a lush dry prairie which is ignited all around its boundary. Gradually the fire fronts are propagated into the inner of the grassland with locally equal speed. The medial axis is then the set of all spots where the firefronts from different parts of the boundary meet and extinguish. This is formalized by the maximal disk definition

$$\begin{aligned} M(X) := & \{x \in X : \exists y_1, y_2 \in \partial X, y_1 \neq y_2 : \\ & \|x - y_1\| = \|x - y_2\| = \min_{y \in \partial X} \|x - y\| \} \end{aligned}$$

for a bounded set $X \subset \mathbb{R}^2$. So there exist at least two points of the boundary with the same Euclidean distance to the medial axis point, which is in passing the minimal distance to the boundary. This can be thought of in terms of a maximal disc around the medial axis point, which is totally contained in X and has at least two distinct contact points with the boundary.

The book [37] provides an implementation for the discrete setting, $X \subset \mathbb{Z}^2$, in terms of basic morphological operations. For this, an iteration of erosion is defined as

$$X \ominus kY := \underbrace{[[\dots [X \ominus Y] \ominus Y] \dots \ominus Y]}_{k \text{ times}} \ominus Y.$$

The skeleton pixels as discrete implementation of the medial axis are defined as the loss of details extending into the background at the successively eroded set:

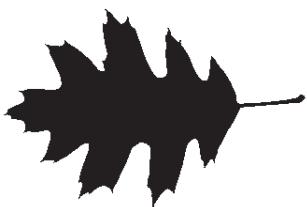
$$S_k(X) := (X \ominus kY) \setminus ((X \ominus kY) \circ Y).$$

The skeleton itself is the union of all losses of detail before cancellation by erosion

$$S(X) := \bigcup_{k=0}^m S_k(X)$$

with m such that

$$X \ominus mY \neq \emptyset, X \ominus (m+1)Y = \emptyset.$$



(a)



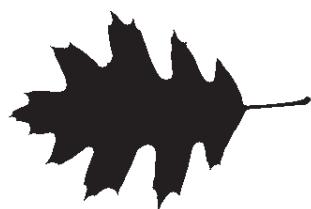
(b)



(c)



(d)



(e)

Figure 19: Effect of disk structuring element $B_5(\mathbf{0})$ on a binary image (a) with the silhouette of a leaf, erosion (b), dilation (c), opening (d) and closing (e).

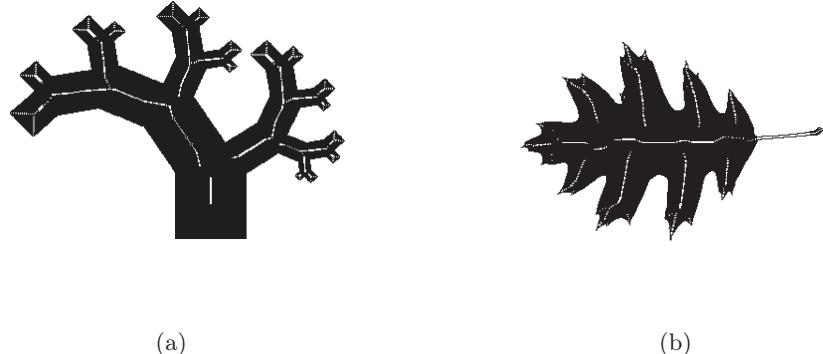


Figure 20: Morphological skeletons by accumulated loss of details at successive erosion, $X \subset \mathbb{Z}^2$ in black, $S(X) \subset X$ in white. The structuring element used is $Y = B_{\sqrt{2}}(\mathbf{0})$. Note the incomplete connectivity of the branches.

The resulting pixel set for a disk structuring element is close to the medial axis of a continuous set with its interpolated boundary, see [37]. Still the property of the medial axis that it forms a network of continuous curves, is lost, see Figure 20.

5.2 Specialized Operations

Binary images are very limited in representing features in degraded images. The segmentation of a foreground set is mostly an unnecessary loss of information in an image that is already an abstraction from real world scenes. Still given a good segmentation the morphological medial axis can be strongly improved. Furthermore mathematical morphology conveys into scalar valued images with few complications and allows for nonlinear filtering techniques to facilitate feature preserving image improvement. The presentation follows the lines of [39].

5.2.1 Nonlinear Filtering

In transfer of morphological operators to scalar valued images

$$f : \Omega \rightarrow \mathbb{R}$$

also the operators can be generalized

$$g : Y \rightarrow \mathbb{R}.$$

Then the basic operations can be defined as convolutions in the tropical algebra

$$[f \ominus g](x) := \min_{y \in Y} \{f(x - y) + g(y)\}$$

for the erosion and

$$[f \oplus g](\mathbf{x}) := \max_{\mathbf{y} \in Y} \{f(\mathbf{x} + \mathbf{y}) - g(\mathbf{y})\}$$

for the dilation. This is consistent with the binary case in the notion of a binary image as

$$f(\mathbf{x}) := \mathbf{1}_X(\mathbf{x}) := \begin{cases} 1 & \text{if } \mathbf{x} \in X, \\ 0 & \text{otherwise} \end{cases}$$

and a binary structuring element with

$$g \equiv 0.$$

Then erosion is comprised as

$$\mathbf{f} \ominus \mathbf{g} = \mathbf{1}_{X \ominus Y}$$

and dilation as

$$\mathbf{f} \oplus \mathbf{g} = \mathbf{1}_{X \oplus Y}.$$

This is similar to the median-filter in that it selects its value from an ordered list of brightness values in a neighborhood, but the first and the last respectively instead of the central element. Opening and closing are defined in the same manner as in the binary case. The opening

$$\mathbf{f} \circ \mathbf{g} := (\mathbf{f} \ominus \mathbf{g}) \oplus \mathbf{g}$$

denotes the loss of bright details. The closing-operator

$$\mathbf{f} \bullet \mathbf{g} := (\mathbf{f} \oplus \mathbf{g}) \ominus \mathbf{g}$$

denotes loss of dark details. With the top-hat transform

$$\mathrm{T}_{\mathbf{g}}(\mathbf{f}) := \mathbf{f} - (\mathbf{f} \circ \mathbf{g})$$

this allows for extraction of bright details and thus negligence of varying background illumination. For an example of a confocal laser scanning microscopic image, see Figure 21.

5.2.2 Homotopic Skeletons

The book [39] points out that there is a single medial axis but a multitude of morphological skeletons: “Several formal definitions for the skeleton of a Euclidean set are available. Fortunately, they all lead to similar thin lines. However, their extension to discrete sets has lead to a wide variety of skeletons. Moreover, the properties of the resulting discrete skeletons are not all identical to those of their continuous counterpart.” The following treatment

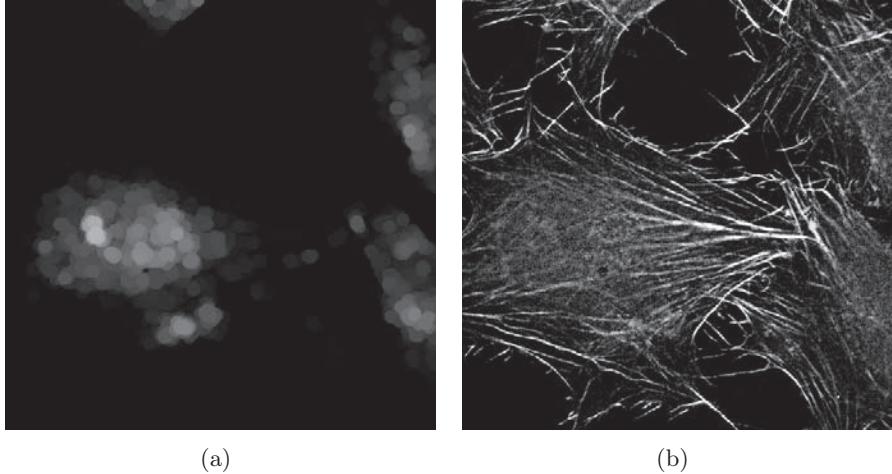


Figure 21: Morphological top-hat transform applied to a confocal laser scanning microscopic image of the actin filament of a human MG-63 osteoblast in Figure 6. Parts (a) and (b) show the opening and the top-hat transform of this microscopic image with a disc structuring element of radius 8 pixels centered at the origin.

is lead by [13] which is a strict evaluation of different skeletons of discrete sets. Two of these illustrate the argument.

The medial axis or skeleton of a Euclidean set, can be e.g. alternatively defined for $X \subset \mathbb{R}^n$ as

$$M(X) := \{\mathbf{x} \in X : \#(\sup_r \{B_r(\mathbf{x}) \subset X\} \cap \partial X) > 1\}.$$

Despite of the disparity of notation this leads to the same set as the definition in Section 5.1.3. Recently the medial axis is used for computer graphics applications where the set $X \subset \mathbb{R}^2$ is defined as a polygonal bounded body. For the application to binary images one must concentrate on properties of the medial axis to transfer into a discrete setting. A method both computationally efficient and capable of preserving the homotopy bases on parallel thinning. This consists of criteria applied to the boundary pixels of the discrete set for whether they can be removed without loss of connectivity or branches. Formally

$$R(X) := \{\mathbf{x} \in X : \mathbf{x} \text{ irrelevant boundary pixel}\}, X \subset \mathbb{Z}^2,$$

is defined without regard to the ordering of the pixel set. The operator

$$A(X) := X \setminus R(X)$$

implements a thinning of the current set and the skeleton is then the m -th

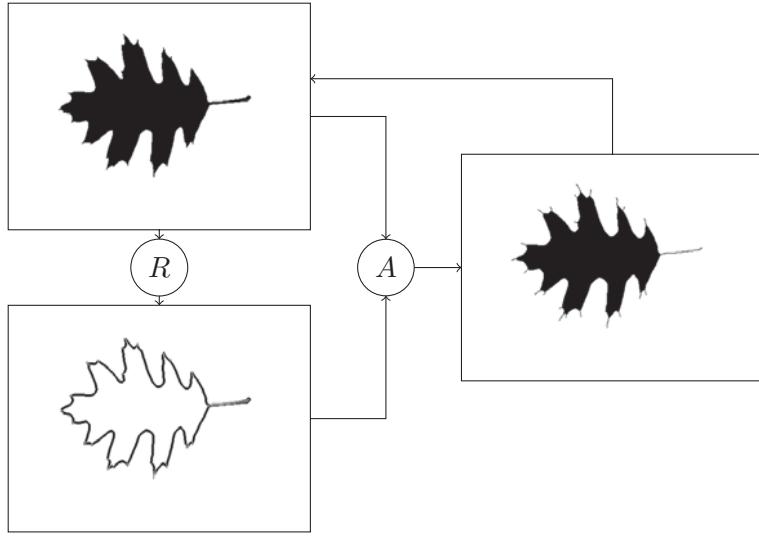


Figure 22: Scheme of parallel thinning algorithms. For visualization of the effect on boundary pixels, the lower and the right image are the accumulation of four subsequent applications of the operators.

iteration

$$S(X) := A^m(X), \text{ such that } A^{m-1}(X) \neq A^m(X) = A^{m+1}(X).$$

The result strongly depends on the notion of the operator R . Figure 22 illustrates the scheme. Typically the notion of relevance involves the preservation of the homotopy, i.e. in 2 dimensions the number of holes in the skeleton is the same as in the original set. Also the preservation of significant branches is part of the notion of relevance of a pixel. This is where there are obvious differences in the methods proposed in the literature.

In [13] there are several approaches compared with regard to similarity to the medial axis and homotopy preservation, ordered by publishing authors and year of publication: Pavlidis '81, Chin et al. '87, Holt et al. '87, Hall '89, Guo & Hall '92, Jang & Chin '93, Eckhardt & Manderlechner '93 and Bernard & Manzanera '99. Figure 23 shows two examples with different branching behavior. The method proposed by Guo and Hall 1992 (GH '92) is reported to perform well and thus chosen for the coming Section as a substantiation of the state of the art, which was published without a requirement on the exact algorithm to be used.

5.2.3 Ridge Detection by Skeletonization

This Section defines an alternative to the method described in Section 3.2.3. The data is assumed to be preprocessed, such that all brightness information

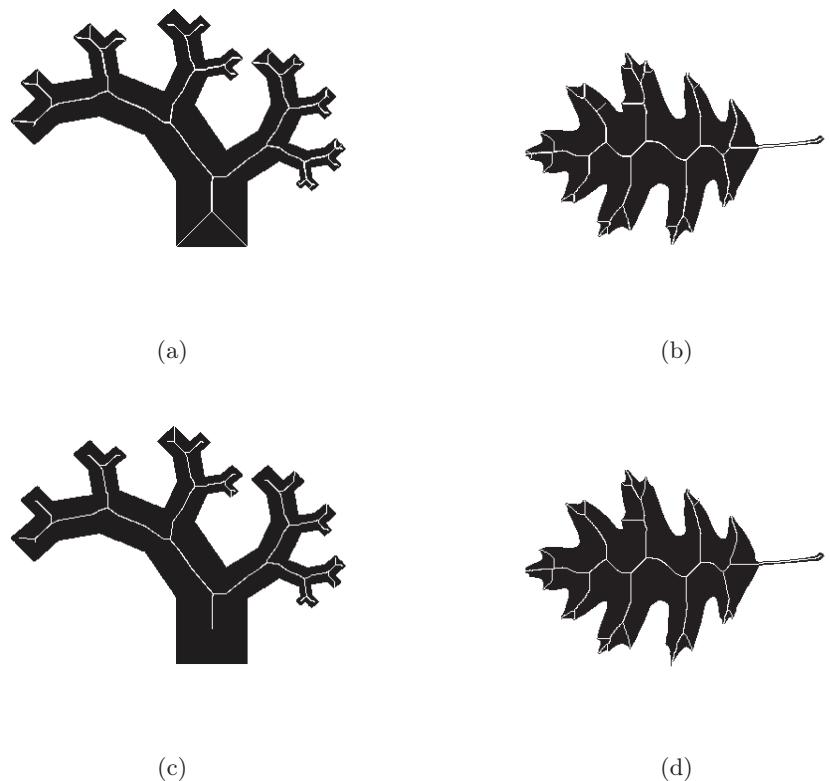


Figure 23: Morphological skeletons by parallel thinning approaches, the original set $X \subset \mathbb{Z}^2$ in black and the skeleton $S(X) \subset X$ in white. Parts (a), (b) result from the application of Pavlidis '81, parts (c), (d) from Guo & Hall '92. The results differ in that lines are not always one pixel thin and there are more branches with Pavlidis' approach.

can be assumed to be determined by the feature. A segmentation X , e.g. by brightness thresholding or other methods, must be processed with the methods described in Section 5.2.2 such that a binary mask feature $S(X)$ is given. Furthermore this binary mask must be one pixel thin except for line crossings, i.e. only crossing pixels have more than 2 eight-neighbors. With this input the following algorithms proposed in [2] are reproduced here for a presentation of the state of the art. The output is a geometrically embedded graph, to be quantitatively described with the methods introduced in Section 2.

Algorithm 5 defines a subgraph to the *grid-graph* of the given skeleton, i.e. the graph with edges between 8-neighbors of the skeleton. A subgraph is found by a combinatorial optimization with a geometrical and a technical objective.

Let vertices of degree greater than 2 in the grid graph be denoted as cluster vertices. Most cluster vertices should loose all superfluous edges while paths between non-cluster vertices which visit only cluster vertices in between are possibly longer but not broken. For this, *clusters* are defined as connected components of the subgraph of the grid-graph induced by vertices of degree other than 2. These clusters are reduced to spanning trees of the graph induced by the cluster vertices and their neighbors. Among all spanning trees, the one with the greatest objective function of the **PAST** is selected, see Section 2.2.1. This is a choice, which is not proposed in the literature, but which obviously optimizes the result for the quantification conducted in the application of this thesis. In case of non-uniqueness, the lexicographically first graph is selected as follows. Let $\mathbf{v}, \mathbf{w} \in \mathbb{Z}^2$ be distinct vertices. Then $\mathbf{v} < \mathbf{w}$, if and only if $v_1 < w_1$ or $v_1 = w_1$ and $v_2 < w_2$. Let $e = \mathbf{v}\mathbf{w}$, $e' = \mathbf{v}'\mathbf{w}'$ be distinct edges with $\mathbf{v} < \mathbf{w}$ and $\mathbf{v}' < \mathbf{w}'$. Then $e < e'$ if $\mathbf{v} < \mathbf{v}'$ or $\mathbf{v} = \mathbf{v}'$ and $\mathbf{w} < \mathbf{w}'$. As the vertex set does not vary, the ordering of edges is sufficient for sorting graphs. Let $G = (V, E)$ and $G' = (V, E')$ be two distinct graphs. Then $G < G'$ if $\min\{E \setminus E'\} < \min\{E' \setminus E\}$.

Algorithm 5 Graph from Segmentation

```

 $V := S(X).$ 
 $E := \{\mathbf{v}\mathbf{w} \in V \times V : \max\{|v_1 - w_1|, |v_2 - w_2|\} \leq 1\}.$ 
 $G := (V, E).$ 
 $V_C := \{\mathbf{v} \in V : \deg(\mathbf{v}) \neq 2\}.$ 
for all connected components  $G' = (V', E')$  of  $G[V_C]$  do
     $V'' := V' \cup \bigcup_{\mathbf{v} \in V'} N_G(\mathbf{v}).$ 
     $G'' := G[V''].$ 
    Let  $G''' := (V''', E''')$  be the lexicographically first spanning tree of  $G''$  with maximal objective function value of the PAST.
     $G := (V, (E \setminus E'') \cup E''').$ 
end for

```

Algorithm 6 simplifies the graph by joining vertices of degree other than 2 with weight-bounded shortest paths and deleting short cycles with exactly one crossing point by a length criterion with a tuning parameter ρ_1 . In [2] the value $\rho_1 = 8$ pixel is proposed. This can be taken as a value, which less depends on the scale of objects displayed by the image and rather from the result of skeletonization algorithms dependent on the pixel grid.

Pairs of vertices of degree other than 2 are sorted by shortest path length in the Euclidean length weighted graph. The lexicographically first, closest pair is joined and the procedure is repeated until no pair has distance less than ρ_1 . After all joining operations short cycles are deleted. For this, all cycles with only one vertex with degree greater than 2 are deleted if their length with Euclidean weighting is less than ρ_1 . The resultant graph is then a meaningful representation of filamentous objects. However, connectivity which is missing in the segmentation, is also missing in the result. The complexity strongly depends on the size of clusters for which spanning trees are enumerated. There might be reasonable heuristics for restriction of this complexity. However, a worst case analysis would require more information about the skeleton in terms of forbidden 8-neighborhoods. This is out of the scope of this presentation.

Algorithm 6 Graph Simplification

```

 $G := (V, E).$ 
repeat
   $V_C := \{v \in V : \deg(v) \neq 2\}.$ 
   $\gamma : E \rightarrow \mathbb{R}_+, \gamma(vw) = \|v - w\|.$ 
  Let  $v, w \in V_C$  be the lexicographically first pair of vertices with a shortest path not longer than the one between every other pair. Let  $\gamma_0$  be the weight of this path.
  if  $\gamma_0 \leq \rho_1$  then
     $G := (V \setminus \{v, w\} \cup \{(v + w)/2\}, (E \setminus (\{v\} \times N(v) \cup \{w\} \times N(w)) \cup \{(v + w)/2\} \times (N(v) \cup N(w)))).$ 
  end if
  until  $\gamma_0 > \rho_1$ 
   $V_C := \{v \in V : \deg(v) \neq 2\}.$ 
  for  $(v_0, v_1, \dots, v_m, v_0)$  is a cycle in  $G$  with  $v_0 \in V_C$ ,  $m \geq 1$  and  $v_1, \dots, v_m \in V \setminus V_C$  do
    if  $\gamma(\{v_0v_1, \dots, v_mv_0\}) \leq \rho_1$  then
       $G := G[V \setminus \{v_1, \dots, v_m\}].$ 
    end if
  end for

```

6 Experimental Validation

At this point the problem has been seen from different perspectives in the field of mathematical image processing. Numerical experiments are necessary for novel developments in all of these fields. They show both the relevance of the theories and reflect the bias of their viewpoint. Usually, there is no improvement of the state of the art without any disadvantages concerning computational complexity, stability or systematic errors.

The tools for the numerical experiments were a Win32 application FilaQuant, which is an implementation of parameter estimators described in Section 2 based on ridge detection proposed in Chapter 3 with a preprocessing described in Sections 4 and 5. The experimental implementation is compiled with Embarcadero® C++ Builder 2010®. It is applied to images stored in uncompressed TIFF files. The preparation of the statistics was done with Matlab R2008b. For the execution, a Microsoft Windows XP Professional, Version 2002, Service Pack 3 system was used. For statistical evaluation the Analyse-it plug-in for Microsoft Office Excel 2007 was used as an implementation of the U-test and the T-test.

6.1 Qualitative Consistency

The parameters for the description of the actin filament of cells grown on well understood surfaces serve as a benchmark for the quantification. In [31, 15] there are qualitative results concerning cellular response to stochastic biomaterial surfaces. The sample images of these studies are used with FilaQuant and the quantitative results are qualitatively assessed by statistical tests with and without distribution assumptions.

A necessary condition for sufficient quantitative accuracy is consistency with established qualitative results. Therefore a set of sample data from qualitative research about the influence of chemically treated biomaterial surfaces on the actin filament of MG-63 osteoblasts is used. There are 3 to 7 images of each population. The established observations concern maximal length and number of filaments. The filament was automatically tagged in the images and decomposed by the software FilaQuant.

Two hypotheses are tested for small sets of routinely generated images. The images have been selected for representation of the actin filament phenotype. Note that results of quantitative imaging will require less selected data. FilaQuant is capable of processing stacks of images, each in a time proportional to the pixel number. The images must be acquired with a filament width of at least one pixel. This selection of images and the selection of evaluation parameters about contrast, filament width, filament curvature are manual input so far. The one-sided student T-test, see Table 2, and the Mann-Whitney U-Test, see Table 3, are applied. Table 1 shows the characterization of the data. For the use of these statistical methods see e.g.

Table 1: Characterization of the samples with populations sizes n_1, n_2 and the corresponding variances of the estimated parameters σ_1, σ_2 . The naming convention for the data stems from the application for which the images were acquired and qualitatively assessed for a confirmation of the hypotheses.

Data	n_1	n_2	σ_1^2	σ_2^2
number Col30 vs. CD30	5	4	7843.7	1174.92
length Ti180 vs. Ti30	7	3	5556.95	241.48

Table 2: One sided T-test of hypotheses about length and number of filaments. They can all be accepted with a probability of first kind error less than 0.01. The values μ_1, μ_2 are the sample means. The length is given in μm .

T-test $H_0 : \mu_1 > \mu_2$	μ_1	μ_2	p-value
number Col30 vs. CD30	27.63	4.18	0.0089
length Ti180 vs. Ti30	27.71	8.25	0.002

[45].

It is clear that the small sample size and the great variance are not sufficient for valid statistical results. Also there are arguments against the use of the T-test for hypotheses about integer random variables. The U-test is closer to practical application for this kind of data. However, the sample size is proposed to be equal and greater than the examples here. Still, these examples show, that FilaQuant is accurate enough to be applied to greater data-sets without violation of well-known qualitative behavior of actin filaments imaged through confocal laser scanning microscopy.

Table 3: One sided Mann-Withney U-test of hypotheses about length and number of filaments. They can all be accepted with a probability of first kind error less than 0.05. The values m_1, m_2 are the sample medians. The length is given in μm .

U-test $H_0 : m_1 > m_2$	m_1	m_2	p-value
number Col30 vs. CD30	209	13.5	0.0159
length Ti180 vs. Ti30	167.5	55.2	0.0083

6.2 Random Model

For the verification of an accurate result, images with known ground truth are required. They are provided by a random model, which randomly defines a geometrically embedded graph and defines a corresponding image. Hence, the result of the ridge tracking algorithm, see Section 3, can be checked against the true graph. The model could be defined arbitrarily realistic. The following models are restricted to ridges with fixed width and contrast, thus the optimal α is known in advance. Application images are degraded by noise, uneven background illumination and varying contrast. The random models, see Figure 24, verify the performance of the ridge tracking algorithm both after perfect elimination of noise and after a preprocess.

There are different random models for tessellation features in an experimental state, see [3]. Ideas for graph features are adopted with the extension that the image domain is not fully decomposed by the feature lines. The first model, EDGE, is the random choice of $m \in \mathbb{N}$ pairs of pixels from the greatest digital disc fitting the image domain and their connection by an edge. The second and the third model, TRAIL and ELLIPSE, are based on the Delaunay-triangulation, see [1], of a $3m$ -pixel random point-process in a digital ellipse. Recall that the Delaunay-triangulation of these pixels is defined as their connection by edges to a maximal planar triangle graph with the triangles chosen such that their circum circles enclose no vertex from the pixel-set except for the triangle corners. The center of gravity of the ellipse is at the center of the image. One semi-axis is one half of the row number of the image. The other semi-axis is of equal length for TRAIL and of half the length for ELLIPSE respectively. Subsequently, $3m$ pixels are drawn from the ellipse with equal probability. The triangle graph by the Delaunay-triangulation is decomposed into almost straight trails, see Section 2.2, of which one is chosen with equal probability. This procedure is m times repeated so that m curved trails are defined. Given a trail E_j , the corresponding image is defined by

$$f(\mathbf{x}) = \lfloor 128 \max_{\mathbf{v}, \mathbf{w} \in E_j} \max\{g_\alpha(\rho(\mathbf{x}; \mathbf{v}, \mathbf{w})), g_\alpha(\|\mathbf{x} - \mathbf{v}\|), g_\alpha(\|\mathbf{x} - \mathbf{w}\|)\} \rfloor.$$

with

$$g_\alpha(s) := \exp\left(-\frac{s^2}{2\alpha^2}\right).$$

This means, the brightness is modelled with Gaussian function decay with distance from edges and vertices of the graph from half the maximal gray value. The images of all m edges in the EDGE model and the trails in the TRAIL and ELLIPSE models are added and cut above 255, so crossings are brighter than isolated trails. A fourth model, noisy ELLIPSE, has a modification of trail superposition. Let f be the image of the ELLIPSE

model. Then the noise model \hat{f} is given by

$$\hat{f}(\mathbf{x}) = \frac{1}{100} X \left(\left\lfloor 100 \left(\frac{1}{3} + \frac{2}{3} f(\mathbf{x}) \right) \right\rfloor, \mathbf{x} \right),$$

where $X(t, \mathbf{x})$ is a Poisson distributed random number with parameter t for every $\mathbf{x} \in \Omega$. Recall that this means $X(t, \mathbf{x})$ is an nonnegative integer, with the probability of the realization $k \in \mathbb{N}$ given by

$$P(X(t, \mathbf{x}) = k) = \frac{t^k \exp(-t)}{k!}.$$

The image \hat{f} is cut above 255. This models an illuminated background and shot-noise as image degradations. For the subsequent calculations α is chosen 1.5 pixels and the **PAST**-threshold β_0 is $\arccos(-0.9)$.

In order to assess the performance of the proposed algorithm the random model is used for elementary error statistics. For this, 100 sample images are drawn from the random models for each experiment. All images are automatically processed with certain tuning parameters substantiating Algorithms 2, 5 and 6. The error is measured relative to the true value for the total length, greatest length and angular deviation. The number of filaments which is expected to be close to m is given as the total value. The preferred orientation is given as an absolute angular error. The distribution of errors over the population of 100 samples are given by the quartiles. For EDGE and TRAIL images with 10, 20, 30 and 40 trails are assessed for density influence. Another experiment highlights the influence of tuning parameters. Furthermore there is a comparison of the proposed method with a state of the art technique with TRAIL, ELLIPSE, each with and without noise.

6.2.1 EDGE

Table 4 shows the error quartiles for the EDGE model. Each parameter error has three rows with the average of 25-th and 26-th, the 50-th and 51-st and the 75-th and 76-th value in the ascending order. The row "length error 75%" e.g. reads, the relative error of the total length parameter for the average of 75-th and 76-th sample in the list ordered by the error in ascending order, is 2.6% for $m = 10$, 4.3% for $m = 20$, 8.4% for $m = 30$ and 13.0% for $m = 40$.

Obviously the trail number was overestimated systematically. This is hard to avoid without post-processing. One reason is that superposition interferes with the exact reconstruction of ground truth. It might even be impossible, e.g. for more than two edges with exactly the same end-points. Crossings with very acute angles over a long distance can reduce the response of the ridgeness along a line such that one or both edges are not properly detected. The effects of superposition increase with the number of trails. The parameter choice was $(\ell_0, r_0, \rho_1) = (16, 60, 512)$. The greatest edge length ρ_1 was

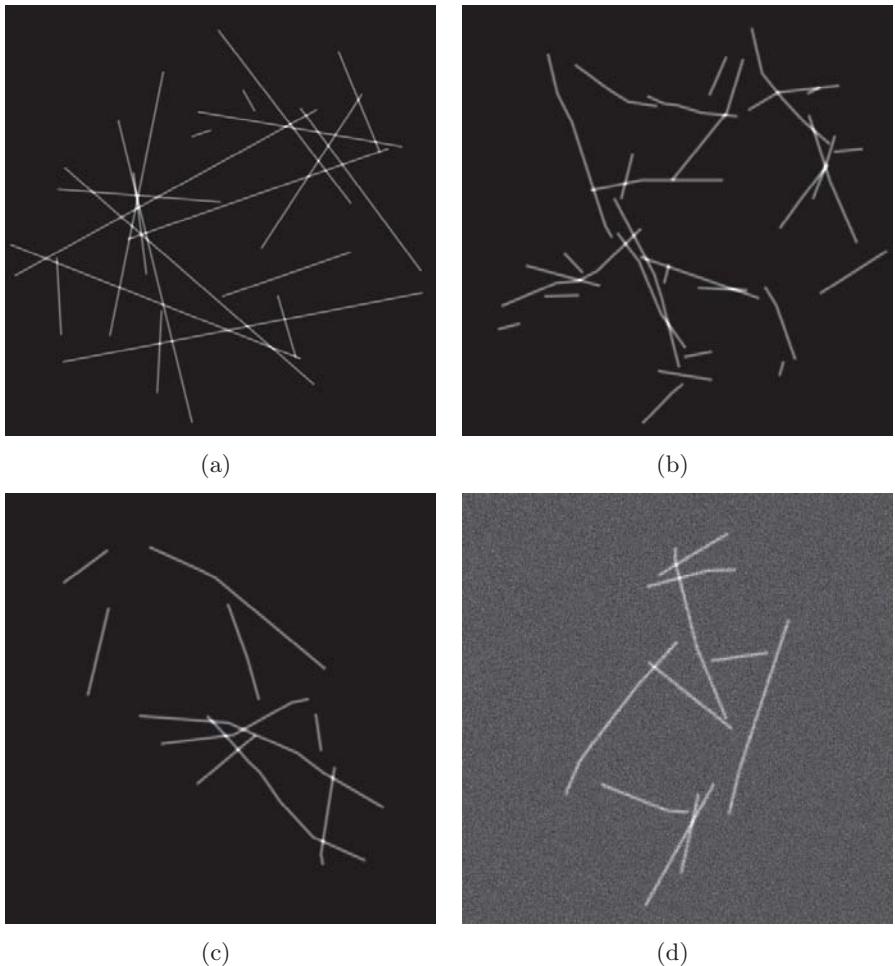


Figure 24: One realization of the random models for microscopic images showing superposed fibers which mimic biological variation, part (a) by the EDGE model, part (b) by the TRAIL model, part (c) by the ELLIPSE model and part (d) by the NOISE model.

Table 4: EDGE model with increasing density

trail-number truth	10	20	30	40
trail-number 25%	10	20	32	46
trail-number 50%	10	21	33	49
trail-number 75%	10	22	38	54.5
length error 25%	0.8%	2.5%	4.8%	8.3%
length error 50%	1.6%	3.3%	6.1%	10.3%
length error 75%	2.6%	4.3%	8.4%	13.0%
error longest 25%	0.6%	0.5%	0.5%	0.5%
error longest 50%	0.7%	0.7%	0.6%	1.3%
error longest 75%	0.8%	3.6%	3.5%	5.0%

as great as possible. This implies a great computational burden because the cost-dominating straight ridge criterion is not fulfilled for most end-points selected from a large neighborhood. The ridge criterion for lines was more restrictive than the pixelwise blob criterion because more information is given by the context.

6.2.2 TRAIL

Table 5 shows the results for the TRAIL model, where the curves are smooth but not entirely straight, see Figure 24(b). The effects leading to false detection are similar to EDGE. The errors of the longest trail parameter were greater for small and accordingly overestimated trail numbers. For greater true trail number, the effects of close parallel trails are reduced. Still the longest trail length showed greater errors. The parameters were chosen as $(\ell_0, r_0, \rho_1) = (12, 80, 32)$. The maximal edge length ρ_1 was smaller because there is no benefit in detecting very long edges. Long edges are unlikely in the Delaunay-triangulation of a point-process with the given density.

The algorithm depends on a suitable choice of the tuning parameters. Table 6 shows the performance for the TRAIL model with $m = 40$ and for alternative choices for the ridgeness and the nodeness thresholds. The insight is twofold. Table 5 has suitable tuning parameters, because the errors increase in almost every instance when changing them. Stability is greater for the nodeness threshold than for the ridgeness threshold.

As a comparison Table 7 shows the results of the state of the art method based on skeletonization. The errors are generally greater compared with the ridge tracking results. Especially the longest filament estimation error strongly increases in m .

Table 5: TRAIL model with increasing density and Algorithm 2

trail-number truth	10	20	30	40
trail-number 25%	10	21	33	44
trail-number 50%	10	22	36	48
trail-number 75%	12	25	39	51
length error 25%	0.3%	0.7%	0.7%	1.0%
length error 50%	0.6%	1.1%	1.3%	1.5%
length error 75%	1.1%	1.5%	2.2%	2.2%
error longest 25%	0.2%	0.3%	0.4%	0.3%
error longest 50%	0.5%	0.6%	0.7%	0.7%
error longest 75%	0.7%	2.5%	7.6%	8.8%

Table 6: TRAIL model with different tuning parameters

ℓ_0	8	16	12	12	12
r_0	80	80	60	100	80
trail-number 25%	45	44	60	56	44
trail-number 50%	48	48	65.5	61	48
trail-number 75%	52	52	74.5	66.5	51
length error 25%	1.1%	0.9%	7.8%	3.1%	1.0%
length error 50%	1.6%	1.5%	9.2%	3.7%	1.5%
length error 75%	2.4%	2.2%	10.5%	4.5%	2.2%
error longest 25%	0.4%	0.3%	1.1%	2.9%	0.3%
error longest 50%	1.0%	0.9%	8.6%	15.6%	0.7%
error longest 75%	9.3%	14.5%	22.2%	27.6%	8.8%

Table 7: TRAIL model with increasing density and Algorithms 5 and 6

trail-number truth	10	20	30	40
trail-number 25%	11	24	38	50.5
trail-number 50%	12	27	41	55
trail-number 75%	14	30	44	59.5
length error 25%	3.3%	1.7%	1.3%	1.1%
length error 50%	4.2%	3.1%	2.1%	2.0%
length error 75%	5.0%	4.4%	3.0%	3.1%
error longest 25%	4.3%	6.3%	6.6%	6.7%
error longest 50%	7.1%	14.0%	13.8%	18.2%
error longest 75%	21.0%	24.9%	29.7%	31.1%

6.2.3 ELLIPSE without and with noise

Tables 8 and 9 show the error quartiles of the former parameters together with two additional parameters about length and orientation of edges, see Section 2.3. All eight experiments only use the most complex random models, namely TRAIL, noisy TRAIL, ELLIPSE and noisy ELLIPSE. Noisy TRAIL is realized as noisy ELLIPSE with equal semi-axis, which produces the same spatial distribution of trails as in the TRAIL model. For all four random models 100 samples with $m = 10$ trails are drawn. The noisy realizations are treated with top-hat transform with disk structuring element radius 4.5, see Section 5.2.1, and total variation of second order weight $\lambda = 0.1$ for the image with intensity pixel-wise normalized to range $[0, 1]$ by division by 255, see Section 4. Furthermore the tuning parameters of the denoised images are chosen differently as $\alpha = 1.2$ and $r_0 = 25$ for the ridges are effectively thinner at the piecewise linear reconstruction of a gaussian shaped crossection. These images were then evaluated by the novel ridge tracking method, see Section 3, for the results in Table 8 and by the state of the art skeletonization method, see Section 5.2.3, for the results in Table 9. This offers different views on the accuracy of the quantification compared to the ground truth of random images. Note that the results for skeletonization are slightly different from the first column of Table 7 as the sample data is different.

Concerning the different approaches to ridge detection the tables show larger errors for the skeletonization approach for all parameters in the scope of this thesis and all random models. The overestimation of the filament number is common to both approaches. This high level feature is strongly context sensitive and slightly better quantified with the novel approach. Overall, the benefit of the novel approach is strongly sensitive for the random model and hence for the noise level and the orientation distribution.

The total length has a stable error for all models with the skeletonization approach. For the novel approach there is a cumulative bias due to noise and elongation. This shows that the error has a disadvantageous tendency to quantify strongly dissimilar observations with a different accuracy, so that only the worst accuracy for noisy ELLIPSE can be assumed for application data. In this respect the benefit of the novel approach is not great for the total length parameter.

The high level feature of the greatest length of a trail has a smaller relative error than the total length for the novel approach and noiseless data. With noise the median is still advantageous, but the 75% quartile reveals a tendency to measurement outliers. This leads to a demand for careful statistical evaluation of this parameter. With skeletonization this error is more stable towards directedness but overall on a higher level. The impact of noise is common to both approaches. The advantages of the ridge tracking for noiseless and sparse ridges is no surprise. Modelling the feature as a geometrically embedded graph and extracting with respect to similarity to the impression

Table 8: TRAIL, ELLIPSE, noisy TRAIL and noisy ELLIPSE models with $m = 10$ and Algorithm 2

trail-number 25%	10	10	10	10
trail-number 50%	11	11	10	10
trail-number 75%	12	12	11	11
length error 25%	0.5%	0.6%	0.5%	0.5%
length error 50%	0.8%	1.1%	0.9%	1.0%
length error 75%	1.1%	1.9%	1.4%	1.9%
error longest 25%	0.2%	0.3%	0.3%	0.3%
error longest 50%	0.5%	0.5%	0.5%	0.6%
error longest 75%	0.9%	1.3%	0.8%	1.2%
error dispersion 25%	0.1%	0.1%	0.2%	0.2%
error dispersion 50%	0.3%	0.5%	0.3%	0.5%
error dispersion 75%	0.6%	0.9%	0.7%	0.9%
error mean angle 25%	0.1°	0.1°	0.1°	0.3°
error mean angle 50%	0.2°	0.3°	0.4°	0.5°
error mean angle 75%	0.5°	0.7°	1.1°	1.2°

of a human observer leads to a disadvantage, the more textural the feature is. It is more dense with ELLIPSE and less separated from the background with noise. A stabilization of this parameter would require more complex algorithms.

Figure 25 illustrates the main sources of errors for skeletonization as a basis for a **PAST**-coloring. The merging of two trails crossing and the wrong guidance of a trail are disadvantageous for the longest filament parameter. Length and orientation are rather low level features, which are also accessible with less sophisticated feature models. Surprisingly the accuracy does not benefit from the support for the distribution assumption for the ELLIPSE models compared to the TRAIL models.

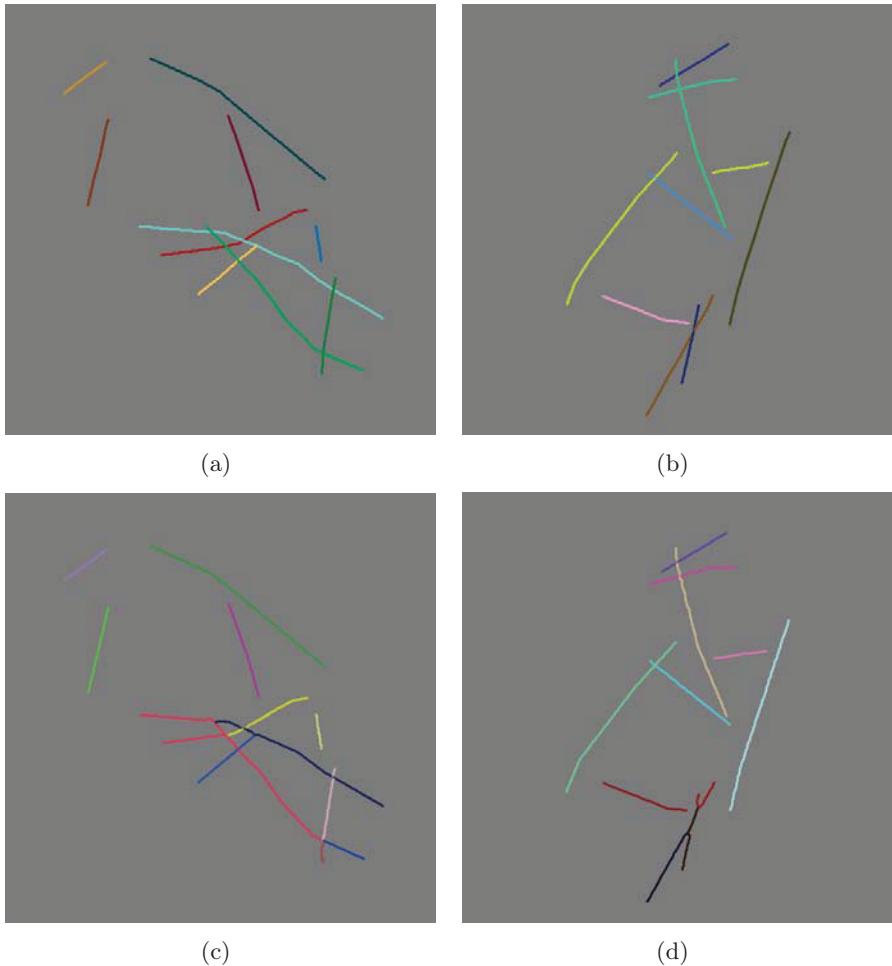


Figure 25: **PAST**-coloring of line-structures in random image in Figure 24(c) processed by ridge tracking, see Section 3.2.3, in part (a) and by skeletonization, see Section 5.2.3, in part (c). Parts (b) and (d) show the result of ridge tracking and skeletonization for the noisy ELLIPSE model image in Figure 24(d).

Table 9: TRAIL, ELLIPSE, noisy TRAIL and noisy ELLIPSE models with $m = 10$ and Algorithms 5 and 6

trail-number 25%	11	11	13	12.5
trail-number 50%	12	12	14	14
trail-number 75%	14	14	17	16
length error 25%	2.7%	2.5%	2.2%	2.1%
length error 50%	4.0%	3.8%	3.4%	3.6%
length error 75%	5.1%	4.9%	4.6%	4.7%
error longest 25%	4.7%	5.6%	6.2%	5.0%
error longest 50%	7.2%	9.9%	12.8%	14.1%
error longest 75%	20.6%	20.2%	28.9%	28.3%
error dispersion 25%	2.2%	3.3%	2.8%	3.5%
error dispersion 50%	4.0%	6.0%	4.5%	6.4%
error dispersion 75%	7.8%	8.9%	7.6%	9.5%
error mean angle 25%	0.7°	0.6°	0.9°	0.7°
error mean angle 50%	1.8°	1.4°	1.7°	1.5°
error mean angle 75%	3.6°	2.8°	3.3°	3.8°

7 Summary and Conclusions

Quantification of ridge features in microscopic images is challenging. An accurate solution is appreciated, because merely qualitative methods suffer from great uncertainty and subjectivity. The problem of detecting ridges with unknown position and connectivity in images is not limited to biological data. Also material inspection of fabrics, recognition of fingerprints and surveying of molecules are treated with similar models. The graph-based approach presented here is little dependent on the image grid by elementary discretization methods of a simple continuous image model. It is shown to cope with exemplary application data and with a random model of it. Additionally this random model highlights the performance of a state of the art method compared to the novel approach. This comparison reveals an improvement with the conducted effort. However, it must be noted that the measurements of accuracy are not identical to the ones in the setting of the state of the art method. The improvement is supported by a confirmation of established hypotheses with the automatic methods.

The application problem to quantify the phenotypical response of the actin filament of osteoblasts produced a fruitful interdisciplinary cooperation, of which only the mathematical side is presented here. The requirements of accurate computer vision based on degraded microscopic image data lead directly to a combinatorial approach for feature detection and indirectly to appropriate methods from mathematical image processing. Whereas mathematical morphology, which is already well established and used for the state of the art, its application for the novel approach is limited to a nonlinear preprocess, the top-hat-transform, to equilibrate the background. Hence, the reduction of information is less crucial than directly proceeding with a segmentation. For noise reduction the model of Rudin, Osher and Fatemi is used in a generalized form and with an efficient numerical realization. The generalization is revealed to be exceptionally suitable for the computer vision task conceived as ridge detection. As total variation of first order has a broad range of applications there is probably a greater potential of the insights concerning total variation of second order.

These tools then facilitate the derivation of geometrically meaningful parameters concerning length and orientation. The measurement of the longest fiber in a network of superposed fibers naturally defines a combinatorial optimization heuristic, which is useful throughout this thesis. The description of orientation preferences in a line-process leads to problems of circular statistics. The development in this field likely allows for a further development of the methods of this thesis. A first generalization of the von Mises distribution keeps the statistic closer to the phenomenon and is shown to be still treatable with efficient estimators.

As a byproduct a software solution entitled FilaQuant was specified to provide a tool for applied scientists with open access. This software is designed

to be extendible for the purpose of other research projects with focus on image data with ridge features. Especially the denoising part might have a broad field of applications. A special challenge is the generalization to 3d and the combination with scale-space ridge detectors. Also the theoretical results can be extended. A continuation of applied mathematical methods for microscopic images requires more emphasis on statistical parameters, for which the feature detection is replaced with different models, according to different image acquisition techniques.

Acknowledgements

This Thesis was funded by the DFG Research Training Group 1505/1 *welisa*. The interdisciplinary environment was a stimulating basis of my research. My supervisors Konrad Engel and Barbara Nebe gave me strong impulses to follow the main tracks of my scientific endeavor. Image processing can be interesting with a set of images and an expert who identifies a feature with a hypothesis about it. The unconscious capabilities of a human observer to accurately tag the information content of an image and the need for an automation in the face of complex biological structures have already been introduced to me by Oliver Schmitt. Claudia Matschegewski introduced me to keeping the goals in perspective. I am grateful to Jochen Merker and Antje Kiesel for the fruitful discussions about analysis and combinatorial optimization out of the scope of this thesis. Also Friedrich Liese always offered me an attentive ear for my curiosity about applied statistical models.

References

- [1] C. Bradford Barber, David P. Dobkin, and Hannu Huhdanpaa. The quickhull algorithm for convex hulls. *ACM Transactions on Mathematical Software*, 22:469–483, 1996.
- [2] Michael Beil, Hans Braxmeier, Frank Fleischer, Volker Schmidt, and Paul Walther. Quantitative analysis of keratin filament networks in scanning electron microscopy images of cancer cells. *Journal of Microscopy*, 220(Pt 2):84–95, 2005.
- [3] Michael Beil, Stefanie Eckel, Frank Fleischer, Hendrik Schmidt, Volker Schmidt, and Paul Walther. Fitting of random tessellation models to keratin filament networks. *Journal of Theoretical Biology*, 241(1):62–72, 2006.
- [4] Benjamin Berkels, Martin Burger, Marc Droske, Oliver Nemitz, and Martin Rumpf. Cartoon extraction based on anisotropic image classification. In *Vision, Modeling, and Visualization Proceedings*, pages 293–300, 2006.
- [5] Harald Birkholz. Extracting the ridge set as a graph for actin filament quantification through confocal laser scanning microscopy. In Jean-Charles Pinoli Johan Debayle Yann Gavet Frédéric Gruy Claude Lambert, editor, *10th International Conference on Quality Control by Artificial Vision*, Proceedings of SPIE, pages 800006–1–800006–8. SPIE, 2011.
- [6] Harald Birkholz. A unifying approach to isotropic and anisotropic total variation denoising models. *Journal of Computational and Applied Mathematics*, 235(8):2502–2514, 2011.
- [7] Harald Birkholz and Konrad Engel. Partition into almost straight trails. *Discrete and Applied Mathematics*, submitted.
- [8] Harald Birkholz, Roger Labahn, Barbara Nebe, and Konrad Engel. P05: automatische auswertung von Aktinfilament-Strukturen in konfokalmikroskopischen bildern. In *BMT 2010 - Reguläre Beiträge*, Rostock-Warnemünde, Germany, 2010.
- [9] Harald Birkholz, Claudia Matschegowski, J. Barbara Nebe, and Konrad Engel. Quantification of actin filament organization by estimating graph structures in confocal microscopic images. In Olaf Dössel and Wolfgang C. Schlegel, editors, *World Congress on Medical Physics and Biomedical Engineering, September 7 - 12, 2009*, volume 25/IV of *IFMBE Proceedings*, pages 1932–1935, Munich, Germany, 2009. Springer.

- [10] Harry Blum. A transformation for extracting new descriptors of shape. In *Models for the perception of speech and visual form : Proceedings of a Symposium (on Models for the Perception of Speech and Visual Form, 1964)*, Cambridge Mass., 1967. M.I.T. Press.
- [11] Antonin Chambolle. An algorithm for total variation minimization and applications. *Journal of Mathematical Imaging and Vision*, 20(1):89–97, 2004.
- [12] Tony F. Chan and Jianhong Shen. *Image processing and analysis : variational, PDE, wavelet, and stochastic methods*. Society for Industrial and Applied Mathematics, Philadelphia, PA, 2005.
- [13] Michel Couplie. Note on fifteen 2d parallel thinning algorithms. In *Internal Report, Université de Marne-la Vallée*, 2006.
- [14] David Eberly. *Ridges in image and data analysis*. Kluwer, Dordrecht, 1996.
- [15] Birgit Finke, Frank Luethen, Karsten Schroeder, Petra D. Mueller, Claudia Bergemann, Marion Frant, Andreas Ohl, and Barbara J. Nebe. The effect if postitively charged plasma polymerization on initial osteoblastic focal adhesion on titanium surfaces. *Biomaterials*, 28:4521–4534, 2007.
- [16] Nicholas I. Fisher. *Statistical analysis of circular data*. Cambridge University Press, Cambridge [u.a.], 1993.
- [17] Rafael C. Gonzalez and Richard E. Woods. *Digital image processing*. Pearson Education. Pearson/Prentice Hall, Upper Saddle River, NJ, 3. international edition, 2008.
- [18] Gösta H. Granlund and Hans Knutsson. *Signal processing for computer vision*. Kluwer Academic Publishers, Dordrecht, 1995.
- [19] Robert M. Haralick. Ridges and valleys on digital images. *Computer Vision, Graphics, and Image Processing*, 22(1):28–38, 1983.
- [20] Jean-Baptiste Hiriart-Urruty and Claude Lemaréchal. *Convex Analysis and Minimization Algorithms I: Fundamentals*, volume 305 of *A Series of Comprehensive Studies in Mathematics*. Springer, Berlin, 1993.
- [21] Charles Hirsch. *Fundamentals of numerical discretization*, volume 1 of *Wiley series in numerical methods in engineering*. Wiley, Chichester, reprint edition, 1997.
- [22] Rainer Höger. Bildbedeutung und strukturelle bildmerkmale. In *Vom Realismus der Bilder : interdisziplinäre Forschungen zur Semantik bildhafter Darstellungsformen*, Magdeburg, 2000. Scriptum-Verlag.

- [23] Richard A. Johnson and Thomas E. Wehrly. Some angular-linear distributions and related regression models. *Journal of the American Statistical Association*, 73(363):602–606, 1978.
- [24] Bela Julesz. Textons, the elements of texture perception, and their interactions. *Nature*, 290(5802):91–97, 1981.
- [25] D. Jungnickel. *Graphs, Networks and Algorithms*. Springer, Berlin, Heidelberg, New York, 2004.
- [26] Michael Krause, Ralph M. Alles, Bernhard Burgeth, and Joachim Weickert. Retinal vessel detection via second derivative of local radon transform. Technical report, Universität des Saarlandes - Fachrichtung 6.1 Mathematik, 2008.
- [27] Tony Lindeberg. Edge detection and ridge detection with automatic scale selection. *International Journal of Computer Vision*, 30:465–470, 1996.
- [28] J. Pablo Marquez. Fourier analysis and automated measurement of cell and fiber angular orientation distributions. *International Journal of Solids and Structures*, 43(21):6413–6423, 2006.
- [29] Yves Meyer. *Oscillating patterns in image processing and nonlinear evolution equations : the fifteenth Dean Jacqueline B. Lewis memorial lectures*, volume 22 of *University lecture series*. American Mathematical Society, Providence, RI, 2001.
- [30] Jason Miller and Jacob D. Furst. The maximal scale ridge. In Mads Nielsen, Peter Johansen, Ole Fogh Olsen, and Joachim Weickert, editors, *Scale-Space*, volume 1682 of *Lecture Notes in Computer Science*, pages 93–104. Springer, 1999.
- [31] Barbara J. G. Nebe, Frank Luethen, Regina Lange, and Beck Ulrich. Interface interactions of osteoblasts with structured titanium and the correlation between physicochemical characteristics and cell biological parameters. *Macromolecular Bioscience*, 7:567–578, 2007.
- [32] Alexander Rigort, David Günther, Reiner Hegerl, Daniel Baum, Britta Weber, Steffen Prohaska, Ohad Medalia, Wolfgang Baumeister, and Hans-Christian Hege. Automated segmentation of electron tomograms for a quantitative description of actin filament networks. *Journal of Structural Biology*, 2011.
- [33] I. Rodrigues, J.M. Sanches, and J. Bioucas Dias. Denoising of medical images corrupted by poisson noise. In *15th IEEE International Conference on Image Processing*, pages 1756–1759, 2008.

- [34] Leonid I. Rudin, Stanley Osher, and Emad Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, 60(1-4):259 – 268, 1992.
- [35] John C. Russ. *The image processing handbook*. A CRC Handbook. Springer, Heidelberg, third edition edition, 1999.
- [36] Otmar Scherzer. Denoising with higher order derivatives of bounded variation and an application to parameter estimation. *Computing*, 60(1):1–27, 1998.
- [37] Jean Serra. *Image analysis and mathematical morphology*. Academic Press, London, 1982.
- [38] Simon Setzer, Gabriele Steidl, and Tanja Teuber. Infimal convolution regularizations with discrete ℓ^1 -type functionals. *Communications in Mathematical Sciences*, 9(3):797–872, 2011.
- [39] Pierre Soille. *Morphological image analysis : principles and applications*. Springer, Berlin, 1999.
- [40] Gabriele Steidl. A note on the dual treatment of higher-order regularization functionals. *Computing*, 76(1):135–148, 2006.
- [41] Gabriele Steidl, Stephan Didas, and Julia Neumann. Splines in higher order tv regularization. *International Journal of Computer Vision*, 70(3):241–255, 2006.
- [42] G. Neville Watson. *A treatise on the theory of Bessel functions*. Cambridge mathematical library. Cambridge University Press, 1995.
- [43] Joachim Weickert. *Anisotropic diffusion in image processing*. Teubner, Stuttgart, 1998.
- [44] Ting Xu, Hongsheng Li, Tian Shen, N. Ojkic, D. Vavylonis, and Xiaolei Huang. Extraction and analysis of actin networks based on open active contour models. In *Biomedical Imaging: From Nano to Macro, 2011 IEEE International Symposium on*, pages 1334–1340, 2011.
- [45] Jerrold H. Zar. *Biostatistical Analysis*. Pearson Education, 5. edition, 2010.
- [46] Mingqiang Zhu and Tony Chan. An efficient primal-dual hybrid gradient algorithm for total variation image restoration. *UCLA CAM Report*, 2008.
- [47] Mingqiang Zhu, Stephen J. Wright, and Tony F. Chan. Duality-based algorithms for total-variation-regularized image restoration. *UCLA CAM Report*, 2008.

Thesen zur Dissertation

„Automatic Quantification of the Actin-Filaments in Microscopic Images“

von Harald Birkholz

1. In der Zellbiologie werden konfokale, Laser-Raster-mikroskopische Aufnahmen vom Aktin-Filament humaner Osteoblasten erstellt, um den Fortschritt der zellulären Entwicklung zu überwachen. Die Aktin-Filamente haben in der frühen Entwicklung die Struktur eines geometrisch eingebetteten Graphen, der in der fokalen Ebene nahe Biomaterialoberflächen kontrastreich abgebildet werden kann.
2. Die Quantifizierung von filamentösen Strukturen anhand von Mikroskopieaufnahmen wird bisher durch Segmentierung anhand von Helligkeitskriterien und Skelettierung im Sinn der mathematischen Morphologie gelöst, siehe [1, 6, 9]. Dabei spielen verschiedene Parameter des Bildinhaltes eine Rolle, die sich in den Anwendungen unterscheiden. Die geschätzten Parameter des Bildinhaltes variieren in ihrer Empfindlichkeit für Fehler der Bildgebung, die der menschliche Betrachter unterbewusst kompensiert. Die Stabilisierung interessanter Parameter folgt einer Abwägung von Kosten und Nutzen.
3. Die für den menschlichen Betrachter gewonnenen Bilddaten bedürfen einer Vorverarbeitung für das maschinelle Sehen. Die Aktin-Filament-Bilder sind gekennzeichnet durch unregelmäßig ausgeleuchteten Hintergrund der Fasern, ein leichtes Poisson-Rauschen, siehe [8]. Die Vorverarbeitung kann unter Erhaltung der Fasern ohne vorherige Erkennung die Störungen reduzieren. Die Beleuchtung des Hintergrundes wird klassisch mit der Zylinderhut-Transformation ausgeglichen, siehe [5]. Das erfordert die Eingabe einer Steuerungsvariablen für die geschätzte Dicke der Filamente.
4. Zur Beseitigung des Rauschens eignet sich in besonderem Maße eine speziell angepasste Variante des Rudin-Osher-Fatemi-Modells. Die Formulierung als konkaves Programm mit quadratischem Ähnlichkeitsterm f in festem Verhältnis $\lambda > 0$ zur totalen Variation zweiter Ordnung für glatte Funktionen auf dem Rechengebiet $\Omega \subset \mathbb{R}^2$ lautet

$$\int_{\Omega} (u - f)^2 d\mathbf{x} + \lambda \int_{\Omega} \|\nabla^2 u\| d\mathbf{x} \rightarrow \min_u .$$

Hierbei ist $\|\nabla^2 u\|$ die Spektralnorm der Hessematrix von $u \in C^2(\Omega)$. Es ergibt sich ein effizientes numerisches Verfahren. Die numerische Behandlung des diskreten Problems zur Minimierung der totalen Variation zweiter Ordnung gelingt durch Ausnutzung der Dualität von konvexen und konkaven Programmen [3]. Das Verfahren eignet sich besonders für die Reduzierung von Störungen vor der Markierung von Graten und Tälern.

5. Das maschinelle Sehen ist mit dieser Vorverarbeitung auf das Erkennen des gefilterten Merkmals zurückgeführt. Dabei schwankt der Kontrast innerhalb des vorverarbeiteten Bildes, so dass einfache Helligkeitskriterien versagen. Zudem variiert die Dicke der im Bild erhaltenen Grate. Das Problem mit unbekannte Dicke und gleich bleibendem Kontrast könnte mit Verfahren der Literatur gelöst werden, siehe [4, 7]. Diese Lösungen bedürfen einer Nachverarbeitung für die bisher keine Grundlage für Aktin-Filamente bekannt ist.
6. Mit der Annahme geringer Schwankungen der Dicke erzeugt ein neu entwickelter Spurensuch-Algorithmus mit einem Vorhersage-Korrektur-Schema effizient einen zusammenhängenden geometrisch eingebetteten Graphen als Markierung der hellen Grate im Bild. Die Markierung erlaubt eine Quantifizierung von Filamentlänge und Filament-Ausrichtung mit hoher Genauigkeit.
7. Die Ausrichtungsparameter können unter von Mises Verteilungsannahme gewonnen werden. Dabei kann eine neu erarbeitete Längengewichtete Variante auf dem Halbkreis verwendet werden, aus der die Parameter durch Maximum-Likelihood Schätzer gewonnen werden.
8. Eine nach anwendungsbezogen sinnvollen Parametern optimale Zerlegung in beinahe gerade Pfade des Merkmalsgraphen liefert die Länge der längsten Faser als stark Kontext-bestimmten Parameter mit hoher Genauigkeit.
9. Die automatische Quantifizierung kann qualitative Beobachtungen von Experten der Zellbiologie anhand kleiner Stichproben bestätigen.
10. Ein Zufallsmodell der Daten gibt die Gestalt der Aktinfilamente so wieder, dass die gleichen Fehlerquellen auftreten und ein Vergleich von Wahrheit und Schätzung möglich ist. Anhand des Zufallsmodells kann der neue Ansatz durch Vorverarbeitung, Spurensuche mittels Vorhersage und Korrektur [2], und Quantifizierung anhand von Fehlerraten mit einem Verfahren durch Segmentierung und Skelettierung [1] aus der Literatur verglichen werden. Der neue Ansatz zeigt experimentell eine Verbesserung der kontextsensitiven Längenschätzung und eine leichte Verbesserung der weiteren Parameter.

11. Die gesamte Lösung zur Aktinfilament-Quantifizierung steht der Öffentlichkeit als frei erhältliche WIN32-Anwendung *FilaQuant* zur Verfügung.

Literatur

- [1] Michael Beil, Hans Braxmeier, Frank Fleischer, Volker Schmidt, and Paul Walther. Quantitative analysis of keratin filament networks in scanning electron microscopy images of cancer cells. *Journal of Microscopy*, 220(Pt 2):84–95, 2005.
- [2] Harald Birkholz. Extracting the ridge set as a graph for actin filament quantification through confocal laser scanning microscopy. In Jean-Charles Pinoli Johan Debayle Yann Gavet Frédéric Gruy Claude Lambert, editor, *10th International Conference on Quality Control by Artificial Vision*, Proceedings of SPIE, pages 800006–1–800006–8. SPIE, 2011.
- [3] Harald Birkholz. A unifying approach to isotropic and anisotropic total variation denoising models. *Journal of Computational and Applied Mathematics*, 235(8):2502–2514, 2011.
- [4] David Eberly. *Ridges in image and data analysis*. Kluwer, Dordrecht, 1996.
- [5] Rafael C. Gonzalez and Richard E. Woods. *Digital image processing*. Pearson Education. Pearson/Prentice Hall, Upper Saddle River, NJ, 3. international edition, 2008.
- [6] Michael Krause, Ralph M. Alles, Bernhard Burgeth, and Joachim Weickert. Retinal vessel detection via second derivative of local radon transform. Technical report, Universität des Saarlandes - Fachrichtung 6.1 Mathematik, 2008.
- [7] Tony Lindeberg. Edge detection and ridge detection with automatic scale selection. *International Journal of Computer Vision*, 30:465–470, 1996.
- [8] I. Rodrigues, J.M. Sanches, and J. Bioucas Dias. Denoising of medical images corrupted by poisson noise. In *15th IEEE International Conference on Image Processing*, pages 1756–1759, 2008.
- [9] John C. Russ. *The image processing handbook*. A CRC Handbook. Springer, Heidelberg, third edition edition, 1999.