Geometric Investigations in Loop Theory and in Image Processing

Doktori (PhD) értekezés

Tomán Henrietta

Témavezető: Dr. Hajdu András és

Dr. Nagy Péter Tibor

Debreceni Egyetem
Természettudományi Doktori Tanács
Informatikai Tudományok Doktori Iskola
Debrecen, 2014.

Geometric Investigations in Loop Theory and in Image Processing

Doktori (PhD) értekezés

Tomán Henrietta

Témavezető: Dr. Hajdu András és

Dr. Nagy Péter Tibor

Debreceni Egyetem
Természettudományi Doktori Tanács
Informatikai Tudományok Doktori Iskola
Debrecen, 2014.

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

Debrecen, 2013. június 25.

.....
a jelölt aláírása

Tanúsítom, hogy Tomán Henrietta doktorjelölt - között a fent megnevezett Doktori Iskola Diszkrét matematika, képfeldolgozás és komputergeometria programjának keretében irányításommal végezte munkáját. Az értekezésben foglalt eredményekhez a jelölt önálló alkotó tevékenységével meghatározóan hozzájárult. Az értekezés elfogadását javasolom.

Debrecen, 2013. június 25.

.....
a témavezető aláírása

Geometric Investigations in Loop Theory and in Image Processing

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

Írta: Tomán Henrietta okleveles matematikus

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

Témavezető: Dr. Hajdu András és Dr. Nagy Péter Tibor

A doktori szigorlati bizottság:

elnök:	Dr. Sztrik János	.....
tagok:	Dr. Hoffmann Miklós	.....
	Dr. Fazekas Gábor	.....

A doktori szigorlat időpontja: 2009. december 15.

Az értekezés bírálói:

.....	.....
.....	.....

A bírálóbizottság:

elnök:	.....	.....
tagok:	.....	.....
	.....	.....
	.....	.....
	.....	.....

Az értekezés védésének időpontja:

Acknowledgements

I would like to express my gratitude to my supervisors, Professor Péter Tibor Nagy and Professor András Hajdu for their magnificent ideas, continuous encouragement during the years of the preparation of this work and for advice and remarks to improve this thesis.

I owe thank to my family for their extreme patience and understanding.

Contents

Chapter 1

Introduction

Differential geometry with its long, fruitful history has found new relevance and produced useful achievements in various fields ranging from e.g. machinery design, visualization to the study of DNA.

Differential geometry has become an expanding field of research through the decades, as it plays an increasingly important role in modern theoretical physics and applied mathematics. Three principle examples are emphasized to show the importance of differential geometry in physics. The general theory of relativity developed by Einstein brought the tools of Riemannian geometry into physics. Moreover, Gauge theories (like Maxwell and Yang-Mills theories) are more recent fields of physics that involve the geometrical objects of connections on principle or vector bundles. A principle bundle associates a Lie group to each point on a manifold. In case of Maxwell theory, the Lie group associated to each point is $U(1)$, while in Yang-Mills theory it is $SU(2)$ or $SU(3)$. The Hamiltonian reformulation of classical mechanics provided the motivation for symplectic geometry as it was originally studied by Arnold.

Differential geometry is concerned with understanding shapes and their properties in terms of calculus. The tools of differential geometry are used e.g. to analyze shapes in computer vision, to process and visualize data on non-flat surfaces in image processing. Diffeomorphic transformations are often applied in computational anatomy to describe and analyze physiological processes guaranteeing the topology of the objects to be preserved. Geometric modelling (including computer graphics), architectural and computer-aided geometric design draw on ideas from differential geometry, as well.

Differential geometry is closely tied up with differential equations. Even in cases when explicit solutions to the relevant differential equations do not exist, numerical solutions can often produce a solution shape. The development of computer algebraic systems makes this feasible even for non-experts in computer programming.

The goal of this dissertation e.g. is to show some further investigations, where (differential) geometric tools are applied to other fields of science and mathematics.

The methods of associative algebra (e.g. group theory, linear algebra) became established tools of differential geometry long ago. Later, the effectiveness of the non-associative algebra was shown e.g. in web theory, affine connection and smooth loops. A Lie group is smooth manifold carrying a group structure whose multiplication and its inverse operation are smooth. Several concepts and results from Lie group theory can be investigated for loops, i.e. for non-associative multiplication, as well.

Next, we focus on the geometrical foundations common to the computer graphics and related areas (e.g. medical imaging and visualization). For example, Diffusion Tensor Imaging (DTI) produces a symmetric, positive definite matrix, at each voxel of an imaging volume. This method can be used to track the white matter fibres in brain, demonstrating the higher diffusivity of water in the direction of the fibre.

Chapter 2

Differentiable $n-$loops

This chapter is devoted to the study of canonical coordinate systems and the corresponding exponential maps of differentiable $n-$ary loops and to the discussion of their differentiability properties.
Canonical coordinate systems can be determined by the canonical normal form of the power series expansion of the $n-$th power map $x \to x \cdot \ldots \cdot x = m(x, x, \ldots, x)$.
The exponential map of a $\mathcal{C}^k-$differentiable local $n-$loop can be constructed as in Lie group theory, determined by the integral curves of vector fields defined by the $i-$th translations of tangent vectors at the unit element of the $n-$loop. The exponential map can be also defined by applying the construction of canonical coordinate systems.

2.1 Introduction

The study of the non$-$associative structures was motivated by some investigations in geometry (e.g. regarding coordinate systems of non$-$Desarguesian planes). Another impulse for W. Blaschke to investigate

the loops and quasigroups came from some topological questions of differential geometry (e.g. the topological behavior of geodesic foliations). In the last decades the theory of loops and quasigroups was developed mainly in the following three directions: in the first one the geometry associated with a loop is considered as an important tool (Baer), the second one treats the theory of loops as a part of the general algebra (Bruck), while the third one handles a loop as a section in the group generated by translations (Albert).

V. D. Belousov has studied the loops, the quasigroups and their associated geometry as abstract objects. The investigation of loops from the perspective of the differential geometry, topological algebra, topological geometry gained importance by A. I. Malcev, K. H. Hofmann, H. Salzmann and M. A. Akivis. L. V. Sabinin has shown how useful the analytic methods can be in the theory of loops.

The theory of differentiable quasigroups has numerous applications to different branches of mathematics and physics. We consider only its application to classical algebraic geometry.

In geometry, the differentiable quasigroups are connected with the theory of multidimensional three-webs, while in algebra they are closely related to the theory of Lie groups.

The canonical coordinate systems of Lie groups are essential tools in the investigation of local properties of group manifolds. They can be generalized for non−associative differentiable loops, as well.

The first study of the expansion of analytical loop multiplication in a canonical coordinate system using formal power series was given in the paper [1] by M. A. Akivis. (cf. [18, Chapter 2])

The convergence conditions of power series expansions of loop multiplications were investigated later in [3]. E. N. Kuzmin treated the local Lie theory of analytic Moufang loops using power series expansion in canonical coordinate systems and gave a generalization of the

classical Campbell-Hausdorff formula.[32]

V. V. Goldberg introduced canonical coordinates using power series expansions in analytic local $n-$ary loops. (cf. [18, Chapter 3])

As it is well-known, differentiable groups are automatically (analytic) Lie groups. In the case of non$-$associative loop theory the class of $\mathcal{C}^k-$differentiable loops contains the class of $\mathcal{C}^l$-differentiable loops for any $k < l; k, l = 0, 1, \ldots, \infty$, as a proper subclass. [45]

The theory of normal forms of $\mathcal{C}^\infty-$differentiable $n-$ary loop multiplications has been investigated by J-P. Dufour and P. Jean. [15] S. Sternberg's linearization theorem has been applied to the coordinate representation of $(n+1)-$webs, which are the differential geometric structures determined by the level manifolds of $n-$ary loop multiplication and its inverse operation.

J. Kozma has defined the canonical coordinates of binary $\mathcal{C}^\infty$-loops by linearizing coordinate systems of the square map $x \to x \cdot x = m(x, x)$ of the loop multiplication m. [29] For Lie groups these canonical coordinate systems coincide with the classical systems defined by one-parameter subgroups.

Now, we consider a natural generalization of Kozma's construction to $\mathcal{C}^k-$differentiable $n-$ary loops. [58], [59] According to Sternberg's linearization theorem the linearizing coordinate system of the $n-$th power map $x \to x \cdot \ldots \cdot x = m(x, x, \ldots, x)$ for the loop multiplication m has the same differentiability property as the $n-$ary loop multiplication map if $k \geq 2$. Hence in the following we will assume that the differentiability class $\mathcal{C}^k$ of the investigated $n-$ary loops satisfies $k \geq 2$. Similar construction for canonical coordinate systems was introduced by V. V. Goldberg in case of analytic $n-$loop multiplications using formal power series expansions. [18, Chapter 3]

2.2 Basic concepts

We give a brief overview of definitions and results concerning loops and webs. Webs (that are also called nets in the literature) form the fundamental elements of the geometric loop theory.

A Lie group is an algebraic group that is also a smooth manifold. Similarly, a loop can be considered as an algebraic and as a differential geometric notion, as well. In algebraic sense, the loop can be defined as a quasigroup with unit element:

Definition 2.2.1. *Let H be a non−empty set with the multiplication $m\colon H^2 \to H$, let $e \in H$ be a given element. Then (H, e, m) is called* ***loop*** *with unit element e if*

1. *$m(e, x) = m(x, e) = x$ for all $x \in H$,*
2. *the equations $m(a, x) = b$ and $m(y, a) = b$ are uniquely solvable for all $a, b \in H$.*

Then the uniquely existing solutions x and y can be determined by the left and right divisions:

$$x = a \setminus b \quad \text{and} \quad y = b/a \ .$$

From the fact that the solution of each equation is unique, it follows that each element occurs exactly once in any row or column of the multiplication table on a finite set. This table is obviously a Latin square of order n, if $|H| = n$. Conversely, a Latin square can be interpreted as a quasigroup.

In the definition of $n-$loop, an $n-$ary multiplication has to satisfy similar properties given in the previous binary case:

Definition 2.2.2. *Let H be a non−empty set with the multiplication $m\colon H^n \to H$, let $e \in H$ be a given element. Then (H, e, m) is called $n-$**loop** with unit element e if*

1. $m(\overset{(1)}{e}, \ldots, \overset{(i-1)}{e}, \overset{(i)}{a}, \overset{(i+1)}{e}, \ldots, \overset{(n)}{e}) = a$ *for all* $a \in H$, $1 \leq i \leq n$, *where* $\overset{(i)}{x}$ *means that the $i-$th argument has the value x,*

2. *the equation $m(a_1, \ldots, a_{i-1}, x, a_{i+1}, \ldots, a_n) = b$ is uniquely solvable for all $a_i \in H, 1 \leq i \leq n$, $b \in H$.*

Then this unique solution x can be determined by the $i-$th division $\delta_i\colon H^n \to H$:

$$x = \delta_i(b; a_1, a_2, \ldots, a_{i-1}, a_{i+1}, \ldots, a_n) \ .$$

In special case, when $n = 2$, we get the classical left and right divisions:

$$\delta_1(b; a) = a \setminus b \qquad \text{and} \qquad \delta_2(b; a) = b/a \ .$$

It follows from these definitions that a group ($n-$group) can be considered as an associative loop ($n-$loop).

In differential geometric sense, we can define the $\mathcal{C}^k-$ differentiable global or local $n-$loop.

Definition 2.2.3. *Let H be a differentiable manifold of class $\mathcal{C}^k$, let $e \in H$ be a given element and let $m\colon H^n \to H$, $\delta_i\colon H^n \to H$ be differentiable maps of class $\mathcal{C}^k$, where $i = 1, \ldots, n$.*
*Then $\mathcal{H} = (H, e, m, \delta_1, \ldots, \delta_n)$ is called $\mathcal{C}^k-$**differentiable** $n-$**ary loop** (or shortly $n-$**loop**) with unit element e if the multiplication m and the $i-$th divisions δ_i $(i = 1, \ldots, n)$ satisfy the following identities:*

1. $m(\overset{(1)}{e},\dots,\overset{(i-1)}{e},\overset{(i)}{a},\overset{(i+1)}{e},\dots,\overset{(n)}{e}) = a$ *for all* $a \in H$, $1 \le i \le n$,

2. $m(a_1,\dots,a_{i-1},\delta_i(b;a_1,\dots,a_{i-1},a_{i+1},\dots,a_n),a_{i+1},\dots,a_n) = b$ *for all* $a_i \in H$, $1 \le i \le n, b \in H$,

3. $\delta_i(m(a_1,\dots,a_n);a_1,\dots,a_{i-1},a_{i+1},\dots,a_n) = a_i$ *for all* $a_i \in H$, $1 \le i \le n$.

In the definition of a $\mathcal{C}^k-$differentiable local $n-$loop, the multiplication and the $i-$th divisions are defined only in a neighborhood of the unit element. Then the implicit function theorem provides the $\mathcal{C}^k-$differentiability of the $i-$th divisions locally around the unit element.

Definition 2.2.4. *If H is a differentiable manifold of class $\mathcal{C}^k$, $e \in H$ is a given element and $m\colon H^n \to H$, $\delta_i\colon H^n \to H$ are differentiable maps of class $\mathcal{C}^k$ $(i = 1,\dots,n)$, which are defined in a neighbourhood of $e \in H$, then $\mathcal{H} = (H, e, m, \delta_1, \dots, \delta_n)$ is called $\mathcal{C}^k-$**differentiable local** $n-$**loop** with unit element e, if the multiplication m and the $i-$th divisions δ_i $(i = 1,\dots,n)$ satisfy the following identities:*

1. $m(\overset{(1)}{e},\dots,\overset{(i-1)}{e},\overset{(i)}{a},\overset{(i+1)}{e},\dots,\overset{(n)}{e}) = a$ *for all* $a \in H$, $1 \le i \le n$,

2. $m(a_1,\dots,a_{i-1},\delta_i(b;a_1,\dots,a_{i-1},a_{i+1},\dots,a_n),a_{i+1},\dots,a_n) = b$ *for all* $a_i \in H$, $1 \le i \le n, b \in H$,

3. $\delta_i(m(a_1,\dots,a_n);a_1,\dots,a_{i-1},a_{i+1},\dots,a_n) = a_i$ *for all* $a_i \in H$, $1 \le i \le n$,

in a neighbourhood of $e \in H$.

We give the definition of $n-$web, that is fundamental concept in geometric loop theory. The close relationship between web geometry and many branches of mathematics has been revealed since 1969.

Definition 2.2.5. *A family of foliations $(F_1, \ldots, F_n)$ on a manifold V in general position is called $n-$**web**.*

In other words: each foliation consists of leaves with the following properties:

1. in each foliation, there exists exactly one leaf through a given point of V,
2. for different leaves L_i, L_i' from the same foliation F_i $(L_i, L_i' \in F_i)$: $L_i \cap L_i' = \emptyset$.

In the beginning, our investigation was motivated by the results of $(n+1)-$webs because $n-$loops are associated with $(n+1)-$webs. If we consider an $(n+1)-$web with a family of foliations $(F_1, \ldots, F_{n+1})$ of codimension q (dimension of the leaves) on a manifold V of dimension qn, we can construct a local $n-$loop and reverse, as well. [2]

This construction is shown for a $(2-)$loop and a $3-$web. Let us consider a $(2-)$loop (H, e, m) and $M = H \times H$. Then three curves $x = a$, $y = b$ and $m(x, y) = m(a, b)$ pass through a point (a, b). Thus the three foliations $x =$ constant, $y =$ constant and $m(x, y) =$ constant form a $3-$web on M.

A $3-$web W on a manifold $M = H \times H$ gives a geometric interpretation to the operation of multiplication in the loop. Namely, let $A \in M$, H_1 and H_2 is a leaf of the foliation F_1 and F_2 passing through A, respectively. The leaves of the third foliation F_3 of W establish

bijective correspondence between the leaves H_1 and H_2 in a neighborhood of the point A. Next, let $u \in H_1$ and $v \in H_2$ be the points on the leaves H_1 and H_2. Then we consider the leaves F_u and F_v of F_1 and F_2 passing through the points u and v, respectively. The leaf $w = constant$ of F_3 passing through the common point $F_u \bigcap F_v$ of the leaves F_u and F_v, defines a point w on the leaf H_1, and this point w corresponds to the product of u and v in the loop: $w = m(u, v) = u \cdot v$.

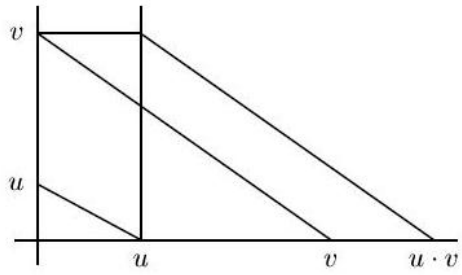


Figure 2.1: The geometric definition of the (coordinate) loop

This construction for the foliations

$$F_1 : x_1 = \text{constant}, F_2 : x_2 = \text{constant}, F_3 : x_1 + x_2 = \text{constant}$$

is shown in Figure 2.1.

J-P. Dufour and P. Jean has proved that any $(n+1)-$web of $\mathbb{R}^{qn}$ of codimension q is smoothly equivalent to a web whose foliations are the following:

$$F_1 : x_1 = \text{constant}, \ldots, F_n : x_n = \text{constant},$$

$$F_{n+1} : x_1 + \ldots + x_n + \alpha(x_1, \ldots, x_n) = \text{constant},$$

where:

$$\alpha(x,\dots,x)=0 \quad \text{and} \quad \alpha(0,\dots,0,\overset{(i)}{x},0,\dots,0)=0$$

for all $1\leq i\leq n$, $x\in H$.

Let (H,e,m) be a smooth binary loop. Then there exist local coordinates near the unit element (vanishing at the unit element) such that:

$$m(x,y)=x+y+\alpha(x,y),$$

where:

$$\alpha(x,x)=\alpha(0,x)=\alpha(x,0)=0$$

for all $x\in H$. It is called **local normal form**.

The theory of normal forms of $\mathcal{C}^\infty-$differentiable $n-$ary loop multiplications has been investigated in the paper of J-P. Dufour and P. Jean. [14]

2.3 Canonical coordinate system

The canonical coordinate systems of Lie groups are essential tools in the investigation of local properties of group manifolds. This concept can be generalized for non−associative differentiable loops, as well.

Definition 2.3.1. *Let $\mathcal{H}=(H,e,m,\delta_1,\dots,\delta_n)$ be a $\mathcal{C}^k-$differentiable local $n-$loop. A coordinate map $\varphi\colon U\to\mathbb{R}^q$ of class $\mathcal{C}^k$ of the open neighbourhood $U\subset H$ of $e\in H$ into the coordinate space $\mathbb{R}^q$ is called* ***canonical coordinate system*** *of $\mathcal{H}$ if $\varphi(e)=0$ and the coordinate function $M\colon\varphi(U)\times\dots\times\varphi(U)\to\mathbb{R}^q$:*

$$M=\varphi\circ m\circ(\varphi^{-1}\times\dots\times\varphi^{-1})$$

of the multiplication map $m\colon H^n \to H$ *satisfies:*

$$M(x, x, \ldots, x) = n\,x \quad \textit{for all} \quad x \in \varphi(U).$$

We will need the following assertions in the investigation of canonical coordinate systems. [55]

First, Poincare's theorem can be rephrased as follows: Given a system of analytic differential equations:

$$dx_i/dt = X_i(x_1, x_2, \ldots, x_n)$$

defined near the origin such that $X_i(0, 0, \ldots, 0) = 0$ and whose matrix of linear terms is diagonalizable with eigenvalues λ_i satisfying:

1. all the eigenvalues λ_i lie in the same open half-plane about the origin,
2. $\lambda_j \neq \sum_{i=1}^{n} m_i\lambda_i$ for any non$-$negative m_i such that $\sum_{i=1}^{n} m_i > 1$,

then there exists an analytic change of coordinates:

$$y_i = y_i(x_1, x_2, \ldots, x_n)$$

transforming the differential equations $dx_i/dt = X_i(x_1, x_2, \ldots, x_n)$ into the linearized form:

$$dy_i/dt = \lambda_i\, y_i.$$

Normal forms for smooth contractions in Euclidean $n-$space, that is, invariants for $\mathcal{C}^n-$contractions under inner automorphisms of the group of local $\mathcal{C}^k-$changes of coordinates can be defined. For the problem of normal forms for differential equations, certain results of Lattes on analytic surface transformations can be generalized (to n dimensions and the non$-$analytic case). A generalization of the results of [54] on invariant curves to n dimensions can also be obtained.

Lemma 2.3.2. *Let ϕ be a local $\mathcal{C}^k-$diffeomorphism of $\mathbb{R}^q$ $(k \geq 2)$ defined in some neighbourhood of $0 \in \mathbb{R}^q$, keeping $0 \in \mathbb{R}^q$ $(q \in \mathbb{N})$ fixed. Let $\phi_*|_{(0)}$ denote the tangent map of ϕ at $0 \in \mathbb{R}^q$. If ϕ satisfies that $\phi_*|_{(0)} = \lambda id_{\mathbb{R}^q}$ with $\lambda \neq 0, 1, -1$ then there exists a unique local $\mathcal{C}^k-$diffeomorphism ρ of $\mathbb{R}^q$ with the following properties:*

1. *it keeps $0 \in \mathbb{R}^q$ fixed,*
2. *$\rho \cdot \phi \cdot \rho^{-1} = \phi_*|_{(0)}$,*
3. *$\rho_*|_{(0)} = id_{\mathbb{R}^q}$.*

Proof. The existence of a local $\mathcal{C}^k-$diffeomorphism ρ of $\mathbb{R}^q$ satisfying the conditions of the assertion follows from Sternberg's linearization theorem for local contractions (cf. [55]), since either the map ϕ or its inverse ϕ^{-1} is a local contraction, the minimum and maximum of the eigenvalues of its tangent map coincide, $k \geq 2$ and it satisfies the so-called resonance condition $\lambda \neq \lambda^m$ for any $m > 1$. The uniqueness of the map ρ follows from the ideas of the proof of Sternberg's theorem, since the difference of two solutions must be a fixed point of a contractive operator on a linear space of differentiable maps. Hence the difference of these solutions is 0. □

Lemma 2.3.3. *Let $\kappa : W \to \mathbb{R}^q$ be a differentiable map of a star shaped neighbourhood $W \subset \mathbb{R}^p$ with $\kappa(0) = 0$. If there exists a real number $0 < r < 1$ such that $\kappa(r\,x) = r\,\kappa(x)$ holds for all $x \in W$, then κ is the restriction of a linear map.*

Proof. Since the map $\kappa\colon W \to \mathbb{R}^p$ is differentiable, we can define the continuous map $\omega\colon W \to \mathbb{R}^p$ satisfying:

$$\kappa(x) = \kappa_*|_{(0)}(x) + \|x\|\omega(x), \quad \omega(0) = 0.$$

Furthermore we have:

$$\kappa(r\,x) = r\,(\kappa_*|_{(0)}(x) + \|x\|\omega(r\,x)),$$

$$r\,\kappa(x) = r\,(\kappa_*|_{(0)}(x) + \|x\|\omega(x)).$$

From $\kappa(r\,x) = r\,\kappa(x)$ it follows that $\omega(x) = \omega(r^m\,x)$ for any natural number $m \in \mathbb{N}$ and hence:

$$\omega(x) = \lim_{m\to\infty} \omega(r^m) = \omega(0) = 0$$

for all $x \in W$. □

Theorem 2.3.4. *Let $\mathcal{H} = (H, e, m, \delta_1, \ldots, \delta_n)$ be a $\mathcal{C}^k-$differentiable local $n-$loop with $k \geq 2$. Then there exists a canonical coordinate system for $\mathcal{H}$.*

If (U, φ) is a canonical coordinate system of $\mathcal{H}$, then for any linear map $\tau : \mathbb{R}^q \to \mathbb{R}^q$ the pair $(U, \tau \circ \varphi)$ is a canonical coordinate system of $\mathcal{H}$, as well.

If $\varphi\colon U \to \mathbb{R}^q$ and $\psi\colon U \to \mathbb{R}^q$ are coordinate maps of canonical coordinate systems of $\mathcal{H}$ defined on the same neighbourhood U, then $\varphi \circ \psi^{-1}$ is the restriction of a linear map $\mathbb{R}^q \to \mathbb{R}^q$.

Proof. Let $(\overline{U}, \overline{\varphi})$ be a coordinate system of $\mathcal{H}$, let $\overline{M}$ be the coordinate function of the local $n-$loop multiplication m with respect to $(\overline{U}, \overline{\varphi})$. Then we introduce the map $\overline{G}\colon \overline{\varphi}(\overline{U}) \to \mathbb{R}^q$ defined by $\overline{G}(x) = \overline{M}(x, x, \ldots, x)$. Clearly one has $\overline{G}(0) = 0$.

From $\overline{M}(0, \ldots, 0, x, 0, \ldots, 0) = x$ it follows that the tangent map $\overline{G}_*|_0\colon \mathbb{R}^q \to \mathbb{R}^q$ of $\overline{G}$ at the point 0 satisfies $\overline{G}_*|_0 = n\, id_{\mathbb{R}^q}$. The map $\overline{G}$ is of class C^k in a neighborhood of 0 and it has an inverse map in a neighborhood of 0 of the same class $\mathcal{C}^k$. We can apply Lemma 2.3.2 for $\overline{G}^{\,-1}$, so there exists a local $\mathcal{C}^k-$diffeomorphism ρ keeping $0 \in \mathbb{R}^q$ fixed such that $(\rho \circ \overline{G} \circ \rho^{-1})_*|_0 = \rho \circ \overline{G} \circ \rho^{-1}$.

We consider the composed map $\varphi = \rho \circ \overline{\varphi}$ as the coordinate map of a new coordinate system (U, φ) with a suitable neighborhood U. The coordinate function of the multiplication map $m\colon H^n \to H$ satisfies $M = \rho \circ \overline{M} \circ \rho^{-1}$. Let Q be the following function:

$$Q\colon \mathbb{R}^q \to \mathbb{R}^q \times \mathbb{R}^q \times \cdots \times \mathbb{R}^q\colon x \mapsto Q(x) = (x, x, \ldots, x).$$

Then we have:

$$G = M \circ Q = (\rho \circ \overline{M} \circ \rho^{-1})(\rho \circ Q \circ \rho^{-1}) = \rho \circ \overline{G} \circ \rho^{-1} =$$

$$= (\rho \circ \overline{G} \circ \rho^{-1})_*|_0 = n\, id_{\mathbb{R}^q}.$$

Hence (U, φ) is a canonical coordinate system of $\mathcal{H}$.

In the next step, the second statement will be proved. For a canonical coordinate system (U, φ) of the local $n-$loop $\mathcal{H}$, the coordinate function:

$$M\colon \varphi(U) \times \cdots \times \varphi(U) \to \mathbb{R}^q\colon M = \varphi \circ m \circ (\varphi^{-1} \times \cdots \times \varphi^{-1})$$

of the multiplication map $m\colon H^n \to H$ satisfies $M(x, x, \ldots, x) = n\,x$ for all $x \in \varphi(U)$. Then for arbitrary linear map $\tau\colon \mathbb{R}^n \to \mathbb{R}^n$ we have:

$$\tau \circ M(\tau^{-1}y, \ldots, \tau^{-1}y) = \tau(n\,\tau^{-1}y) = n\,y, \quad y \in \tau \circ \varphi(U).$$

It follows that if $\psi = \tau \circ \varphi$, then (U, ψ) is also a canonical coordinate system of $\mathcal{H}$.

Finally, let (U, φ) and (U, ψ) be canonical coordinate systems of $\mathcal{H}$ given on the same neighbourhood U. Then let M_φ and M_ψ be the coordinate functions of the multiplication map $m\colon H^n \to H$, respectively. For all $x \in \varphi(U)$ and $y \in \psi(U)$, we have:

$$M_\varphi(x, x, \ldots, x) = n\,x \quad \text{and} \quad M_\psi(y, y, \ldots, y) = n\,y.$$

We denote $\kappa\colon \psi(U) \to \varphi(U)\colon \kappa = \varphi \circ \psi^{-1}$. Since:

$$M_\varphi(\kappa(y), \kappa(y), \ldots, \kappa(y)) = \kappa(M_\psi(y, y, \ldots, y)),$$

then we obtain: $n\,\kappa(y) = \kappa(n\,y)$. After putting $z = n\,y$ we get $\kappa(r\,z) = r\,\kappa(z)$ for all $z \in \psi(U)$, where $r = \frac{1}{n}$. It follows by Lemma 2.3.3 that the map $\kappa = \psi \circ \varphi^{-1}$ is the restriction of a linear map. □

Example 2.3.5. *The local loop−multiplication $f(x, y) = x^2y(x - y) + x + y$ is non−associative, consequently it is not a group multiplication. It is defined in a canonical coordinate system, as $f(x, x) = 2x$ for all $x \in \varphi(U)$.*

2.4 Exponential map

A Lie group G is a differentiable group, it consists of a group structure and a manifold structure such that the multiplication map and the inversion map are differentiable. The simplest Lie group is the additive group of the real numbers. A differentiable homomorphism of $(\mathbb{R}, +)$ into a given Lie group G is called a one−parameter subgroup of G.

To each tangent vector $X \in T_eG$ at the identity e, there exists a unique one−parameter subgroup $\phi_X : \mathbb{R} \to G$ with X as its initial velocity. There exists a canonical bijection between T_eG and the set of all one−parameter subgroups. We define $exp(X) = \phi_X(1)$ for each $X \in T_eG$. The map $exp : T_eG \to G : X \mapsto \phi_X(1)$ is called the exponential map of G.

Observe that $\mu_c : \mathbb{R} \to \mathbb{R} : \mu_c = c \cdot t$ is a Lie homomorphism, hence $\phi_X \circ \mu_c$ is also a one−parameter subgroup of G. It follows from the chain rule of differentiation that $\phi_X \circ \mu_c = \phi_{cX}$, so $exp(tX) = \phi_{tX}(1) = \phi_X(t)$ holds for all $t \in \mathbb{R}$.

Since $exp : T_eG \to G$ is a local diffeomorphism, we can use the exponential map to define local coordinates for G. Let $X_1, \ldots, X_n$ be a basis of T_eG. The mapping $exp(\alpha_1 X_1 + \ldots + \alpha_n X_n)$ defines a local diffeomorphism between the real n−tuples $((\alpha_1, \ldots, \alpha_n) \in \mathbb{R}^n)$ and $g \in G$ for g being sufficiently near the identity. Then we can consider $\alpha : U \subset G \to \mathbb{R}^n$ as a coordinate mapping, where U is an arbitrarily small neighborhood of the origin. Using this coordinate chart and the left translation, we can construct an entire atlas for the Lie group G. The functions $(\alpha_1, \ldots, \alpha_n)$ around the identity are called canonical coordinates related to the basis $X_1, \ldots, X_n$.

There are several natural possibilities for the definition of the exponential map exp: $W \to H$ with $0 \in W \subset T_eH$ of $\mathcal{C}^k$−differentiable local n−loops. One of them is analogous to the usual construction

applied in Lie group theory, namely the map *exp* could be determined by the integral curves of vector fields defined by the $i-$th translations of tangent vectors at the unit element of the $n-$loop.

In binary Lie groups, these curves are one$-$parameter subgroups, but for smooth loops it is not always the case. (cf. J. Kozma [29]) Another disadvantage of such a construction is that one can expect only $\mathcal{C}^{k-1}-$differentiability of the map $exp\colon W \to H$ with $0 \in W \subset T_eH$ which is determined by integral curves of $\mathcal{C}^{k-1}-$differentiable vector fields defined by the $i-$th translations of tangent vectors.

Definition 2.4.1. *Let λ_x^i denote the $i-$th translation with x. Let $\gamma_v^i(t)$ be the integral curve of the differential equation:*

$$\dot{\gamma}_v^i(t) = (\lambda^i_{\gamma(t)})_* v, \text{ where } \gamma_v^i(0) = e, \dot{\gamma}_v^i(0) = v.$$

*Then the mapping $exp^{(i)} : T_eH \to H$, where $exp^{(i)}(v) = \gamma_v^i(1)$, is called $i-$**th exponential map**.*

Example 2.4.2. *Let $f(x,y) = x + y + x^2y(x-y)$.*
This loop$-$multiplication is considered in a canonical coordinate system, as $f(x,x) = 2x$ for all $x \in \varphi(U)$. This loop$-$multiplication does not generate a group because it is non$-$associative.
The exponential maps $exp^{(1)}$ and $exp^{(2)}$ are different, because:

$$f_y'(x,0) = 1 + x^3$$
$$f_x'(0,y) = 1.$$

Then $\gamma^1(t)$ and $\gamma^2(t)$ satisfy the following equations:
$ln(|1+\gamma^1(t)|^{\frac{1}{3}} \cdot |\gamma^1(t)^2 - \gamma^1(t) + 1|^{\frac{1}{2}}) + \frac{1}{\sqrt{3}} \cdot \arctan(\frac{2}{\sqrt{3}}\gamma^1(t) - \frac{1}{\sqrt{3}}) = t$
and

$$\gamma^2(t) = t.$$

An alternative natural method for the definition of the exponential map is given by using the construction of canonical coordinate systems studied in the previous section.

Theorem 2.4.3. *Let $\mathcal{H} = (H, e, m, \delta_1, \ldots, \delta_n)$ be a $\mathcal{C}^k-$differentiable local $n-$loop. Then there exists a unique local $\mathcal{C}^k-$diffeomorphism* $\exp\colon W \to H$, *where W is a neighbourhood of $0 \in T_eH$, such that the following conditions hold:*

(i) $\exp(n\,x) = m(\exp(x), \ldots, \exp(x))$,

(ii) $\exp(0) = e$,

(iii) $\exp_*|_0 = id_{T_eH}$.

Proof. Let $\varphi\colon U \to \mathbb{R}^q$ be the coordinate map of a canonical coordinate system (U, φ) of the local $n-$loop $\mathcal{H}$. According to Theorem 2.3.4, $(U, \varphi_*|_0^{-1} \circ \varphi)$ is also a canonical coordinate system of $\mathcal{H}$, where the vector space T_eH is the coordinate space and $\varphi_*|_0^{-1} \circ \varphi\colon U \to T_eH$ is the coordinate map. Let $W \subset \varphi_*|_0^{-1} \circ \varphi(U)$ be a neighbourhood of $0 \in T_eH$. Then the coordinate function $M\colon W \times \cdots \times W \to T_eH$:

$$M = \varphi_*|_0^{-1} \circ \varphi \circ m \circ ((\varphi_*|_0^{-1} \circ \varphi)^{-1} \times \cdots \times (\varphi_*|_0^{-1} \circ \varphi)^{-1})$$

of the multiplication map $m\colon H^n \to H$ satisfies $M(x, \ldots, x) = n\,x$, or equivalently:

$$m(\varphi^{-1} \circ \varphi_*|_0(x), \ldots, \varphi^{-1} \circ \varphi_*|_0(x)) = \varphi^{-1} \circ \varphi_*|_0(nx)$$

for any $x \in W$. Moreover one has $(\varphi^{-1} \circ \varphi_*|_0)_*|_0 = id_{T_eH}$. Hence we can define $\exp = \varphi^{-1} \circ \varphi_*|_0$ and this map satisfies the conditions given in the assertion.

For proving the uniqueness, let us assume the map $\widetilde{\exp}\colon W \to H$ fulfills the conditions (i), (ii) and (iii). Then $(\widetilde{\exp}(W), \widetilde{\exp}^{-1})$ is a canonical coordinate system of the $n-$loop $\mathcal{H}$ and according to Theorem 2.3.4 the map $\widetilde{\exp}^{-1} \circ \exp\colon W \to T_eH$ is the restriction of a linear map $\varepsilon\colon T_eH \to T_eH$. Since both of the maps $\widetilde{\exp}$ and exp satisfy even the condition (iii), then the linear map $\varepsilon\colon T_eH \to T_eH$ must be the identity map. Hence $\widetilde{\exp} = \exp\colon W \to H$, which proves that the map $\exp\colon W \to H$ is determined uniquely. □

Theorem 2.4.4. *Let* $\exp\colon W \to H$, $\exp'\colon W' \to H'$ *be the corresponding exponential maps for the* $\mathcal{C}^k-$*differentiable local* $n-$*loops* $\mathcal{H} = (H, e, m, \delta_1, \ldots, \delta_n)$ *and* $\mathcal{H}' = (H', e', m', \delta'_1, \ldots, \delta'_n)$*, respectively, where* $W \subset T_eH$ *and* $W' \subset T_{e'}H'$*. If* $\alpha\colon \mathcal{H} \to \mathcal{H}'$ *is a continuous local homomorphism then the composed map* $\exp'^{-1} \circ \alpha \circ \exp\colon W \to T_{e'}H'$ *is locally linear.*

Proof. Let us consider the $\mathcal{C}^k-$differentiable binary local loops $\widetilde{\mathcal{H}}$ and $\widetilde{\mathcal{H}'}$ which are determined by the multiplication and division maps of $\mathcal{H}$ and $\mathcal{H}'$ in such a way that in the multiplication and division functions the $j-$th variable ($j \geq 3$) is replaced by the identity element $e \in H$ and $e' \in H'$, respectively. The map $\alpha\colon H \to H'$ is clearly a continuous local loop homomorphism. According to the result of R. Bödi and L. Kramer [11], the map $\alpha\colon H \to H'$ is $\mathcal{C}^k-$differentiable.

Then according to Lemma 2.3.3, we have:

$$\exp'^{-1} \circ \alpha \circ \exp(nx) = n \exp'^{-1} \circ \alpha \circ \exp(x),$$

or equivalently:

$$\exp'^{-1} \circ \alpha \circ \exp(ry) = r \exp'^{-1} \circ \alpha \circ \exp(y)$$

with $y = nx$ and $r = \frac{1}{n}$, and it implies the assertion. □

2.5 Conclusion

In this section we give the generalized definition of $n-$ary loop, first in algebraic sense, then we have defined the $\mathcal{C}^k-$differentiable (local) $n-$ary loop in differential geometrical sense, as well. The concept of canonical coordinate system, which is an essential tool for Lie groups, has been generalized for non$-$associative differentiable loops. It has been shown that there always exists a canonical coordinate system for a $\mathcal{C}^k-$differentiable local $n-$loop ($k \geq 2$).

There are several natural options for the definition of the exponential map of $\mathcal{C}^k-$differentiable local $n-$loops. One of these methods is analogous to the usual construction in Lie group theory, namely the map *exp* could be determined by the integral curves of vector fields defined by the $i-$th translations of tangent vectors at the unit element of the $n-$loop. We can also define the exponential map using the construction of canonical coordinate systems by proving its unique existence.

Chapter 3

Diffusion Tensor Imaging

In the next two chapters it has been shown how useful the tools of (differential) geometry can be in medical imaging and in image processing problems, as well. For example, by the previous chapter, the problem of image registration can be interpreted as finding a parameterizing velocity field v, where the diffeomorphic transformation $\exp(v)$ minimizes a distance between moving and fixed image with respect to a desired smoothness of the transformation.

In this chapter, the directional information of diffusion tensor maps is used by fiber tracking to estimate connection pathways in white matter. When the diffusion is anisotropic, a scalar diffusion measure is insufficient for describing diffusion properties. In this case, the diffusion can be characterized by a second-order diagonally symmetric tensor, called the diffusion tensor. This tensor model of diffusion is able to get full description of the directional diffusion information: molecular mobility along each direction and correlation between these directions. The diffusion displacement profile may be represented as an ellipsoid with the length of principal axes determined by the eigen-

values of the diffusion tensor and the directions given by eigenvectors of the diffusion tensor.

3.1 Introduction

Diffusion Tensor Imaging (DTI) is an emerging Magnetic Resonance Imaging (MRI) technique based on water diffusion.[8] The spatial properties of molecular diffusion processes can be characterized by diffusion tensor imaging.[43], [64] The application of this technique to the central nervous system has revealed that the diffusion of water in white matter is anisotropic.[44] This directionality has been attributed to constraints imposed upon water motion by the ordered structure of axons and myelin sheaths.[10], [48], [49]

The white matter of human brain has a complex structure and plays an essential role in brain function. In spite of the fact, that a fair amount of information is available today about white matter, not all the aspects of its structure are completely known and understood. We know even less about how the white matter structure is affected by neurological diseases, tumors or traumas.

Not only the magnitude of anisotropy but the orientation in which water preferentially diffuses also can be quantified using diffusion tensor imaging. By combining these two parameters (anisotropy and orientation), DTI provides new and unique opportunities for studying the white matter architecture.

Furthermore, MRI provides access to both superficial and deep organs with high resolution and does not interfere with the diffusion process itself: diffusion is an intrinsic physical process that is totally independent of the MR effect or the magnetic field.

Fiber tracking, also called White Matter Tractography (WMT), uses the directional information of diffusion tensor maps to estimate connection pathways in white matter. This method is presented in [37].

The diffusion tensor field can be simplified to the vector field of the main eigenvector by fiber tracking. If this vector field is considered as a velocity field dropping a free particle on it, then a trajectory will be followed due to the velocity field. The found trajectory can be visualized as a fiber that represents a bundle of nerve fibres in the brain or muscle fibres.

The diffusion model assumes homogeneity and linearity of the diffusion within each voxel. Some diffusion anisotropy measures such as the fractional anisotropy can be computed from the diffusion tensor. Moreover, the principal direction of the diffusion tensor can be used to gather the white-matter connectivity of the brain.

Nowadays, several brain pathologies may be best detected by regarding particular measures of anisotropy and diffusivity. The diffusion process (by Brownian motion) causes water molecules to move out from a central point, and in the anisotropic (isotropic) case the surface of an ellipsoid (sphere) is reached gradually. The ellipsoid formalism operates also as a mathematical method to organize tensor data.

The ellipsoid itself has a principal long axis and then two more small axes describing its width and depth. All three axes are perpendicular to each other crossing at the center point of the ellipsoid. The axes of the ellipsoid represent the eigenvectors of the diffusion tensor and the length of each axis is related to one of the eigenvalues of the diffusion tensor. The length of the longest one pointing along the axon direction is denoted by λ_1 and the two small axes have lengths of λ_2 and λ_3. In the setting of the DTI tensor ellipsoid, each of them can be considered as a measure of the diffusivity along each of the three primary axes of the ellipsoid.

The diffusivity along the principal axis, λ_1 is also called the longitudinal diffusivity (the axial or even the parallel diffusivity). Historically, this is the closest to what Richards originally measured with the vector length.

The diffusivities in the two minor axes are often averaged to produce a measure of radial (or perpendicular) diffusivity $\frac{\lambda_2+\lambda_3}{2}$. This quantity characterizes the degree of restriction due to membranes and other effects and it is proved to be a sensitive measure of degenerative pathology in some neurological conditions.

Another widely used measure summarizing the total diffusivity is the trace $\lambda_1 + \lambda_2 + \lambda_3$, that is the sum of the eigenvalues. After dividing this sum by three we have the mean diffusivity $\frac{\lambda_1+\lambda_2+\lambda_3}{3}$.

Our aim is to reconstruct the fiber tracts of the human brain from measurements of fiber orientation and visualize them on the image of the brain. [62], [63] Generally, the surface model clipped by orthogonal sections (coronal, axial and sagital) is shown. We are capable to visualize the surface model clipped by (even more than the usual three) planes having arbitrary directions. It is able to promote the recognition of the brain diseases and after the diagnosis it helps to select the appropriate way of cure. When surgical intervention is needed, the point and the direction of the permeation can be determined more exactly. This component can be parallelized by using the shading languages. [50]

3.2 Theoretical background

During an MRI scanning process, radio waves are sent through the brain which are 10 000 to 30 000 times stronger than Earth's magnetic field. This method forces the nuclei into a different position and after moving back into their place they send out radio waves of their

own. These signals are picked up by the scanner and their strength as numeric values are recorded into a file.

Diffusion MRI measures the diffusion of water molecules in biological tissues. In an isotropic medium (e.g. inside a glass of water), water molecules naturally move randomly according to Brownian motion. However, in biological tissues the diffusion may be anisotropic (the diffusion properties vary with orientation). The recent development of Diffusion Tensor Imaging [57] enables diffusion to be measured in multiple directions and the fractional anisotropy in each direction to be calculated for each voxel. The most important concepts are reviewed, while a complex overview of anisotropic water diffusion is presented by Beaulieu. [9]

It is well-known that the diffusion in white matter is the largest along fiber directions. When diffusion is anisotropic, a scalar diffusion measure is insufficient for describing diffusion properties. It has been shown that the diffusion in this case can be described by a second-order diagonally symmetric tensor, called the diffusion tensor. This tensor model of diffusion can be used to get full description of the directional diffusion information: molecular mobility along each direction and correlation between these directions.

$$D = \begin{pmatrix} D_{xx} & D_{xy} & D_{xz} \\ D_{yx} & D_{yy} & D_{yz} \\ D_{zx} & D_{zy} & D_{zz} \end{pmatrix}$$

The six independent elements of the symmetric diffusion tensor D can be estimated from a series of diffusion-weighted images. When diffusion weighted measurements are performed along N directions, the following matrix equation can be constructed:

$$B\vec{d}^{\,T} = \vec{A}^{\,T} \ ,$$

where:

$$\vec{A} = \begin{pmatrix} \ln \frac{S_1}{S_0} & \ln \frac{S_2}{S_0} & \dots & \ln \frac{S_N}{S_0} \end{pmatrix}$$

is the vector of the corresponding logarithmic signal ratios and:

$$B = \begin{pmatrix} \vec{b_1} \\ \vec{b_2} \\ \vdots \\ \vec{b_N} \end{pmatrix}$$

includes the influences of all the encoding gradients. [46]

The same matrix can be applied simultaneously to describe the shape and orientation of an ellipse. Futhermore, we can use this matrix in a third way for matrix mathematics to sort out eigenvectors and eigenvalues as explained below.

3.3 The process of calculation

In practice, 25 files are provided usually containing the results of diffusion weighted measurements, and we need also a separate file describing the baseline data. All of them serve as input data of the algorithm and they are referred as volume of voxels further on.

3.3.1 Diffusion tensor

During the calculation special volume iterators are used, that makes more convenient to reach all the data belonging to the same voxel at the same time in a very effective way. This makes also possible that the algorithm remains independent from the size of volumes. The names of input files are used as command line parameters, that is why the number of input volumes can be handled dynamically. A function was constructed to determine the six tensor elements of a voxel in the following way:

1. If the input data make possible the gradients themselves are yielded or default values are used instead.

2. The components of vector $\vec{A}$ are calculated based upon the signal values. (S_0 denotes the value derived from the baseline data.)

3. The equation itself is solved by using the appropriate function of GSL. The GNU Scientific Library (GSL) is a numerical library for C and C++ programmers, which is a free software under the GNU General Public License and provides a wide range of mathematical routines. The selected GSL function is able to find the least squares solution to our over-determined system.

4. Finally the resulted six values (the six independent elements of diffusion tensor) are to be stored in appropriate position of new volumes.

3.3.2 Parameters of diffusion ellipsoid

A principal frame of directions (x', y' and z') can be defined by the eigenvectors of the diffusion tensor for each voxel. The diffusion displacement profile may be represented as an ellipsoid with the length of principal axes described by the tensor eigenvalues λ_1, λ_2 and λ_3 (principal diffusivities) and the directions given by the tensor eigenvectors ($\vec{e_1}$, $\vec{e_2}$ and $\vec{e_3}$). Figure 3.1 shows the geometric meanings of the computed values.

The diffusion eigenvectors are generally not aligned with the laboratory frame. In the principal component frame, the displacements along x', y' and z' appear uncorrelated and the diagonal elements of the tensor are equal to tensor eigenvalues. The major axes are given by the diffusion tensor eigenvectors. The length of each axis is proportional with the square root of the tensor eigenvalues.

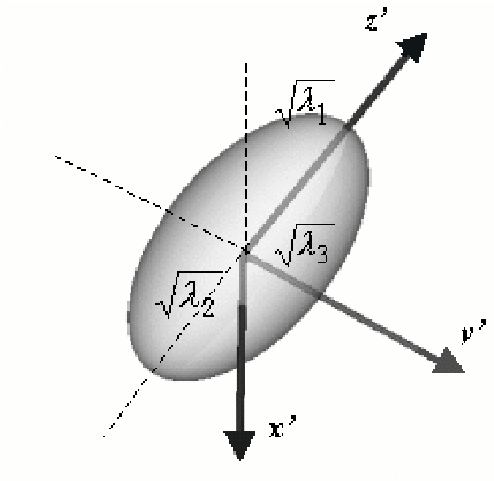


Figure 3.1: The diffusion ellipsoid.

Depending on the relation between the eigenvalues volume, three types of diffusion and corresponding ellipsoidal shapes can be differentiated:

a) Isotropic diffusion: $\lambda_1 \approx \lambda_2 \approx \lambda_3$
Diffusion in gray matter and fluids generally appears isotropic. The corresponding diffusion ellipsoid has a spherical shape.

b) Planar diffusion: $\lambda_1 \approx \lambda_2 \gg \lambda_3$
Planar diffusion is generally associated with diffusion in sheets or it may describe regions of crossing fibers. The corresponding diffusion ellipsoid has a special disk shape.

c) Prolate diffusion: $\lambda_1 \gg \lambda_2 > \lambda_3$
Prolate or uniaxial diffusion is observed in highly organized white

matter regions. The corresponding diffusion ellipsoid has a prolate shape.

White matter structures (corpus callosum and corticospinal tract) are generally characterized by uniaxial diffusion. Planar diffusion is dominant in regions of crossing or fanning fibers. For the visualization of the stream directions at each voxel ellipsoids are used. These ellipsoids can be incorporated with the brain model. Figure 3.2 shows how the brain can be clipped by three arbitrary planes that is not common in usual softwares.

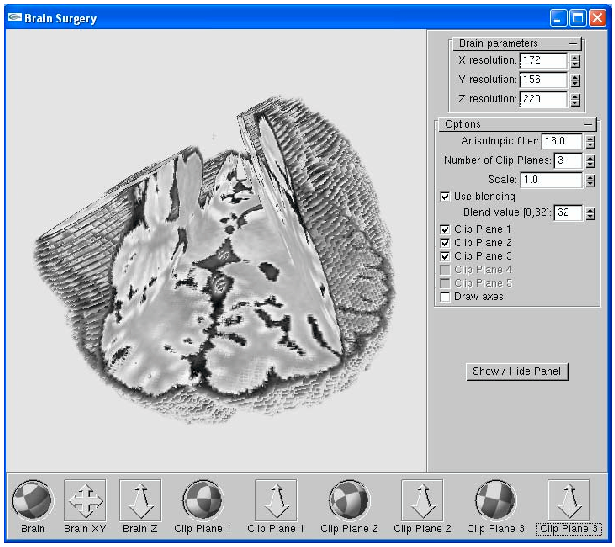


Figure 3.2: Brain clipped by three arbitrary positioned planes.

After the computation, the fibers can be built up by the diffusion ellipsoids [43]. In case of fiber crossings, only the stronger track is

followed. For the dynamic modelling of the fibers, a particle system is applied. [41] The molecule groups inspected by the measurements can be modelled by points moving along the actual fiber. These fibers make up a complex brain structure. For example, the tumors distort the fibers, so change in the fiber structure prognosticates the illness.

Moreover, we can handle more pictures (e.g. one individual and one picture that is common for a population) and blend them together in any ratio. Besides, we can enhance a strip of the brain around any given intensity value.

3.4 Conclusion

The fiber tracts of the human brain can be reconstructed from the measurements of fiber orientation and can be visualized on the image of the brain. Generally, the surface model clipped by orthogonal sections (coronal, axial and sagital) is shown. We are capable to visualize the surface model clipped by (even more than the usual three) planes having arbitrary directions.

The standard algorithms are improved and new methods are developed considering the branching and merging problems, as well. The algorithms can be massively parallelized by using heavily the shader languages (vertex, geometry and pixel shaders).

Beyond the completed new methods, the ellipsoids and the particle system can be integrated with the brain model having arbitrary clipping planes.

Chapter 4

Generalization of the majority voting scheme

Generating ensembles from multiple individual classifiers is a popular approach to raise the accuracy of the decision. As a rule for decision making, majority voting is a usually applied model. In this chapter, we generalize the classical majority voting by incorporating probability terms $p_{n,k}$ to constrain the basic framework. These terms control whether a correct or false decision is made if k correct votes are present among the total number of n.

This generalization is motivated by object detection problems, where the members of the ensemble are image processing algorithms giving their votes as pixels in the image domain. In this scenario, the terms $p_{n,k}$ can be specialized by a geometric constraint. Namely, the votes should fall inside a region matching the size and shape of the object to vote together. We give several theoretical results in this new model for both dependent and independent classifiers, whose individ-

ual accuracies may also differ. As a real world example, we present our ensemble-based system developed for the detection of the optic disc in retinal images. For this problem, experimental results are shown to demonstrate the characterization capability of this system.

4.1 Introduction

Ensemble-based systems are rather popular to raise the decision accuracy by combining the responses of different sources (voters, classifiers). Regarding pattern recognition, the idea of combining the decisions of multiple classifiers has also been studied [34]. As corresponding examples, we can mention neural networks [12], [21], decision trees [26], sets of rules [4] and other models [23], [24], [66]. As a specific application field, now we will focus on object detection in digital images which is a vivid field, as well. [16], [47], [53]

A usual way for information fusion is to consider the majority of the votes of the classifiers as the basis of the decision. The current literature is quite rich regarding both theoretical results and applications of such systems (ensembles). Strong focus is set to the combination of votes of binary (correct/false) values. The related decision may take place based on simple majority [21],[30],[31], weighted majority [30], or using some other variants. [25],[38]

In the research of majority voting, a cardinal issue is the assumptions on the dependency of the voters. Several results are achieved for independent voters, and the minimal and maximal accuracies of such majority voting systems are also studied for the dependent case. In this chapter, we investigate how such voting systems behave if we apply some further constraints on the votes. Namely, we generalize the classical majority voting scheme by introducing real values

$0 \leq p_{n,k} \leq 1$ for the probability that a good decision is made if we have k correct votes out of the n ones. In other words, in our case it will be possible that a good decision is made even if the good votes are in minority (less than half).

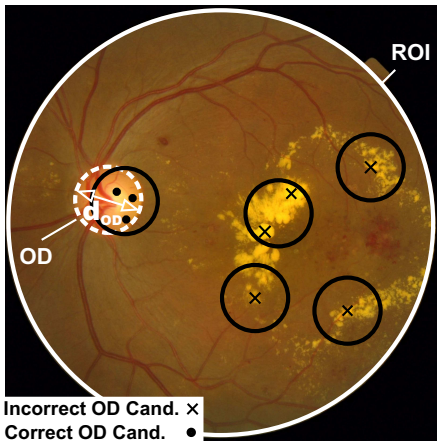


Figure 4.1: The optic disc (OD) of diameter d_{OD} in a retinal image and the OD center candidates (3 correct, 5 false) of individual detector algorithms. Candidates inside the black circles can vote together for possible OD locations.

The development of this new model is motivated by a retinal image processing problem – the detection of the optic disc (OD), which appears as a bright circular patch within the region of interest (ROI) in a retinal image (see Figure 4.1). Namely, in a former work it was observed that organizing more individual OD detector algorithms into an ensemble may raise detection accuracy [22]. In the voting system applied here, each individual OD algorithm votes in terms of a single pixel as its candidate for the OD center. The application of exist-

ing majority voting models are not adequate here, since they consider only the correctness of the votes, which concerns falling into the true OD region in this scenario. However, in our case, the spatial behavior of the votes is also important, since they vote together for a specific location of the OD, only if they fall within a region matching the OD geometry. Consequently, we should consider discs of diameter of the OD $d_{OD} \in \mathbb{R}_{\geq 0}$ covering the candidates of the individual detector algorithms as shown in Figure 4.1. The diameter d_{OD} can be derived by averaging the manual annotations made by clinical experts on a dataset and can be adjusted to the resolution of the image. As a final decision, the disc having diameter d_{OD} with maximal number of candidates included is chosen for the OD location. In this combined system, we can make a good decision even if the false candidates have majority such as in the case illustrated in Figure 4.1. A bad decision is made only when a subset of false candidates with larger cardinality than the number of correct ones can be covered by a disc having diameter d_{OD}.

In this chapter, we propose the generalization of the classical majority voting model by incorporating the probability terms $p_{n,k}$ mentioned before. With an appropriate geometric constraint, our generalized model can be specialized to be applicable for the above detection scenario, as well. Namely, the corresponding values $p_{n,k}$ will be adjusted by requiring that the candidates should fall inside a disc of a fixed diameter d_{OD} to vote together. With the help of this model, we can characterize our detector ensemble and gain information on further improvability issues, as well.

The rest of the chapter is organized as follows. Section 4.2 recalls the basic concepts of classical majority voting, which will provide fundamentals for our more general framework. In section 4.3, we show how to incorporate the probability terms $p_{n,k}$ to constrain the basic

formulation. We present theoretical results for the case of independent voters. Since in applications independent detector algorithms can hardly be expected, we also generalize the method to the dependent case in section 4.4. As a main focus, we investigate the possible lowest and highest accuracy of constrained ensembles. Moreover, we both consider equal and different individual accuracies for the members of the ensemble. Section 4.5 contains our empirical results regarding a real world application (optic disc detection), where we apply this new model to characterize our current OD detector ensemble. In section 4.6, the final decision rule of the ensemble is modified to result in further improvement of the system accuracy. In section 4.7, we discuss our results and draw some conclusions regarding other test datasets and detection problems.

4.2 Majority voting

Let $D_1, D_2, \ldots, D_n$ be a set of classifiers (voters), the $i-$th classifier $D_i : \Lambda \to \Omega$ $(i = 1, \ldots, n)$, where Λ can be any domain, and Ω is a set of finite class labels. The majority voting rule assigns the class label supported by the majority of the classifiers $D_1, \ldots, D_n$ to $\alpha \in \Lambda$. Usually, ties (same number of different votes) are broken randomly.

In the literature, the classifier outputs are generally expected to be independent. It means that for any subset of classifiers $(D_{i_1}, \ldots, D_{i_k})$:

$$P(D_{i_1} = \omega_{i_1}, \ldots, D_{i_k} = \omega_{i_k}) = \prod_{j=1}^{k} P(D_{i_j} = \omega_{i_j}), \tag{4.1}$$

where $\{i_1, i_2, \ldots, i_k\} \subseteq \{1, 2, \ldots, n\}$ $(1 \leq k \leq n)$, $\omega_i \in \Omega$ is the label suggested for $\mathbf{x}$ by the classifier D_i.

In [31] Kuncheva et al. discuss exhaustively the following special case. Let n be odd, $|\Omega| = 2$ (each classifier has a binary (correct/false) output value) and all classifiers are independent and have the same classification accuracy p. A correct class label is given by majority voting if at least $\lceil \frac{n}{2} \rceil$ classifiers give correct answers. The majority voting rule with independent classifier decisions gives an overall correct classification accuracy calculated by the following formula:

$$P = \sum_{k=\lceil \frac{n}{2} \rceil}^{n} \binom{n}{k} p^k (1-p)^{n-k}. \tag{4.2}$$

The following result which is known as the Condorcet Jury Theorem (1785) supports the intuition that we can expect improvement over the individual accuracy p only when p is higher than 0.5.

- If $p > 0.5$, then P is monotonically increasing and $P \to 1$ as $n \to \infty$.
- If $p < 0.5$, then P is monotonically decreasing and $P \to 0$ as $n \to \infty$.
- If $p = 0.5$, then $P = 0.5$ for any natural number n.

This result proves that the majority voting method is guaranteed to give a higher accuracy than the individual classifiers if the classifiers are independent and $p > 0.5$ holds for the individual accuracy.

Several interesting results can be found in [34] applying the majority voting in pattern recognition. Different methods for combining the decisions of classifiers for different types of outputs have been described by Lam and Suen. [36] The combination methods can be applied by various architectures. For each type of output, combination methods have been developed from simple operations requiring no prior training, to complex and highly tailored methods that can

produce higher recognition rates. However, these better recognition rates may be accompanied by higher costs in terms of computation requirements, quantity of training data, and difficulty of theoretical analysis.

Lam and Suen proceed to analyze the case of even n and the effect on the ensemble accuracy of adding or removing classifiers. For example, it can confidently be predicted that an even number $2n$ of classifiers would produce more reliable combined recognition results than can be obtained by adding another classifier, or by eliminating one of the classifiers. This conclusion is valid whether the classifiers are independent or not. Shapley and Grofman note in [52] that the result is valid even for unequal p, provided the distribution of the individual accuracies p_i is symmetrical about the mean.

4.3 Generalization to constrained voting

As it has been discussed in the introduction, we generalize the classical majority voting approach by considering some constraints that must be also met by the votes. To give a more general methodology beyond geometric considerations, we model this type of constrained voting by introducing values $0 \leq p_{n,k} \leq 1$ describing the probability of making a good decision, when we have exactly k good votes from the n voters. Then, in section 4.5 we will adopt this general model to our practical problem with spatial constraints.

As we have summarized in the introduction, several theoretical results are achieved for independent voters in the current literature, so we start with generalizing them to this case. However, in the vast majority of applications, we cannot expect independency among algorithms trying to detect the same object. Thus, later we extend the model to the case of dependent voters with generalizing such formerly investigated concepts that have high practical impact, as well.

4.3.1 The independent case

In our model, we consider a classifier D_i with accuracy p_i as a random variable η_i of Bernoulli distribution, i.e.:

$$P(\eta_i = 1) = p_i, \quad P(\eta_i = 0) = 1 - p_i \quad (i = 1, \ldots, n).$$

Here $\eta_i = 1$ means correct classification by D_i. In particular, the accuracy of D_i is just the expected value of η_i, that is, $E\eta_i = p_i$ $(i = 1, \ldots, n)$.

Let $p_{n,k}$ $(k = 0, 1, \ldots, n)$ be given real numbers with:

$$0 \leq p_{j0} \leq \cdots \leq p_{jj} \leq 1 \quad (j = 1, \ldots, n),$$

and let the random variable ξ be such that:

$$P(\xi = 1) = p_{n,k} \quad \text{and} \quad P(\xi = 0) = 1 - p_{n,k},$$

where $k = |\{i \ : \ \eta_i = 1\}|$. That is, ξ represents the modified majority voting of the classifiers $D_1, \ldots, D_n$: if k out of the n classifiers give a correct vote, then we make a good decision (i.e. we have $\xi = 1$) with probability $p_{n,k}$.

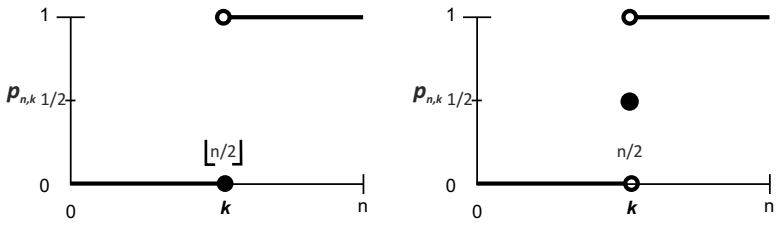
Note that, in the special case, where:

$$p_{n,k} = \begin{cases} 1, & \text{if } k > \frac{n}{2}, \\ \frac{1}{2}, & \text{if } k = \frac{n}{2}, \\ 0, & \text{otherwise}, \end{cases} \tag{4.3}$$

this model results in the classical majority voting scheme.

The values $p_{n,k}$ as a function of k corresponding to the classical majority voting can be observed in Figure 4.2 for both an odd and an even n, respectively.

The ensemble accuracy of the classical majority voting system is shown in Table 4.1 for different number of classifiers n for some equal individual accuracies p (see also [30]).



(a) n is odd (b) n is even

Figure 4.2: The graph of $p_{n,k}$ for classical majority voting for (a) an odd, and (b) an even number of voters n.

	$n = 3$	$n = 5$	$n = 7$	$n = 9$
$p = 0.6$	0.6480	0.6826	0.7102	0.7334
$p = 0.7$	0.7840	0.8369	0.8740	0.9012
$p = 0.8$	0.8960	0.9421	0.9667	0.9804
$p = 0.9$	0.9720	0.9914	0.9973	0.9991

Table 4.1: Ensemble accuracy for classical majority voting.

As the very first step of our generalization, we show that similarly to the individual voters, ξ is of Bernoulli distribution, as well. We also provide its corresponding parameter q, that represents the accuracy of the ensemble in our model.

Lemma 4.3.1. *The random variable ξ is of Bernoulli distribution with parameter q, where:*

$$q = \sum_{k=0}^{n} p_{n,k} \left(\sum_{\substack{I \subseteq \{1,\dots,n\} \\ |I|=k}} \prod_{i \in I} p_i \prod_{j \in \{1,\dots,n\} \setminus I} (1 - p_j) \right). \tag{4.4}$$

Proof. Since for any $k \in \{0, 1, \ldots, n\}$ we have:

$$P(|\{i \ : \ \eta_i = 1\}| = k) = \sum_{\substack{I \subseteq \{1,\ldots,n\} \\ |I|=k}} \prod_{i \in I} p_i \prod_{j \in \{1,\ldots,n\} \setminus I} (1 - p_j),$$

the statement immediately follows from the definition of ξ. □

The special case assuming equal accuracy for the classifiers received strong attention in the literature, so we investigate this case first. That is, in the rest of section 4.3, we suppose that $p = p_1 = \ldots = p_n$. Then, (4.4) reads as:

$$q = \sum_{k=0}^{n} p_{n,k} \binom{n}{k} p^k (1-p)^{n-k}. \tag{4.5}$$

Thus, if n is odd then by the particular choice (4.3) for the values $p_{n,k}$, we get $q = P$, where P is given in (4.2). In order to have our generalized majority voting model be more accurate than the individual decisions, we have to guarantee that $q \geq p$. The next statement yields a guideline along this way.

Proposition 4.3.2. *Let $p_{n,k} = \frac{k}{n}$ $(k = 0, 1, \ldots, n)$. Then, we have $q = p$, and consequently $E\xi = p$.*

Proof. Since by Lemma 4.3.1 the random variable ξ is of Bernoulli distribution with parameter q, we have $E\xi = q$. Thus, we just need to show that $q = p$ whenever $p_{n,k} = \frac{k}{n}$ $(k = 0, 1, \ldots, n)$. By our settings, from (4.5) we have:

$$q = \sum_{k=0}^{n} \frac{k}{n} \binom{n}{k} p^k (1-p)^{n-k} = \frac{1}{n} \sum_{k=0}^{n} k \binom{n}{k} p^k (1-p)^{n-k} = p.$$

Observe that the last sum just expresses the expected value np of a random variable of binomial distribution with parameters (n, p). Thus, we have $q = p$, and the statement follows. □

The above statement shows that if the probabilities $p_{n,k}$ increase uniformly (linearly), then the ensemble has the same accuracy as the individual classifiers. As a trivial consequence we obtain the following corollary.

Corollary 4.3.3. *Let $p_{n,k} \geq \frac{k}{n}$ for all $k = 0, 1, \ldots, n$. Then $q \geq p$, and consequently $E\xi \geq p$.*

The next result helps us to compare our model constrained by the terms $p_{n,k}$ with the classical majority voting scheme.

Theorem 4.3.4. *Suppose that $p \geq \frac{1}{2}$ and for any k with $0 \leq k \leq \frac{n}{2}$ we have:*

(i) $p_{n,k} + p_{n,n-k} \geq 1$,

(ii) $p_{n,n-k} \geq \frac{n-k}{n}$.

Let q be given by (4.5). *Then, $q \geq p$, and consequently $E\xi \geq p$.*

Proof. We can write:

$$q = \sum_{k=0}^{n} p_{n,k} \binom{n}{k} p^k (1-p)^{n-k} = \sum_{k=0}^{\lfloor \frac{n}{2} \rfloor} \left(p_{n,k} \binom{n}{k} p^k (1-p)^{n-k} + \right.$$

$$\left. + p_{n,n-k} \binom{n}{n-k} p^{n-k} (1-p)^k \right) + p_{n,\frac{n}{2}} \binom{n}{\frac{n}{2}} p^{\frac{n}{2}} (1-p)^{\frac{n}{2}}.$$

Here if n is odd, then the last term should be considered to be zero.

Now by our assumptions $p \geq \frac{1}{2}$, together with (i) and (ii), using also the identities $\binom{n}{k} = \binom{n}{n-k}$ and $\frac{k}{n} + \frac{n-k}{n} = 1$, for any k with $0 \leq k < \frac{n}{2}$ we have:

$$
\begin{aligned}
&p_{n,k}\binom{n}{k}p^k(1-p)^{n-k} + p_{n,n-k}\binom{n}{n-k}p^{n-k}(1-p)^k \geq \\
&\quad \geq (1-p_{n,n-k})\binom{n}{k}p^k(1-p)^{n-k} + p_{n,n-k}\binom{n}{n-k}p^{n-k}(1-p)^k = \\
&\quad = \frac{k}{n}\binom{n}{k}p^k(1-p)^{n-k} + \frac{n-k}{n}\binom{n}{n-k}p^k(1-p)^{n-k} + \\
&\quad + \ p_{n,n-k}\binom{n}{n-k}\left(p^{n-k}(1-p)^k - p^k(1-p)^{n-k}\right) \geq \\
&\quad \geq \frac{k}{n}\binom{n}{k}p^k(1-p)^{n-k} + \frac{n-k}{n}\binom{n}{n-k}p^k(1-p)^{n-k} + \\
&\quad + \frac{n-k}{n}\binom{n}{n-k}\left(p^{n-k}(1-p)^k - p^k(1-p)^{n-k}\right) = \\
&\quad = \frac{k}{n}\binom{n}{k}p^k(1-p)^{n-k} + \frac{n-k}{n}\binom{n}{n-k}p^{n-k}(1-p)^k.
\end{aligned}
$$

In the last inequality, we use (ii) and the non-negativity of the term $p^{n-k}(1-p)^k - p^k(1-p)^{n-k}$. Furthermore, if n is even, by (ii) we also have:

$$p_{n,\frac{n}{2}}\binom{n}{\frac{n}{2}}p^{\frac{n}{2}}(1-p)^{\frac{n}{2}} \geq \frac{\frac{n}{2}}{n}\binom{n}{\frac{n}{2}}p^{\frac{n}{2}}(1-p)^{\frac{n}{2}}.$$

Thus, we obtain:

$$q \geq \sum_{k=0}^{n}\frac{k}{n}\binom{n}{k}p^k(1-p)^{n-k} = p.$$

Here, the last equality follows from the proof of Proposition 4.3.2. Since $E\xi = q$, we have the inequality $E\xi \geq p$. □

As a specific case, we obtain the following corollary concerning the classical majority voting scheme [31].

Corollary 4.3.5. *Suppose that n is odd, $p \geq \frac{1}{2}$ and:*

$$p_{n,k} = \begin{cases} 1, & \text{if } k > \frac{n}{2}, \\ 0, & \text{otherwise.} \end{cases}$$

holds for all $k = 0, 1, \ldots, n$. Then $q \geq p$, and consequently $E\xi \geq p$.

Proof. Observing that by the above choice for the values $p_{n,k}$ both properties (i) and (ii) of Theorem 4.3.4 are satisfied, the statement immediately follows from Theorem 4.3.4. □

Of particular interest is the case, when the ensemble makes exclusively good decisions after t executions. That is, we are curious to know the conditions to have a system with accuracy 100%. So write $\xi^{\otimes t}$ for the random variable obtained by repeating ξ independently t times, and counting the number of one values (correct decisions) received, where t is a positive integer. Then, as it is well-known, $\xi^{\otimes t}$ is a random variable of binomial distribution with parameters (t, q) with q given by (4.5). Now we are interested in the probability $P(\xi^{\otimes t} = t)$. In case of using an individual classifier D_i (that is, a random variable η_i) with any $i = 1, \ldots, n$, we certainly have $P(\eta_i^{\otimes t} = t) = p^{\,t}$. To make the ensemble better than the individual classifiers, we need to choose the probabilities $p_{n,k}$ so that $P(\xi^{\otimes t} = t) \geq p^{\,t}$. In fact, we can characterize a much more general case. For this purpose we need the following lemma, due to Gilat [17].

Lemma 4.3.6. *For any integers t and s with $1 \leq s \leq t$ the function:*

$$f(x) = \sum_{k=s}^{t} \binom{t}{k} x^k (1-x)^{t-k}$$

is strictly monotone increasing on $[0, 1]$.

Note that, for any $x \in [0, 1]$ we obviously have:

$$\sum_{k=0}^{t} \binom{t}{k} x^k (1-x)^{t-k} = 1.$$

As a simple consequence of Lemma 4.3.6, we obtain the following result.

Theorem 4.3.7. *Let t and s be integers with $1 \leq s \leq t$. Then:*

$$P(\xi^{\otimes t} \geq s) \geq P(\eta_1^{\otimes t} \geq s),$$

if and only if $q \geq p$, i.e. $E\xi^{\otimes t} \geq tp$.

Proof. Let t and s be as given in the statement. Then, we have:

$$P(\xi^{\otimes t} \geq s) = \sum_{k=s}^{t} \binom{t}{k} q^k (1-q)^{t-k},$$
$$P(\eta_1^{\otimes t} \geq s) = \sum_{k=s}^{t} \binom{t}{k} p^k (1-p)^{t-k}.$$

Thus, by Lemma 4.3.6, we obtain $P(\xi^{\otimes t} \geq s) \geq P(\eta_1^{\otimes t} \geq s)$, if and only if, $q \geq p$, and the theorem follows. □

4.4 The dependent case

We have assumed earlier that the coordinates η_i of η are independent random variables. The classifiers are generally considered independent if they produce independent errors. In our special application, different algorithms detecting the optic disc are regarded as classifiers. These algorithms can not be considered independent in all cases because it can happen that the performance of the algorithms is based on very similar conditions. In case of dependent algorithms we have to decide how to measure the dependencies of the algorithms. We can consider the joint distribution of the outputs of the algorithms that is a well-known method to calculate the dependency. It can be shown that, similarly to the independent case, the overall performance of the system is equal to the individual accuracies if linear p_{nk} is given.

Theorem 4.4.1. *Let $\eta = (\eta_1, \ldots, \eta_n)$ be an $n-$dimensional random variable, where $E\eta_i = p$ $(i = 1, \ldots, n)$. We consider the joint distribution of $(\eta_1, \ldots, \eta_n)$ such that*

$$c_{a_1,\ldots,a_n} = P(\eta_1 = a_1, \ldots, \eta_n = a_n).$$

Let $p_{n,k} = \frac{k}{n}$ $(k = 0, 1, \ldots, n)$. Then we have $E\xi = p$.

Proof. It follows from rearranging the sums in the following way:

$$E\xi = \sum_{k=0}^{n} \sum_{a_1+\ldots+a_n=k} \frac{k}{n}\, q_{a_1,\ldots,a_n} = \frac{1}{n} \sum_{i=1}^{n} \sum_{a_i=1} q_{a_1,\ldots,a_n} =$$

$$= \frac{1}{n} \sum_{i=1}^{n} P(\eta_i = 1) = p.$$

□

More realistically, the different classifiers in general make errors with different probabilities.

Remark 4.4.2. *If the random variables η_i of η have different expected values, that is $E\eta_i = p_i$ $(i = 1, \ldots, n)$, then we get that:*

$$E\xi = \frac{p_1 + p_2 + \ldots + p_n}{n},$$

if $p_{n,k} = \frac{k}{n}$ $(k = 0, 1, \ldots, n)$.

It means that the overall performance of the system is equal to the average of the individual accuracies if linear $p_{n,k}$ is considered.

The calculations of the majority voting error for such a system (with classifiers making errors with different probabilities) is no longer as easily tractable. Namely, it requires probability calculations for all combinations of classifiers being in error in majority voting sense. Computationally, this process is very complex (complexity of the order of 2^n) and calculations of majority voting errors tend to be intractable very quickly even for small values of n.

In this section, we investigate how dependencies among the voters influence the accuracy of the ensemble; for related results, see e.g. [5, 30]. For this purpose, we generalize some concepts that were introduced for classical majority voting to measure the extremal behavior (minimal and maximal accuracies) of an ensemble. First, we consider *pattern of success* and *pattern of failure* [30] which are such realizations of the votes in a series of experiments that lead to the possible highest and lowest accuracy of the ensemble, respectively. It is worth noting that to define these measures, a rather serious restriction considering discretization of the model is needed to be applied. Namely, not only the accuracies of the individual classifiers are given, but also the precise numbers of successful decisions during the experiment are

fixed: e.g. for a classifier having accuracy $p = 0.6$ we consider 6 correct votes in 10 experimental runs.

Though there are some results in the literature for the case of different accuracies p_i of the classifiers D_i (or, in other words, for the case $E\eta_i = p_i$ $(i = 1, \ldots, n)$) (see e.g. [21], [40], [65] and the references there), the vast majority of the results (such as e.g. in [31]) concern the case $p = p_1 = \ldots = p_n$. Hence in section 4.4.1, we shall make the latter assumption, too. However, in section 4.4.2, we give a much more general framework which handles both dependencies without the restriction considering discretization, and also different accuracies of classifiers that makes the model realistic for applications.

4.4.1 Pattern of success and pattern of failure

In this section, we suppose that the individual classifier accuracies coincide ($p = p_1 = \ldots = p_n$). Repeat the experiments $\eta_1, \ldots, \eta_n$ t times, with some positive integer t, and write $\eta_i^{(j)}$ for the $j-$th realization of η_i $(j = 1, \ldots, t)$. Suppose (as a rather strong, but standard assumption) that we have:

$$|\{j \ : \ \eta_i^{(j)} = 1\}| = r \quad \text{for all } i = 1, \ldots, n. \tag{4.6}$$

Here r is a positive integer with $r = np$. We are interested in the behavior (accuracy) of ξ repeated t times, or in other words in the value $E\xi^{\otimes t}$, under the condition (4.6). Write $\xi^{(j)}$ for the $j-$th realization of ξ $(j = 1, \ldots, t)$. Then, we clearly have $E\xi^{\otimes t} = E\xi^{(1)} + \ldots + E\xi^{(t)}$.

The number of one values is fixed for η_i, however, their positions can be freely changed. For simplicity, we shall describe the situation by a table T of size $n \times t$: in the $(i, j)-$th entry $T(i, j)$ of T we write 0 or 1, according to the actual value of $\eta_i^{(j)}$ $(1 \leq i \leq n, 1 \leq j \leq t)$. Our first result in this interpretation concerns the case of linear $p_{n,k}$.

Proposition 4.4.3. *If $p_{n,k} = \frac{k}{n}$ for all $k = 0, 1, \ldots, n$, then we have that $E\xi^{\otimes t} = r$.*

Proof. Denote the number of ones in the $j-$th column of the table T by u_j for $j = 1, \ldots, t$. Then, we have $E\xi^{(j)} = \frac{u_j}{n}$. Thus:

$$E\xi^{\otimes t} = E\xi^{(1)} + \ldots + E\xi^{(t)} = \frac{u_1}{n} + \ldots + \frac{u_t}{n}. \tag{4.7}$$

Since $u_1 + \ldots + u_t$ is just the total number of ones in T, we have:

$$u_1 + \ldots + u_t = nr. \tag{4.8}$$

Combining (4.7) and (4.8) we obtain $E\xi^{\otimes t} = r$, and the statement follows. □

In view of the proof of Proposition 4.4.3, we see that in case of a general system $p_{n,k}$ we have:

$$E\xi^{\otimes t} = \sum_{j=1}^{t} p_{n,u_j}. \tag{4.9}$$

Then, to describe the pattern of success (the highest accuracy) and the pattern of failure (the lowest accuracy), we need to maximize and minimize the above quantity, respectively.

Our next result concerns the pattern of success. Here we consider the problem only under some further assumptions, which in fact are not necessary to study and describe the situation as it will be shown in section 4.4.2. However, the statement together with its proof already show the basic idea for construction. Furthermore, the former results usually consider these assumptions, so in this way our model can be fitted to the existing literature, as well. In section 4.4.2, we describe the general method, which works without any technical restrictions.

Theorem 4.4.4. *Let the probabilities $p_{n,k}$ be arbitrary, up to $p_{n,0} = 0$. Let $k_1 \neq 0$ be an index such that $\frac{p_{n,k_1}}{k_1} \geq \frac{p_{n,k}}{k}$ for all $k = 1, \ldots, n$. Then, $E\xi^{\otimes t} \leq \frac{nrp_{n,k_1}}{k_1}$. Further, if $tk_1 = nr$ then the maximum can be attained.*

Proof. Using (4.9) and our assumption $\frac{p_{n,k_1}}{k_1} \geq \frac{p_{n,k}}{k}$ for all $k = 1, \ldots, n$, we get:

$$E\xi^{\otimes t} = \sum_{j=1}^{t} p_{n,u_j} = \sum_{\substack{j=1 \\ u_j \neq 0}}^{t} u_j \frac{p_{n,u_j}}{u_j} \leq \sum_{j=1}^{t} u_j \frac{p_{n,k_1}}{k_1} =$$

$$= \frac{p_{n,k_1}}{k_1} \sum_{j=1}^{t} u_j = \frac{nrp_{n,k_1}}{k_1},$$

which implies the first part of the statement.

Assume now that we also have $tk_1 = nr$. Fill in the $n \times t$ table T with zeros and ones arbitrarily, such that we have r ones in each row. If there is a column containing less than k_1 ones, then by $tk_1 = nr$ there is another column with more than k_1 ones. Write j_1 and j_2 for the indices of these columns, respectively. Then there exists a row say with index i, such that $T(i, j_1) = 0$ and $T(i, j_2) = 1$. Change these zero and one values, and continue this process as long as possible. Since $tk_1 = nr$, finally we end up with a table T containing r ones in each row and k_1 ones in each column. Then, we have:

$$E\xi^{\otimes t} = \sum_{j=1}^{t} p_{n,k_1} = tp_{n,k_1} = tk_1 \frac{p_{n,k_1}}{k_1} = \frac{nrp_{n,k_1}}{k_1},$$

and the theorem follows. □

Our next theorem describes the pattern of failure, in a similar fashion as the previous statement.

Theorem 4.4.5. *Let the probabilities $p_{n,k}$ be arbitrary, up to $p_{n,0} = 0$. Let $k_2 \neq 0$ be an index such that $\frac{p_{n,k_2}}{k_2} \leq \frac{p_{n,k}}{k}$ for all $k = 1, \ldots, n$. Then, $E\xi^{\otimes t} \geq \frac{nrp_{n,k_2}}{k_2}$. Further, if $tk_2 = nr$ then the minimum can be attained.*

Proof. Since the proof follows the same lines as that of Theorem 4.4.4, the details are omitted. □

Similarly to the independent case in section 4.3.1, we also investigate the case, when only good decision is made by the ensemble. In other words, we would like to describe the situation, where:

$$P(\xi^{\otimes t} = t) = \prod_{j=1}^{t} p_{n,u_j} \tag{4.10}$$

is maximal. Note that, in this case one can easily obtain a table T with $P(\xi^{\otimes t} = t) = 0$. So now finding the minimum (i.e. investigating the pattern of failure) does not make sense.

For the special case of $p_{n,k} = \frac{k}{n}$, we have the following result.

Theorem 4.4.6. *Let $p_{n,k} = \frac{k}{n}$ for all $k = 0, 1, \ldots, n$, and assume that $nr \geq t$. Then $P(\xi^{\otimes t} = t)$ is maximal for the tables T in which:*

$$\left\lfloor \frac{nr}{t} \right\rfloor \leq u_j \leq \left\lceil \frac{nr}{t} \right\rceil \quad (1 \leq j \leq t),$$

where u_j denotes the number of ones in the $j-$th column of T. Further, all these tables T can be explicitly constructed.

Proof. Let T be an arbitrary table having r ones in each row such that T has no column consisting only of zeros. Since $nr \geq t$, such a table T exists (and can be easily constructed). In view of the proof of Proposition 4.4.3, for the corresponding $\xi^{\otimes t}$ in (4.10) we have:

$$P(\xi^{\otimes t} = t) = \frac{1}{n^t} \prod_{j=1}^{t} u_j.$$

Then $(u_{j_1} - 1)(u_{j_2} + 1) > u_{j_1} u_{j_2}$ clearly holds, if for some indices $1 \leq j_1, j_2 \leq t$ we have $u_{j_1} - u_{j_2} \geq 2$. Hence moving a one from the j_1-th column to the j_2-th column (keeping its row; just as at the end of the proof of Theorem 4.4.4), the new value for $P(\xi^{\otimes t} = t)$ will be larger than the previous one. Continuing this process as long as possible, finally we obtain a table T, where for any indices $1 \leq j_1, j_2 \leq t$ we have $|u_{j_1} - u_{j_2}| \leq 1$. Obviously, this is equivalent to the following inequalities:

$$\left\lfloor \frac{nr}{t} \right\rfloor \leq u_j \leq \left\lceil \frac{nr}{t} \right\rceil \quad (1 \leq j \leq t).$$

Observing that for all such tables T the values $P(\xi^{\otimes t} = t)$ coincide, and these tables differ from each other only by a permutation of their columns, the theorem follows. □

Note that, if $t > nr$ then T necessarily has a column with all zero entries, whence $P(\xi^{\otimes t} = t) = 0$ in this case. For general values $p_{n,k}$, we have the following result.

Theorem 4.4.7. *Let the probabilities $p_{n,k}$ be arbitrary, up to $p_{n,0} = 0$ and $p_{n,k} > 0$ for $0 < k \leq n$. Let $k_0 \neq 0$ be an index such that*

$\frac{\ln p_{n,k_0}}{k_0} \geq \frac{\ln p_{n,k}}{k}$ *for all* $k = 1, \ldots, n$. *Then:*

$$P(\xi^{\otimes t} = t) \leq p_{n,k_0}^{\frac{nr}{k_0}}.$$

Further, if $tk_0 = nr$ *then the maximum can be attained.*

Proof. First, we have:

$$P(\xi^{\otimes t} = t) = \prod_{j=1}^{t} p_{n,u_j} = \exp\left(\sum_{j=1}^{t} \ln p_{n,u_j}\right).$$

On the other hand, by the assumption $\frac{\ln p_{n,k_0}}{k_0} \geq \frac{\ln p_{n,k}}{k}$ $(k = 1, \ldots, n)$:

$$\sum_{j=1}^{t} \ln p_{n,u_j} = \sum_{\substack{j=1 \\ u_j \neq 0}}^{t} \frac{u_j \ln p_{n,u_j}}{u_j} \leq \sum_{j=1}^{t} \frac{u_j \ln p_{n,k_0}}{k_0} =$$

$$= \frac{\ln p_{n,k_0}}{k_0} \sum_{j=1}^{t} u_j = \frac{nr \ln p_{n,k_0}}{k_0}$$

holds. Thus:

$$P(\xi^{\otimes t} = t) \leq p_{n,k_0}^{\frac{nr}{k_0}},$$

which implies the first part of the statement. The second part can be proved by following the argument at the end of the proof of Theorem 4.4.4. □

4.4.2 Extremal accuracies by linear programming

In this section, we drop the condition (4.6), and give a compact tool based on linear programming to calculate the minimal and maximal

ensemble accuracies. We assumed earlier that the random variables η_i $(i = 1, \ldots, n)$ are independent. In our application, we consider different algorithms detecting the optic disc as classifiers. These algorithms cannot be assumed to be independent in all cases, because it can happen that the operations of the algorithms are based on very similar principles. In case of dependent algorithms, we have to decide how to measure the dependencies of the algorithms. For this aim, we can investigate the joint distribution of the outputs of the algorithms. So let:

$$c_{a_1,\ldots,a_n} = P(\eta_1 = a_1, \ldots, \eta_n = a_n),$$

where $a_i \in \{0, 1, *\}$. The star denotes any of the possible correctness values, that is, $* = 0$ or 1. The problem to determine the combination of voters achieving the best/ the worst ensemble performance is equivalent to maximize/minimize the function:

$$q(c_{a_1,\ldots,a_n}) = \sum_{k=0}^{n} \left(p_{n,k} \cdot \sum_{a_1+\ldots+a_n=k} c_{a_1,\ldots,a_n} \right) \tag{4.11}$$

under the following conditions:

$$\begin{aligned} &\sum_{a_i=1} c_{*,\ldots,*,a_i,*,\ldots,*} = p_i \quad (i = 1, \ldots, n), \\ &\sum_{a_1,\ldots,a_n} c_{a_1,\ldots,a_n} = 1, \\ &c_{a_1,\ldots,a_n} \geq 0, \end{aligned} \tag{4.12}$$

where $E\eta_i = p_i$ $(i = 1, \ldots, n)$ is the accuracy of the $i-$th classifier. Observe that this is just a linear programming problem for the variables $c_{a_1,\ldots,a_n}$, which can be solved by standard tools.

In the special case, when $(\eta_1, \ldots, \eta_n)$ are totally independent, we have:

$$c_{a_1,\ldots,a_n} = P(\eta_1 = a_1) \ldots P(\eta_n = a_n). \tag{4.13}$$

That is, the entries of the contingency table can be determined by the probabilities $p_1, \ldots, p_n$. In this case, the ensemble performance q is simply given by (4.4).

From a practical point of view, it is very important to study the improvability of an existing ensemble regarding its accuracy. To address this issue, we investigate to what extent the addition of a new classifier D_{n+1} with accuracy p_{n+1} may improve the system. For this study, we observe both the change of the system accuracy q and the interval $[q_{min}, q_{max}]$ for the minimal and maximal system accuracy.

After adding a new algorithm to the existing system, the new system accuracy depends not only on the accuracies $p_1, \ldots, p_{n+1}$, but also on the values $p_{n+1,k}$.

As an estimation for $p_{n+1,k}$, from the definition of $p_{n,k}$ we have:

$$p_{n,k} \geq p_{n+1,k}, \tag{4.14}$$

$$p_{n,k} \leq p_{n+1,k+1}. \tag{4.15}$$

In (4.14), the added vote is supposed to be false, so the probability of good decision after the extension cannot be greater than in the existing system. The estimation (4.15) describes the case of adding a correct vote to the system. To sum up (4.14) and (4.15), we get the following properties for $p_{n+1,k}$:

$$p_{n,k-1} \leq p_{n+1,k} \leq p_{n,k}. \tag{4.16}$$

Applying inequalities (4.16), the values $p_{n+1,k}$ can be estimated from the values $p_{n,k}$.

4.5 Application – optic disc detection

Now we turn to show, how our generalized model supports real-world problems in a clinical field. Progressive eye diseases can be caused

by diabetic retinopathy (DR) which can lead even to blindness. One of the first essential steps in automatic grading of the retinal images is to determine the exact location of the main anatomical features, such as the optic disc. The locations of these features play important role in making diagnosis in the clinical protocol. In this section, for the OD detection task, we start with showing how the general formulation considering the probabilities $p_{n,k}$ is restricted for this specific challenge using geometric constraints defined by anatomic rules. Then, we present the accuracy of our current ensemble, characterize it by the achieved results and discuss the possibilities of its further improvement.

4.5.1 Constraining by shape characteristics

In our application, the votes are required to fall inside a disc of diameter d_{OD} to vote together. For the calculation of the values $p_{n,k}$ for our proposed method, the k correct votes must fall inside the true OD region, however, the $n-k$ false ones can fall within discs with diameter d_{OD} anywhere else within the ROI (region of interest in the image). That is, more false regions are possible to be formed which gives the possibility to make a correct decision even if the true votes are in minority. Note that, a candidate of an algorithm is considered to be correct if its distance from the manually selected OD center is not larger than $\frac{d_{OD}}{2}$. For this configuration, see Figure 4.4.

If we assume independency among the algorithms, for our application the behavior of the values $p_{n,k}$ as a function of k for a given n is shown in Figure 4.3 for $n = 9$ and $p = 0.9$.

This function has been determined empirically by dropping random pixels on the disc in a large number of experiments.

Figure 4.3 shows that the $p_{n,k}$ increases exponentially in k for a given n. This is a consequence of the following result of [6] and [7]:

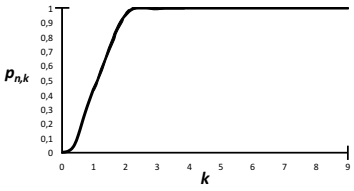


Figure 4.3: The graph of $p_{n,k}$ for a fixed $n = 9$ and $p = 0.9$ with our geometric constraint to fall within a disc of diameter d_{OD}.

Theorem 4.5.1. *Let $U_1, U_2, \ldots, U_n, \ldots$ be a sequence of independent random points in the closed Euclidean disk B of unit radius (centered at the origin). Let $H_n = max\{h(U_i, U_j) : 1 \leq i \leq j \leq n\}$ denote the h-diameter of the set $\{U_i, i = 1, \ldots, n\}$, that is the maximum pair-wise h-distance among the first n points, where h is a symmetric measurable function. Then we get that*

$$\lim_{n\to\infty} P(n^{\frac{4}{5}}(2 - H_n) > \omega) = exp(-\frac{4\omega^{\frac{5}{2}}}{5\pi}).$$

It says the probability that the diameter of a point set is not less than a given constant decreases exponentially if the number of points tends to infinity. In our case, this point set consists the pixel outputs of the algorithms for the optic disc detection. Note that, this diameter corresponds again to the diameter d_{OD} of the OD.

The ensemble accuracy for our spatially constrained system is measured empirically by the help of a set of test images. The obtained data are enclosed in Table 4.2 for different number of independent classifiers (n) for some equal individual accuracies (p).

From Table 4.2 we can see a rapid increase in the ensemble accuracy. From trivial geometric considerations, it can be also seen why

	$n = 3$	$n = 5$	$n = 7$	$n = 9$
$p = 0.6$	0.6435	0.9076	0.9654	0.9893
$p = 0.7$	0.7889	0.9631	0.9938	0.9985
$p = 0.8$	0.9029	0.9906	0.9986	0.9997
$p = 0.9$	0.9697	0.9994	1.0000	1.0000

Table 4.2: Measured ensemble accuracy under the geometric constraint.

an ensemble with few members (e.g. $n = 3$) performs bad.

Now, to describe the spatially constrained case in detail, let us assign the probability $(1-p_i)r_i$ with $r_i \in [0, 1]$ to the $i-$th independent classifier. This probability means that the $i-$th voter makes false individual decision (term $1 - p_i$) and participates in making a false ensemble decision (term r_i). For the algorithm D_i with accuracy p_i giving a false candidate having coordinates (x_i, y_i) for the OD center, we consider that the distribution of (x_i, y_i) is uniform outside the true OD region for all $i = 1, \ldots, n$. With this setup, we have:

$$r_1 = \ldots = r_n = \frac{T_0}{T - T_0}, \tag{4.17}$$

where T_0 and T are the area of the OD and the ROI, respectively, so in this case r_i is the same predetermined constant for all $i = 1, \ldots, n$. For better understanding, see also Figure 4.4.

For the interpretation of the values $p_{n,k}$ for this case, let us consider $n - k = k_1 + \ldots + k_l$, i.e. the decomposition of the number of false candidates , where all the false votes are covered by the l disjoint discs of diameter d_{OD}, and k_i is the cardinality of the false votes covered by the $i-$th disc. Without the loss of generality, $k_1 \geq \ldots \geq k_l$ may be assumed. To determine the values $p_{n,k}$, we introduce the probability

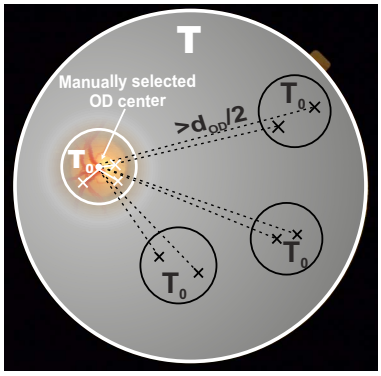


Figure 4.4: The geometric constraint applied to the candidates of the algorithms: they should fall inside a disc of a fixed diameter d_{OD} to vote together.

$P(n, k, k_1, \ldots, k_l)$ for the good decision in case of a concrete realization of the n votes:

$$P(n, k, k_1, \ldots, k_l) =$$

$$\frac{n!}{k!k_1! \ldots k_l!} p_1 \ldots p_k (1 - p_{k+1}) \ldots (1 - p_n) \left(1 - \frac{T_0}{T}\right)^{k_1} \ldots \left(1 - \frac{lT_0}{T}\right)^{k_l}.$$

Applying the geometric constraint, false decision is made only when $k_1 > k$ hence $p_{n,k} = 0$ for $k_1 > k$, while $p_{n,k} = 1$ for $k > k_1$ should hold. The case $k_1 = k$ is broken randomly. Based on these considerations and summing for the possible distribution of the $n - k$ false votes among the discs, we can calculate the corresponding values $p_{n,k}$ as

follows:

$$p_{n,k} = \sum_{k_1+\ldots+k_l=n-k, k>k_1} P(n,k,k_1,\ldots,k_l) + \\ +\frac{1}{2} \sum_{k_1+\ldots+k_l=n-k, k=k_1} P(n,k,k_1,\ldots,k_l). \tag{4.18}$$

The values $p_{n,k}$ calculated by (4.18) and the ones shown in Figure 4.3 slightly differ. The reason for this difference is that in our geometric derivation to have the closed form (4.18), we have considered only disjoint discs that completely fall inside the ROI, as well. However, these differences are minor, and both approaches have exponential trends.

4.5.2 An ensemble-based OD detector

To take advantage of the theoretical foundations of the previous sections for efficient OD detection, we have collected eight corresponding individual algorithms to create an ensemble from. Then, with a brute force approach (i.e. checking all the possible combinations) we select such an ensemble which maximizes the accuracy of the combined system. For measuring the accuracy of both the individual algorithms and the ensembles, we used the dataset MESSIDOR [42] containing 1200 digital images, where the OD centers were manually labelled by clinical experts. The images are losslessly compressed with 45° FOV and of different resolutions (1440×960, 2240×1488, and 2304×1536 pixels) that were re-scaled to 1500×1152 for normalization. For this specific resolution, we get $d_{OD} = 184$ pixels from averaging the manual annotations of clinical experts for this dataset. As a result of brute force selection, we composed an ensemble from six OD−detectors. To have an impression about the similarities and differences between these

approaches, next we give a short description for each of them. Each individual accuracy (p_i) has been measured on the dataset MESSIDOR. [19]

- *Based on pyramidal decomposition:* Lalonde et al. created an algorithm which generates a pyramid with simple Haar-based discrete wavelet transform. The pixel with the highest intensity value in the low-resolution image (4th or 5th level of decomposition) is considered as the center of the OD. $p_1 = 0.767$

- *Based on edge detection:* This method uses edge detection algorithm which is based on Rayleight-based CFAR threshold. Next, Hausdorff distance is calculated between the set of edge points and a circular template like the average OD. The pixel with the lowest distance value is selected for OD center. $p_2 = 0.958$

- *Based on entropy measurement:* Sopharak et al. proposed this method which applies a median and a CLAHE filter on the retinal image. In a neighborhood of each pixel, the entropy of intensity is calculated; the pixel with the largest entropy value is selected as the OD center. $p_3 = 0.315$

- *Based on kNN classification:* Niemeijer et al. extracted features (number, width, orientation and density of vessels and their combination), and applied a kNN classifier to decide whether a pixel belongs to the OD region. The centroid of the largest component found is considered as the OD center. $p_4 = 0.759$

- *Based on fuzzy convergence of blood vessels:* This method thins the vessel system and models each line-shape segment with a fuzzy segment. A voting map of these fuzzy segments is created and the pixel receiving the most votes is considered as OD center. $p_5 = 0.977$

- *Based on Hough transformation of vessels:* Ravishankar et al. proposed to fit lines to the thinned vessel system by Hough transformation. The intersection of these lines results in a probability map. A weighting is also applied considering the intensity values corresponding to the intersection points. The pixel having the highest probability is considered as OD center. $p_6 = 0.647$

As for the decision of the ensemble, we select the disc of the fixed diameter d_{OD} containing the largest number of algorithm candidates. Then, as the final OD center, we consider the centroid of these candidates. The final OD center is correctly found, if it falls inside the disc aligned to the manually selected OD center and having diameter d_{OD}.

4.5.3 Characterizing and comparing OD-ensemble accuracies

A natural question regarding the ensemble of the detectors is what accuracies we can expect as the best or worst based on the given individual detector accuracies. Then, we can see where the accuracy of our current ensemble falls within this interval, and can also check how it relates to a system which would contain independent ensemble members.

In our application, the values $p_{n,k}$ for calculating the above characterizing ensemble accuracies as a function of k for $n = 6$ is calculated empirically and shown in Figure 4.5. Note that, though our system naturally contains dependencies among its members, the exponential behavior of the independent ensemble (see Figure 4.3) can be observed here, as well.

Using the linear programming technique described in section 4.4.2, we have the following minimal and maximal ensemble accuracies:

$$q_{min} = 0.899, \qquad q_{max} = 1$$

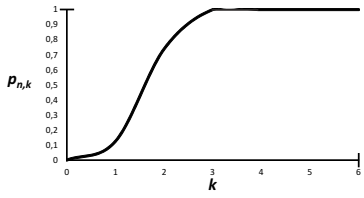


Figure 4.5: The graph of $p_{n,k}$ for a fixed $n = 6$ in our OD detector ensemble.

for the given individual accuracies, respectively.

Based on our experimental tests, the ensemble accuracy for our system has been found to be:

$$q = 0.981,$$

which is quite close to the possible maximal accuracy $q_{max} = 1$. However, if we calculate the system accuracy using (4.11) under the conditions (4.12) and with the assumption (4.13) on the independency of the detectors, we have:

$$q_{ind} = 0.998.$$

That is, an ensemble of independent algorithms with the given individual accuracies $p_1, \ldots, p_6$ would lead to nearly perfect results regarding accuracy. On the other hand, it is not surprising that our current system performs worse, since in this specific detection task it is quite challenging to find algorithms based on different (independent/diverse) principles.

4.6 Modifications on the decision rule

In this section, we modify the final decision rule of the ensemble which will result in further improvement of the system accuracy. Our generalization is based on the assignment of weights to the ensemble members (classifiers) [60].

Weighted majority voting is widely examined in the literature (see e.g. [21], [30]). For characterizing the accuracy of the weighted system to our application a corresponding theoretical model is needed. If we consider the bounding circle with the maximal weight sum, then similarly to a traditional weighted majority setup, we can make a good decision even in the case when the bad candidates have pure majority in number. In [61], we have generalized the classical majority voting to our problem. Now, just as in the traditional case, we check how weighted majority can outperform classical majority voting. In the non-weighted generalized voting system bad decision can be made only when a subset of bad candidates with larger cardinality than the number of good ones can be bounded by a circle with an appropriate radius such as in the case shown in Figure 4.6. In the weighted generalized voting system we make a wrong decision only in that case when a subset of bad candidates having larger sum of weights than the sum of weights assigned to the good ones can be bounded by a circle with an appropriate radius. In the case demonstrated in Figure 4.6. good decision is made applying the weighted generalized voting system.

These observations motivated us to work out a corresponding theoretical model, where bad votes can overcome good ones only if a further (e.g. geometrical) condition is fulfilled. This additional condition is the spatial closeness of the candidates in the above application. With this model we generalize the classical non-weighted and weighted majority voting scheme, since in the case of less good votes we may make

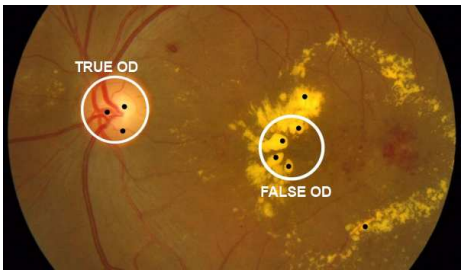


Figure 4.6: Results of the different OD detecting algorithms

a good decision. This generalized method can be applied to several problems corresponding to spatial location with additional constraints (e.g. detecting a certain pixel or region).

First, we recall the necessary procedure for finding the weights in classical majority voting (see e.g. [30]). Then, we derive how the optimal weights can be found for our generalized voting case.

4.6.1 Weighted voting system

Let us consider a set of the classifiers $\{D_1, D_2, \ldots, D_n\}$ with accuracies $(p_1, p_2, \ldots, p_n)$, respectively, for weighted voting system. Then, let $d_{i,j}$ be defined in the following way: $d_{i,j} = 1$, if the classifier D_i labels $\mathbf{x}$ in the class ω_j, and $d_{i,j} = 0$, otherwise. In case of weighted voting, the discriminant function for class ω_j is given as:

$$g_j(\mathbf{x}) = \sum_{i=1}^{n} b_i d_{i,j},$$

where the weight b_i corresponds to the classifier D_i.

Note that the following discriminant functions are equivalent for the given decision rule:

$$g_j(\mathbf{x}) = P(s|\omega_j)P(\omega_j), \quad g_j(\mathbf{x}) = \log(P(s|\omega_j)P(\omega_j)),$$

where $s = [s_1, \ldots, s_n]$ is the vector with the label output of the ensemble. Here $s_i \in \Omega$ is the label suggested for $\mathbf{x}$ by the classifier D_i and $P(\omega_j)$ is the prior probability for class ω_j.

In a weighted majority voting system, the class label ω_k is chosen for $\mathbf{x}$ if

$$g_k(\mathbf{x}) = \max_{j=1,\ldots,n} g_j(\mathbf{x}) = \sum_{i=1}^{n} b_i d_{i,k}.$$

In a weighted majority system a natural question is that how to choose the optimal weights for the classifiers. If we consider independent classifiers, then the system accuracy is maximized by assigning weights (see e.g. [30]):

$$b_i \propto \log \frac{p_i}{1-p_i} (i = 1, \ldots, n).$$

Note that, conditional independence is assumed here, that is:

$$P(s|\omega_j) = \prod_{i=1}^{n} P(s_i|\omega_j),$$

where $s = [s_1, \ldots, s_n]$ is the same as above.

The weights $b_i \propto log\frac{p_i}{1-p_i}$ do not guarantee the minimum classification error, because the prior probabilities for the classes $P(\omega_j)$ have to be taken into account, as well. More precisely, if the individual classifiers are independent, and the a priori likelihood is that each choice is equally likely to be correct, then the decision rule that maximizes the system accuracy is a weighted majority voting rule obtained by

assigning weights $b_i \propto \log \frac{p_i}{1-p_i}$.

In contrast to the classical majority voting, we equip each classifier output with different weights b_i, where $0 \leq b_i \leq 1$ $(i = 1, \ldots, n)$. It seems natural to give the classifiers with larger accuracies larger importance in making the final decision. Note that the classical majority voting scheme can be considered as a special case of the weighted voting system since in the majority rule the weight of each vote given by a classifier is constrained to be $b_i = 1$ for all $i = 1, \ldots, n$.

4.6.2 Generalized weighted voting system

We can also assign weights to the classifiers within our generalized voting scheme presented earlier. If the classifiers $(D_1, D_2, \ldots, D_n)$ with respective accuracies $(p_1, p_2, \ldots, p_n)$ and weights $(b_1, \ldots, b_n)$ are considered, then the final decision is made by choosing the maximal sum of weights, where some additional constraints (e.g. a geometrical one for OD detection) have to be fulfilled by the classifier outputs.

Similarly as in a previous section, let us consider the probability $(1 - p_i)r_i$ with $r_i \in [0, 1]$ for the $i-$th classifier that means that the $i-$th classifier makes wrong classification and participates in making a bad decision.

In our application, we choose the maximal sum of those weights of the algorithms whose outputs can be bounded by a circle with an appropriate radius. An algorithm takes part in making a bad decision if its output falling outside the optic disc meets other bad candidates. For the algorithm D_i with accuracy p_i giving a bad candidate (x_i, y_i) for the optic disc, we consider that the distribution of (x_i, y_i) is uniform

outside the optic disc for all i ($i = 1, \ldots, n$). In this case, we have:

$$r_1 = \ldots = r_n = \frac{T_0}{T - T_0}, \tag{4.19}$$

where T_0 and T are the area of the optic disc and the ROI (whole useful image domain), respectively, so r_i is a predetermined constant and same for all i ($i = 1, \ldots, n$).

The next theorem gives the answer on how to select the weights in our generalized weighted majority voting model.

Theorem 4.6.1. *If independent classifiers $(D_1, D_2, \ldots, D_n)$ are given (conditional independence is assumed), then the optimal weight b_i for the classifier D_i with accuracy p_i can be calculated as:*

$$b_i \propto \log \frac{p_i}{(1 - p_i)^2 r_i (1 - r_i)}. \tag{4.20}$$

Proof. Let $s = [s_1, \ldots, s_n]$ denote the vector with the label output of the ensemble, where $s_i \in \Omega$ is the label suggested for $\mathbf{x}$ by the classifier D_i. A Bayes-optimal set of discriminant functions based on the outputs of the n classifiers is:

$$g_j(\mathbf{x}) = \log P((\omega_j) P(s|\omega_j)), \quad (j = 1, \ldots, c).$$

From the conditional independence, for the discriminant functions

$g_j(\mathbf{x})$ we get:

$$\log P(\omega_j)P(s|\omega_j) = \log\left[P(\omega_j)\prod_{i=1}^{n} P(s_i|\omega_j)\right] =$$

$$= \log P(\omega_j) + \log\left(\prod_{i,s_i=\omega_j} P(s_i|\omega_j)\prod_{i,s_i\neq\omega_j} P(s_i|\omega_j)\right) =$$

$$= \log P(\omega_j) + \log\left(\prod_{i,s_i=\omega_j} p_i \prod_{i,s_i\neq\omega_j}(1-p_i)r_i \prod_{i,s_i\neq\omega_j}(1-p_i)(1-r_i)\right) =$$

$$= \log P(\omega_j) +$$

$$+ \log\left(\prod_{i,s_i=\omega_j}\frac{p_i(1-p_i)}{1-p_i}\prod_{i,s_i\neq\omega_j}(1-p_i)r_i \prod_{i,s_i\neq\omega_j}(1-p_i)(1-r_i)\right) =$$

$$= \log P(\omega_j) +$$

$$+ \log\left(\prod_{i,s_i=\omega_j}\frac{p_i}{1-p_i}\prod_{i,s_i\neq\omega_j}(1-p_i)r_i(1-r_i)\prod_{i=1}^{n}(1-p_i)\right) =$$

$$= \log P(\omega_j) +$$

$$+ \sum_{i,s_i=\omega_j}\log\frac{p_i}{1-p_i} + \sum_{i,s_i\neq\omega_j}\log((1-p_i)r_i(1-r_i)) + \sum_{i=1}^{n}\log(1-p_i).$$

The last term does not depend on the class label j, so we can

reduce the discriminant function to:

$$g_j(\mathbf{x}) = \log P(\omega_j) + \sum_{i,s_i=\omega_j} \log \frac{p_i}{1-p_i} + \sum_{i,s_i\neq\omega_j} \log((1-p_i)r_i(1-r_i)) =$$

$$\log P(\omega_j) + \sum_{i=1}^{n} d_{i,j} \log \frac{p_i}{1-p_i} + \sum_{i=1}^{n} (1-d_{i,j}) \log((1-p_i)r_i(1-r_i)) =$$

$$\log P(\omega_j) + \sum_{i=1}^{n} d_{i,j} \log \frac{p_i}{(1-p_i)^2 r_i(1-r_i)} + \sum_{i=1}^{n} \log((1-p_i)r_i(1-r_i)).$$

The last term of the summation is also independent from the class label j, so it can be omitted. If we consider the equations:

$$g_j(\mathbf{x}) = \log P(\omega_j) + \sum_{i=1}^{n} d_{i,j} \log \frac{p_i}{(1-p_i)^2 r_i(1-r_i)} \ ,$$

and

$$g_j(\mathbf{x}) = \sum_{i=1}^{n} b_i d_{i,j} \ ,$$

we get that the weights b_i $(i = 1, \ldots, n)$, where

$$b_i \propto \log \frac{p_i}{(1-p_i)^2 r_i(1-r_i)},$$

are supposed to maximize the system accuracy. □

Note that, similarly to classical majority voting, the weights given as optimal do not always guarantee the minimum classification error. Only if the individual classifiers are independent and the prior probabilities for the classes $P(\omega_j)$ are equal, the decision rule that maximizes the system accuracy is a weighted majority voting rule, obtained by assigning the above weights.

4.6.3 Weighted majority voting in OD detection

In our application, the output of each OD detecting algorithm is the OD center given as a single pixel with coordinates (x_0, y_0). In our ensemble-based system we have the set of class labels $\{\omega_{(x,y)}|(x,y) \in ROI\}$. For an OD detector (as a classifier) with its output (x_0, y_0), the class label $\omega_{(x_0,y_0)}$ is assigned to the detector. In other words, the classifier voted to the pixel (x_0, y_0) as OD center. The classification is considered to be correct if the output (x_0, y_0) falls inside the true optic disc on the retinal image. We can define the decision rule as the sum of the weights of the OD detecting algorithms, whose outputs can be bounded by a circle of the OD radius. Such a circle with the maximal sum of weights is accepted as the final decision for the OD.

In this application, the condition for the equal prior probabilities for the classes is fulfilled if we suppose uniform distribution of the candidates both inside and outside the optic disc.

In contrast to the non-weighted systems, less conflicting situations can be obtained when the decision is not exact because of the equal number of outputs falling inside the discs of the predetermined radius. Further improvement of this weighted system on majority voting is that there is no need for accuracy constraints $p > 0.5$ on individual algorithms to achieve larger system accuracy. It can be shown that this weighted voting rule always outperforms the classical majority rule because in case of a conflict (when the same number of votes are densified in different discs of a given radius) majority rule decides randomly between the disc candidates, while the weighted voting system can handle the conflict determining to the sum of the weights corresponding the output votes falling inside the discs.

4.6.4 Experimental results

We compare the system accuracies of the classical and the weighted majority voting for different accuracies and different weights. In our tests, we considered three different types of accuracies for the algorithms:

- $A_1 : p_1 = p_2 = \ldots = p_9 = 0.6$,
- $A_2 : p_i = 1 - 0.1i,\ i = 1, \ldots, 9$,
- $A_3 : p_1 = 0.6472, p_2 = 0.9765, p_3 = 0.3205, p_4 = 0.7593,$ $p_5 = 0.3153, p_6 = 0.2276, p_7 = 0.9582, p_8 = 0.7671, p_9 = 0.6432.$

The case A_1 is often examined in the literature with equal weights, A_2 is a theoretical example, while A_3 contains true accuracies of OD detecting algorithms measured on the Messidor test database containing 1200 retinal images.

For comparative studies, we apply the following weights b_i for the $i-$th algorithm having accuracies p_i $(i = 1, \ldots, 9)$:

- $B_1 : b_i = p_i$,
- $B_2 : b_i = log\frac{p_i}{1-p_i}$,
- $B_3 : b_i = \frac{p_i}{(1-p_i)^2 r_i(1-r_i)}$.

That is, in case B_1 each weight is equal to the accuracy of the individual algorithm (such as taken the $i-$th algorithm with accuracy p_i, then it participates in the final decision with weight $b_i = p_i$). B_2 is suggested as optimal for the classical weighted majority voting, while B_3 is the proposed assignment for our generalized weighted majority voting. In this way, we give a practical example to confirm the theoretical derivation of the optimal weights given in Section 4.6.2.

We apply OD detecting algorithms as classifiers, so we can test and compare the overall performance of the different voting systems on classifier output generated artificially. In lack of independent OD detecting algorithms providing these accuracies, we are not able to test and compare the voting systems on retinal images. We generate the classifier outputs in the following way: we consider a disc of radius R (ROI) and a disc of radius R_0 inside the ROI (optic disc), where $R = 712$ and $R_0 = 48$ pixels, respectively. We generate 9 output points with coordinates (x_i, y_i) (as outputs the D_i's), where the probability that the point (x_i, y_i) falls inside the optic disc is p_i and the distribution of (x_i, y_i) is uniform outside the optic disc. Now, the probability r_i $(i = 1, \ldots, 9)$ can be determined as:

$$r_1 = \ldots = r_n = \frac{T_0}{T - T_0} = \frac{R_0^2}{R^2 - R_0^2}. \tag{4.21}$$

In this test we compare the performance of the following voting systems: MV- majority voting, WMV- weighted majority voting, GMV- generalized majority voting, WGMV- weighted generalized majority voting. The system accuracies for the individual accuracy setups A_1, A_2, A_3 with the weight assignments (B_1, B_2, B_3) are given in Figure 4.7(a)., Figure 4.7(b)., Figure 4.7(c)., respectively.

From the tables we can see that if all weights are equal, then it naturally results in the same system accuracy as the non-weighted voting scheme, otherwise, weighted voting outperforms non-weighted voting. Our generalized non-weighted (weighted) voting system has better overall performance than the classical non-weighted (weighted) majority voting scheme.

For the OD detection application, we can test and compare our generalized non-weighted and generalized weighted voting system on a real database of retinal images, as well. The Messidor dataset considered for this aim contains 1200 retinal images. In this test,

A_1	MV	WMV	GMV	WGMV
B_1	0.7323	0.7323	0.9948	0.9996
B_2	0.7380	0.7380	0.9941	0.9991
B_3	0.7326	0.7326	0.9948	0.9989

(a) System accuracies for the set A_1

A_2	MV	WMV	GMV	WGMV
B_1	0.5012	0.8066	0.9889	0.9943
B_2	0.4965	0.9688	0.9901	0.8712
B_3	0.5009	0.7289	0.9877	0.9951

(b) System accuracies for the set A_2

A_3	MV	WMV	GMV	WGMV
B_1	0.8241	0.9526	0.9996	1.0000
B_2	0.8260	0.9926	0.9989	0.9941
B_3	0.8258	0.9481	0.9989	0.9998

(c) System accuracies for the set A_3

Figure 4.7: Overall system accuracies for the set of classifier accuracies

we assigned the optimal weights derived in Section 4.6.2 to the participating algorithms (classifiers) having individual accuracies $p_1 = 0.6472, p_2 = 0.9765, p_3 = 0.3205, p_4 = 0.7593, p_5 = 0.3153, p_6 = 0.2276, p_7 = 0.9582, p_8 = 0.7671, p_9 = 0.6432$ (as given in case $A3$). However, note that we have no information about the dependencies among these algorithms. Despite the unknown dependencies of the algorithms, we found that weighted majority voting with its system accuracy 0.98 outperformed classical majority voting (system accuracy 0.974), and also all the individual accuracies.

4.7 Discussion and conclusions

In this chapter, we have introduced a new model that enables the investigation of majority voting systems in the spatial domain. We have considered independent/dependent ensembles composed by classifiers having not necessarily the same individual accuracies. We have described how a constraint may raise from shape characteristics, and presented an ensemble-based system for optic disc detection in retinal images, where the object has a circular anatomical geometry. The general theory of ensemble-based systems describes several voting methodologies. However, for spatial voting, corresponding models have not been presented yet, and their adaptation is rather challenging to this domain. For instance, the extension of the approach proposed in this chapter is currently under study for weighted spatial majority voting, but for several cases (e.g. dependent voters) it is far from being trivial.

Our detailed experimental studies have been performed on the image dataset MESSIDOR [42]. However, it is well-known in the community working with retinal images that we can expect high variance among retinal image databases, so tests on different datasets are recommended. Thus, to validate more its efficiency, we have tested the proposed ensemble-based approach on a database containing 327 im-

ages provided by the Moorfields Eye Hospital, London from a real mass screening process. Following the presented methodology, we have composed ensembles from the individual OD detector algorithms, and applied the spatially constrained decision rule. The highest accuracy $q = 0.921$ has been found for the ensemble containing the four members having individual accuracies $p_1 = 0.798$, $p_3 = 0.150$, $p_4 = 0.801$, $p_5 = 0.835$, respectively (for the remaining three algorithms we have measured $p_2 = 0.780$, $p_6 = 0.342$, and $p_7 = 0.297$, respectively). Similarly to MESSIDOR, the ensemble performed better than any of its members for the Moorfields dataset, as well. Moreover, we can observe that the individual accuracies have been varied more among the different datasets than that of the ensemble. This observation suggests that we can expect a more stable and calculable behavior if we work with ensembles.

Our approach can be extended to other detection problems with keeping in mind that the presented results are suitable to handle such shapes that can be described by set diameter. To demonstrate the efficiency of our method, we considered another detection problem: the localization of the macula, which is the center of the sharp vision in the retina and appears as a dark, disc-like object of diameter approximately 6mm. That is, we have a very similar scenario to that of the OD detection problem. We have set up an ensemble of five macula detectors [19] having individual accuracies 0.583, 0.870, 0.714, 0.624, 0.962, respectively. By applying the proposed spatially constrained decision scheme, we have found 0.968 for the accuracy of the ensemble for the dataset MESSIDOR. From both the individual and ensemble accuracy values, we can see that the consideration of the macula detector algorithms are less reliable, however, in practical application we can also take advantage of the detection of the OD, since the OD has a fixed distance and direction from the macula. Nevertheless, from these results we can see that our proposed ensemble-based approach has leaded to improvement in this field, as well.

This generalized voting system (when some additional geometrical constraints have to be fulfilled) can be applied in that case when weights are assigned to the classifiers, as well. In our specific application better overall system accuracy is achieved than in case of individual algorithm and weighted voting outperformed the non-weighted one. Same results are expected for all image processing problems where the algorithms vote with a single pixel or range as output. In this application adding a new algorithm to the system seems to be very effective, as well.

It will be a future research direction to characterize how the accuracy can be raised by removing/adding algorithms from/to the existing system in consideration to individual accuracies and dependencies. The performance of combination methods that are variations of the majority vote have been studied in [35]. A Bayesian formulation and a weighted majority vote (with weights obtained through a genetic algorithm) are implemented, and the combined performances of 7 classifiers on a large set of handwritten numerals are analyzed. The genetic algorithm is capable of identifying dependencies among classifiers, by assigning lower weights to those that are less effective in influencing the group decision in the optimal direction. This is a very useful feature, since this kind of knowledge is not available when classifiers are designed and implemented. In general, it is a non-trivial problem to decide whether classifiers are independent, since there can exist many possible overlaps between feature sets and classification methods. Even after a recognizer has been trained and tested, and its individual performance is known, it is still difficult to know how significant a role it would assume as a member of a group. The genetic algorithm can help to obtain such information, which can lead to simpler and more efficient multiple classifier systems.

In the future, the pattern of success and failure can be examined in generalized weighted voting system, as well. This will give useful information in clinical systems since they characterize the expected

value of the system error and the boundary of the system accuracy.

Chapter 5

Összefoglaló

Egy rövid bevezető után a 2. fejezetben számos, a Lie–csoport elméletéből ismert fogalmat és eredményt vizsgálunk meg a loopok szempontjából, ahol elhagyjuk a szorzás asszociatív tulajdonságát.[58], [59] A Lie–csoport egy sima sokaság, amelyen olyan csoportstruktúra adható meg, ahol a szorzásművelet és annak inverze is sima leképezések. A Lie–csoporthoz hasonlóan, egy loopot is definiálhatunk algebrai és differenciálgeometriai struktúraként egyaránt. Az n-loop definíciójában egy n-változós szorzásműveletnek hasonló tulajdonságokat kell kielégítenie, mint a loop definíciójában szereplő bináris műveletnek:

Definíció. *Legyen H egy nemüres halmaz, melyen értelmezve van egy $m\colon H^n \to H$ n-változós szorzásművelet, és adott egy $e \in H$ H-beli elem. Ekkor (H, e, m) n-**loop** az e egységelemmel, ha teljesíti a következő feltételeket:*

1. $m(\overset{(1)}{e}, \ldots, \overset{(i-1)}{e}, \overset{(i)}{a}, \overset{(i+1)}{e}, \ldots, \overset{(n)}{e}) = a$ *minden* $a \in H$ *és* $1 \leq i \leq n$ *esetén, ahol* $\overset{(i)}{x}$ *jelöli azt, hogy az* i*. argumentum értéke* x*,*

2. *az* $m(a_1, a_2, \ldots, a_{i-1}, x, a_{i+1}, \ldots, a_n) = b$ *egyenlet egyértelműen megoldható minden* $a_i \in H, 1 \leq i \leq n$, $b \in H$ *esetén.*

Ezután definiáljuk a kanonikus koordináta-rendszer fogalmát.

Definíció. *Legyen* $\mathcal{H} = (H, e, m, \delta_1, \ldots, \delta_n)$ *egy* $\mathcal{C}^k-$*differenciálható lokális n-loop. Egy* $\varphi\colon U \to \mathbb{R}^q$ $\mathcal{C}^k-$ *osztályú koordinátatérképet, mely az* $e \in H$ *egységelem* $U \subset H$ *nyílt környezetét képezi le az* $\mathbb{R}^q$ *koordinátatérre, a* $\mathcal{H}$ ***kanonikus koordináta-rendszer****ének nevezzük, ha:*

1. $\varphi(e) = 0$,

2. *az* $m\colon H^n \to H$ *szorzás* $M\colon \varphi(U) \times \cdots \times \varphi(U) \to \mathbb{R}^q$ *kooordináta-függvénye*

$$M = \varphi \circ m \circ (\varphi^{-1} \times \cdots \times \varphi^{-1})$$

rendelkezik az alábbi tulajdonsággal:

$$M(x, x, \ldots, x) = n\, x$$

minden $x \in \varphi(U)$ *esetén.*

Tétel. *Legyen* $\mathcal{H} = (H, e, m, \delta_1, \ldots, \delta_n)$ *egy* $\mathcal{C}^k-$*differenciálható lokális* $n-$*loop, ahol* $k \geq 2$. *Ekkor* $\mathcal{H}$*-nak mindig létezik kanonikus koordináta-rendszere.*
Ha (U, φ) *kanonikus koordináta-rendszere* $\mathcal{H}$*-nak, akkor* $(U, \tau \circ \varphi)$ *is kanonikus koordináta-rendszere* $\mathcal{H}$*-nak minden* $\tau : \mathbb{R}^q \to \mathbb{R}^q$ *lineáris leképezés esetén.*

Ha $\varphi\colon U \to \mathbb{R}^q$ *és* $\psi\colon U \to \mathbb{R}^q$ *a* $\mathcal{H}$ *kanonikus koordináta-rendszereinek koordináta-leképezései, melyek ugyanazon* U *környezeten vannak definiálva, akkor* $\varphi \circ \psi^{-1}$ *egy* $\mathbb{R}^q \to \mathbb{R}^q$ *lineáris leképezés leszűkítése.*

A $\mathcal{C}^k-$differenciálható lokális n-loopok esetén többféle természetes lehetőség is adódik a $W \to H$ $(0 \in W \subset T_eH)$ exponenciális leképezés definiálására. Az egyik lehetséges definíció a Lie-csoport elméletében gyakran használt konstrukcióval analóg, nevezetesen az *exp* leképezést az egységelembeli érintővektorok $i.$ eltolásai által létrehozott vektormezők integrálgörbéi határozzák meg.

Definíció. *Legyen* $\gamma_v^i(t)$ *a* $\dot{\gamma}_v^i(t) = (\lambda^i_{\gamma(t)})_* v$ *differenciálegyenlet integrálgörbéje, ahol* $\gamma_v^i(0) = e$ *,* $\dot{\gamma}_v^i(0) = v$ *és* λ_x^i *jelöli az* x *elemmel történő* $i.$ *eltolást. Ekkor az* $exp^{(i)} : T_eH \to H$ *leképezést:* $exp^{(i)}(v) = \gamma_v^i(1)$*, az* i***. exponenciális leképezés****nek nevezzük.*

Az exponenciális leképezés egy természetesen adódó, másik lehetséges definícióját az előbbiekben tárgyalt kanonikus koordináta-rendszer konstrukciója szolgáltathatja.

Tétel. *Legyen* $\mathcal{H} = (H, e, m, \delta_1, \ldots, \delta_n)$ *egy* $\mathcal{C}^k-$*differenciálható lokális* $n-$*loop, ahol* $k \geq 2$*. Ekkor egyértelműen létezik olyan* $\exp\colon W \to H$ *lokális* $\mathcal{C}^k-$*diffeomorfizmus, ahol* W *a* $0 \in T_eH$ *egy környezete, és teljesülnek rá az alábbi tulajdonságok:*

(i) $\exp(0) = e,$

(ii) $\exp(n\,x) = m(\exp(x), \ldots, \exp(x)),$

(iii) $\exp_* |_0 = id_{T_eH}.$

Tétel. *Legyenek* $\mathcal{H} = (H, e, m, \delta_1, \ldots, \delta_n)$, $\mathcal{H}' = (H', e', m', \delta'_1, \ldots, \delta'_n)$ $\mathcal{C}^k-$*differenciálható lokális* n*-loopok. Legyenek továbbá* $\exp\colon W \to H$, $\exp'\colon W' \to H'$ *rendre ezek exponenciális leképezései, ahol* $W \subset T_eH$ *és* $W' \subset T_{e'}H'$. *Ha* $\alpha\colon \mathcal{H} \to \mathcal{H}'$ *egy folytonos lokális homomorfizmus, akkor az* $\exp'^{-1} \circ \ \alpha \circ \exp\colon W \to T_{e'}H'$ *leképezés lokálisan lineáris.*

A 3. fejezetben megmutatjuk, hogy a differenciálgeometria eszközei milyen hasznosak lehetnek akár az orvosi képalkotás terén is. Anizotropikus diffúzió esetén ugyanis a skalár diffúziós együttható már nem elegendő a diffúzió tulajdonságainak leírására. Ebben az esetben a diffúziót egy másodrendű, szimmetrikus tenzorral lehet jellemezni, melyet diffúziós tenzornak nevezünk:

$$D = \begin{pmatrix} D_{xx} & D_{xy} & D_{xz} \\ D_{yx} & D_{yy} & D_{yz} \\ D_{zx} & D_{zy} & D_{zz} \end{pmatrix}.$$

A diffúziós tenzor hat független elemét diffúziós súlyozott képek segítségével tudjuk megbecsülni.

Ha N irány mentén végzünk diffúziós súlyozott méréseket, a következő mátrixegyenlet írható fel:

$$B\vec{d}^{\,T} = \vec{A}^{\,T},$$

ahol

$$\vec{A} = \begin{pmatrix} \ln \frac{S_1}{S_0} & \ln \frac{S_2}{S_0} & \ldots & \ln \frac{S_N}{S_0} \end{pmatrix}$$

a megfelelő logaritmikus jelarányok vektora,

$$B = \begin{pmatrix} \vec{b_1} \\ \vec{b_2} \\ \vdots \\ \vec{b_N} \end{pmatrix}$$

a diffúziós gradiens paramétereit jellemzi.

Ez a diffúziós tenzormodell a diffúziós irányra vonatkozó információk teljes leírását nyújtja.

A diffúziós elmozdulást egy ellipszoid segítségével jellemezhetjük, ahol a főtengelyek hosszát a diffúziós tenzor sajátértékei, míg az irányukat a diffúziós tenzor sajátvektorai írják le.

Az agy rostszálainak rekonstruálása és képi megjelenítése volt a célunk.[62], [63] A modellezés során általában csak három ortogonális síkkal történő metszeteket mutatnak, mely tetszőleges számú és állású vágósíkkal is elvégezhető.

A 4. fejezetben általánosítjuk a klasszikus többségi szavazómodellt a $0 \leq p_{n,k} \leq 1$ valószínűségi értékek bevezetésével, melyek a jó döntés valószínűségét írják le, amennyiben az n (összes) szavazat között pontosan k helyes található. Ezt az általánosítást azok a detektálási problémák motiválták, ahol az összetett rendszer tagjai képfeldolgozó algoritmusok, melyek szavazatként a kép egy pixelét adják vissza. Ebben az esetben a $p_{n,k}$ valószínűségi értékeket geometriai feltételek határozzák meg.[19], [61]

Az általánosított modellünkben a D_i osztályozókat, melyek p_i pontosságúak, η_i Bernoulli-eloszlású valószínűségi változókként tekintjük, azaz:

$$P(\eta_i = 1) = p_i, \quad P(\eta_i = 0) = 1 - p_i \quad (i = 1, \ldots, n).$$

Az $\eta_i = 1$ a D_i általi helyes osztályozást jelöli. Ekkor a D_i osztályozó pontosságát az η_i valószínűségi változó várható értéke adja meg, ahol $E\eta_i = p_i$ $(i = 1, \ldots, n)$.

A $p_{n,k}$ $(k = 0, 1, \ldots, n)$ adott valós számok, ahol:

$$0 \leq p_{j0} \leq \cdots \leq p_{jj} \leq 1 \quad (j = 1, \ldots, n),$$

és a ξ valószínűségi változót a következőképpen definiáljuk:

$$P(\xi = 1) = p_{n,k} \quad \text{és} \quad P(\xi = 0) = 1 - p_{n,k},$$

ahol $k = |\{i \ : \ \eta_i = 1\}|$. Ekkor a $D_1, \ldots, D_n$ osztályozókhoz tartozó módosított többségi szavazást a ξ valószínűségi változó reprezentálja: ha az n osztályozó közül k esetén valósult meg helyes osztályozás, akkor ezen többségi szavazást alkalmazva $p_{n,k}$ valószínűséggel hozunk helyes döntést (azaz $\xi = 1$).

A célunk, hogy az általánosított többségi szavazómodell minden egyedi osztályozónál pontosabb legyen. Ehhez a $q \geq p$ feltételnek kell teljesülnie.

Ezután megvizsgáltuk, hogy az osztályozók függősége mennyire befolyásolja az összetett rendszer pontosságát. Ezért általánosítottuk azokat a klasszikus többségi szavazással kapcsolatos fogalmakat, melyek az összetett rendszer pontosságának maximumát (*pattern of success*), illetve minimumát (*pattern of failure*) jellemzik.

Az előzőekben használt megkötések elhagyhatóak, ha egy olyan átfogó, kompakt eszközt használunk a maximális, illetve minimális pontosságú összetett rendszer feltérképezésére, mely ezt egy lineáris programozási feladat megoldásaként valósítja meg. Az osztályozók függőségének jellemzésére a valószínűségi változók együttes eloszlását tekintjük:

$$c_{a_1,\ldots,a_n} = P(\eta_1 = a_1, \ldots, \eta_n = a_n),$$

ahol $a_i \in \{0, 1, *\}$ és $* = 0$ vagy 1.

Ahhoz, hogy meghatározzuk a legjobban/legrosszabbul teljesítő összetett rendszert, illetve annak pontosságát (q_{max}/q_{min}), maximalizálni/minimalizálni kell a:

$$q(c_{a_1,\ldots,a_n}) = \sum_{k=0}^{n} \left(p_{n,k} \sum_{a_1+\ldots+a_n=k} c_{a_1,\ldots,a_n} \right)$$

függvényt az alábbi feltételek mellett:

$$\sum_{a_i=1} c_{*,\ldots,*,a_i,*,\ldots,*} = p_i \quad (i = 1, \ldots, n),$$

$$\sum_{a_1,\ldots,a_n} c_{a_1,\ldots,a_n} = 1,$$

$$c_{a_1,\ldots,a_n} \geq 0,$$

ahol $E\eta_i = p_i$ $(i = 1, \ldots, n)$ az $i.$ osztályozó pontossága.

Ezután az OD detektálási probléma kapcsán, bemutatjuk, hogy a $p_{n,k}$ valószínűségi értékeket hogyan befolyásolják a geometriai feltételek. Ebben az esetben egyben formalizálni is tudjuk a $p_{n,k}$ valószínűségi értékeket.

Ezt követően módosítjuk a végső döntés meghozatalára vonatkozó szabályt, mely további javulást eredményez az összetett rendszer pontosságában. Az általánosítás arra épül, hogy súlyokat rendelünk az osztályozókhoz.[60]

Végül választ adunk arra a kérdésre, hogy független osztályozókat tekintve, az általánosított súlyozott többségi szavazás esetén hogyan lehet a súlyokat optimálisan megválasztani (a maximális rendszerpontosság elérése érdekében).

Tétel. *Független $(D_1, D_2, \ldots, D_n)$ osztályozók esetén, a p_i pontosságú D_i osztályozóhoz az optimális b_i súlyra teljesül, hogy:*

$$b_i \propto \log \frac{p_i}{(1-p_i)^2 r_i (1-r_i)}$$

ahol a $(1-p_i)r_i$ $(r_i \in [0,1])$ érték annak a valószínűségét jellemzi, hogy az $i.$ osztályozó helytelen osztályozás esetén részt vesz a végső helytelen döntés meghozatalában.

Chapter 6

Summary

After a short introduction, in the second chapter, several concepts and results from the theory of Lie group are investigated for loops, i.e. for non-associative multiplication.[58],[59]

A Lie group is smooth manifold which also carries a group structure whose multiplication and its inverse operation are smooth as maps of manifolds. Similarly to Lie group, we can consider a loop as an algebraic and as a differential geometric notion, as well.

In the definition of n-loop, an n-ary multiplication has to satisfy similar properties as in the binary case:

Definition. *Let H be a non-empty set with the n-ary multiplication $m\colon H^n \to H$, let $e \in H$ be a given element. Then (H, e, m) is called n-**loop** with the unit element e if*

1. $m(\overset{(1)}{e}, \ldots, \overset{(i-1)}{e}, \overset{(i)}{a}, \overset{(i+1)}{e}, \ldots, \overset{(n)}{e}) = a$ *for all* $a \in H$, $1 \leq i \leq n$, *where* $\overset{(i)}{x}$ *means that the i-th argument has the value x,*

2. *the equation* $m(a_1, a_2, \ldots, a_{i-1}, x, a_{i+1}, \ldots, a_n) = b$ *is uniquely solvable for all* $a_i \in H (1 \leq i \leq n)$, $b \in H$.

In the definition of a local $\mathcal{C}^k-$differentiable $n-$loop, the multiplication and the $i-$th divisions are defined only in a neighborhood of the unit element and the implicit function theorem provides the $\mathcal{C}^k-$differentiability of the $i-$th divisions locally around the unit element.

Definition. *If* H *is a differentiable manifold of class* $\mathcal{C}^k$, $e \in H$ *is a given element and* $m\colon H^n \to H$, $\delta_i\colon H^n \to H$ *are differentiable maps of class* $\mathcal{C}^k$ $(i = 1, \ldots, n)$, *which are defined in a neighbourhood of* $e \in H$, *then* $\mathcal{H} = (H, e, m, \delta_1, \ldots, \delta_n)$ *is called* $\mathcal{C}^k-$***differentiable local*** $n-$***loop*** *with unit element* e, *if the multiplication* m *and the* $i-$*th divisions* δ_i $(i = 1, \ldots, n)$ *satisfy the following identities:*

1. $m(\overset{(1)}{e}, \ldots, \overset{(i-1)}{e}, \overset{(i)}{a}, \overset{(i+1)}{e}, \ldots, \overset{(n)}{e}) = a$ *for all* $a \in H$, $1 \leq i \leq n$,

2. $m(a_1, \ldots, a_{i-1}, \delta_i(b; a_1, \ldots, a_{i-1}, a_{i+1}, \ldots, a_n), a_{i+1}, \ldots, a_n) = b$ *for all* $a_i \in H$, $1 \leq i \leq n, b \in H$,

3. $\delta_i(m(a_1, \ldots, a_n); a_1, \ldots, a_{i-1}, a_{i+1}, \ldots, a_n) = a_i$ *for all* $a_i \in H$, $1 \leq i \leq n$,

in a neighbourhood of $e \in H$.

Then, we define the notion of canonical coordinate system.

Definition. *Let* $\mathcal{H} = (H, e, m, \delta_1, \ldots, \delta_n)$ *be a* $\mathcal{C}^k-$*differentiable local* $n-$*loop. A coordinate map* $\varphi\colon U \to \mathbb{R}^q$ *of class* $\mathcal{C}^k$ *of the open*

neighbourhood $U \subset H$ of $e \in H$ into the coordinate space $\mathbb{R}^q$ is called ***canonical coordinate system*** *of $\mathcal{H}$ if $\varphi(e) = 0$ and the coordinate function $M\colon \varphi(U) \times \cdots \times \varphi(U) \to \mathbb{R}^q$ of the multiplication map $m\colon H^n \to H$*

$$M = \varphi \circ m \circ (\varphi^{-1} \times \cdots \times \varphi^{-1})$$

satisfies

$$M(x, x, \ldots, x) = n\, x$$

for all $x \in \varphi(U)$.

Theorem. *Let $\mathcal{H} = (H, e, m, \delta_1, \ldots, \delta_n)$ be a $\mathcal{C}^k-$differentiable local $n-$loop with $k \geq 2$. Then there exists a canonical coordinate system for $\mathcal{H}$.*

If (U, φ) is a canonical coordinate system of $\mathcal{H}$ then for any linear map $\tau : \mathbb{R}^q \to \mathbb{R}^q$ the pair $(U, \tau \circ \varphi)$ is a canonical coordinate system of $\mathcal{H}$, as well.

If $\varphi\colon U \to \mathbb{R}^q$ and $\psi\colon U \to \mathbb{R}^q$ are the coordinate maps of canonical coordinate systems of $\mathcal{H}$ defined on the same neighbourhood U then $\varphi \circ \psi^{-1}$ is the restriction of a linear map $\mathbb{R}^q \to \mathbb{R}^q$.

There are several natural possibilities for the definition of the exponential map $W \to H$ with $0 \in W \subset T_eH$ of $\mathcal{C}^k-$differentiable local n-loops. One of them is analogous to the usual construction in Lie group theory, namely the map *exp* could be determined by the integral curves of vector fields defined by the i-th translations of tangent vectors at the unit element of the n-loop.

Definition. *Let $\gamma_v^i(t)$ be the integral curve of the differential equation*

$$\dot{\gamma}_v^i(t) = (\lambda^i_{\gamma(t)})_* v, \text{ where } \gamma_v^i(0) = e, \ \dot{\gamma}_v^i(0) = v$$

*and λ_x^i denotes the $i-th$ translation with x. Then $exp^{(i)} : T_eH \to H$, where $exp^{(i)}(v) = \gamma_v^i(1)$, is called $i-$**th exponential map**.*

An alternative natural possibility for the definition of the exponential map is given by using the construction of canonical coordinate systems.

Theorem. *Let $\mathcal{H} = (H, e, m, \delta_1, \ldots, \delta_n)$ be a $\mathcal{C}^k-$differentiable local $n-$loop with $k \geq 2$. There exists a unique local $\mathcal{C}^k-$diffeomorphism $\exp\colon W \to H$, where W is a neighbourhood of $0 \in T_eH$, such that the following conditions hold:*

(i) $\exp(0) = e$,

(ii) $\exp(n\,x) = m(\exp(x), \ldots, \exp(x))$,

(iii) $\exp_* |_0 = id_{T_eH}$.

Theorem. *Let $\mathcal{H} = (H, e, m, \delta_1, \ldots, \delta_n)$, $\mathcal{H}' = (H', e', m', \delta'_1, \ldots, \delta'_n)$ be $\mathcal{C}^k-$differentiable local $n-$loops. Let $\exp\colon W \to H$, $\exp'\colon W' \to H'$, where $W \subset T_eH$, $W' \subset T_{e'}H'$, be the corresponding exponential maps, respectively. If $\alpha\colon \mathcal{H} \to \mathcal{H}'$ is a continuous local homomorphism then the composed map $\exp'^{-1} \circ \alpha \circ \exp\colon W \to T_{e'}H'$ is locally linear.*

In the third chapter, it has been shown how useful the tools of differential geometry can be also in biomedical imaging. When the

diffusion is anisotropic, a scalar diffusion measure is insufficient for describing diffusion properties. In this case, the diffusion can be described by a second−order diagonally symmetric tensor, called the diffusion tensor:

$$D = \begin{pmatrix} D_{xx} & D_{xy} & D_{xz} \\ D_{yx} & D_{yy} & D_{yz} \\ D_{zx} & D_{zy} & D_{zz} \end{pmatrix}.$$

The six independent elements of the diffusion tensor can be estimated from a series of diffusion-weighted images. When diffusion weighted measurements are performed along N directions, the following matrix equation can be constructed:

$$B\vec{d}^{\,T} = \vec{A}^{\,T},$$

where

$$\vec{A} = \begin{pmatrix} \ln\frac{S_1}{S_0} & \ln\frac{S_2}{S_0} & \dots & \ln\frac{S_N}{S_0} \end{pmatrix}$$

is the vector of the corresponding logarithmic signal ratios and

$$B = \begin{pmatrix} \vec{b_1} \\ \vec{b_2} \\ \vdots \\ \vec{b_N} \end{pmatrix}$$

includes the influences of all the encoding gradients.

This tensor model of diffusion is able to get full description of the directional diffusion information. The diffusion displacement profile may be represented as an ellipsoid with the length of principal axes described by the eigenvalues of the diffusion tensor and the directions given by eigenvectors of the diffusion tensor.

Our aim is to reconstruct the fiber tracts of the human brain from measurements of fiber orientation and visualize them on the image of

the brain.[62], [63] Generally, the surface model clipped by orthogonal sections is shown. We are capable to visualize the surface model clipped by (even more than the usual three) planes having arbitrary directions.

In the fourth chapter, we propose the generalization of the classical majority voting model by introducing values $0 \leq p_{n,k} \leq 1$ describing the probability of making a good decision, when we have exactly k good votes from the n voters. This generalization is motivated by object detection problems, where the members of the ensemble are image processing algorithms giving their votes as pixels in the image domain. In this scenario, the terms $p_{n,k}$ can be specialized by a geometric constraint.[19], [61]

In our model, we consider a classifier D_i with accuracy p_i as a random variable η_i of Bernoulli distribution, i.e.:

$$P(\eta_i = 1) = p_i, \quad P(\eta_i = 0) = 1 - p_i \quad (i = 1, \ldots, n).$$

Here $\eta_i = 1$ means correct classification by D_i. In particular, the accuracy of D_i is just the expected value of η_i, that is, $E\eta_i = p_i$ $(i = 1, \ldots, n)$.

Let $p_{n,k}$ $(k = 0, 1, \ldots, n)$ be given real numbers with:

$$0 \leq p_{j0} \leq \cdots \leq p_{jj} \leq 1 \quad (j = 1, \ldots, n),$$

and let the random variable ξ be such that:

$$P(\xi = 1) = p_{n,k} \quad \text{and} \quad P(\xi = 0) = 1 - p_{n,k},$$

where $k = |\{i \ : \ \eta_i = 1\}|$. That is, ξ represents the modified majority voting of the classifiers $D_1, \ldots, D_n$: if k out of the n classifiers give a correct vote, then we make a good decision (i.e. we have $\xi = 1$) with probability $p_{n,k}$.

In order to have this generalized majority voting model be more accurate than all the individual decisions, we have to guarantee that $q \geq p$.

Next, we investigate how dependencies among the voters influence the accuracy of the ensemble. For this purpose, we generalize some concepts that were introduced for classical majority voting to measure the extremal behavior (minimal and maximal accuracies) of an ensemble. First we consider *pattern of success* and *pattern of failure*, i.e. such realizations of the votes in a series of experiments that lead to the possible highest and lowest accuracy of the ensemble, respectively.

Then we drop all the previous conditions and give a compact tool (working without any technical restrictions) based on linear programming to calculate the minimal and maximal ensemble accuracies. In case of dependent classifiers, we can investigate the joint distribution of the random variables to measure the dependencies of the classifiers:

$$c_{a_1,\ldots,a_n} = P(\eta_1 = a_1, \ldots, \eta_n = a_n),$$

where $a_i \in \{0, 1, *\}$. The problem to determine the combination of voters achieving the best/ the worst ensemble performance (q_{max}/q_{min}) is equivalent to maximize/minimize the function:

$$q(c_{a_1,\ldots,a_n}) = \sum_{k=0}^{n} \left(p_{n,k} \sum_{a_1+\ldots+a_n=k} c_{a_1,\ldots,a_n} \right)$$

under the following conditions:

$$\sum_{a_i=1} c_{*,\ldots,*,a_i,*,\ldots,*} = p_i \quad (i = 1, \ldots, n),$$

$$\sum_{a_1,\ldots,a_n} c_{a_1,\ldots,a_n} = 1,$$

$$c_{a_1,\ldots,a_n} \geq 0,$$

where $E\eta_i = p_i$ $(i = 1, \ldots, n)$ is the accuracy of the i-th classifier.

For the OD detection task, we start with showing how the general formulation considering the probabilities $p_{n,k}$ is restricted for this specific challenge using geometric constraints defined by anatomic rules. We can calculate the corresponding values $p_{n,k}$ in this case, as well.

At last, we modify the final decision rule of the ensemble which will result in further improvement of the system accuracy. Our generalization is based on the assignment of weights to the ensemble members (classifiers).[60]

We give the answer on how to select optimally the weights (to achieve the best system accuracy) for independent classifiers in our generalized weighted majority voting model.

Theorem. *If independent classifiers $(D_1, D_2, \ldots, D_n)$ are given (conditional independence is assumed), then the optimal weight b_i for the classifier D_i with accuracy p_i can be calculated as:*

$$b_i \propto \log \frac{p_i}{(1-p_i)^2 r_i (1-r_i)},$$

where the probability $(1-p_i)r_i$ with $r_i \in [0,1]$ is considered for the $i-th$ classifier meaning that the $i-th$ classifier makes wrong classification and participates in making a bad decision.

Chapter 7

List of talks

1. *Diversity measures for majority voting in the spatial domain*, 8th International Conference on Hybrid Artificial Intelligence Systems, Salamanca, September 11–13, 2013.

2. *Generalized weighted majority voting with an application to algorithms having spatial output*, 7th International Conference on Hybrid Artificial Intelligence Systems, Salamanca, March 28–30, 2012.

3. *Az orvosi képfeldolgozás és a geometria, avagy mit tudhat egy körlap a diabéteszről*, Budapesti Műszaki és Gazdaságtudományi Egyetem Geometria Tanszék Szemináriuma, Budapest, December 11, 2012.

4. *A generalization of majority voting scheme for medical image detectors*, 6th International Conference on Hybrid Artificial Intelligence Systems, Wroclaw, May 23–25, 2011.

5. *Általánosított többségi szavazás alkalmazása retinaképek elemzésében*, 14. Gyires Béla Informatikai Nap, Debrecen, December 16, 2011.

6. *Bol reflection*, XIth Conference Geometry and Graphics, Ustron, June 28–30, 2010.

7. *Detecting digital intersections using line approximation*, 8th International Conference on Applied Informatics, Eger, January 27–30, 2010.

8. *Thickness-based binary morphological improvement of distorted digital line intersections*, 5th Hungarian Conference on Computer Graphics and Geometry, Budapest, January 26–27, 2010.

9. *New DTI visualization methods*, 13th International Conference on Geometry and Graphics, Dresden, August 4–8, 2008.

10. *New visualization methods in radiology*, 12th Colloquium of Croatian Society for Geometry and Graphics, Vukovar, September 16–20, 2007.

11. *On Bol closure condition*, Loops 07, Prague, August 19–25, 2007.

12. *On $n-$loops and $(n+1)-$webs*, 26th Conference on Geometry and Computer Graphics, Nové Mesto na Morave, September 11–15, 2006.

13. *Exponential maps and canonical coordinate-systems of differentiable $n-$loops*, 9th International Conference on Differential Geometry and Its Applications, Prague, August 30–September 3, 2004.

14. *Exponential maps of differentiable $n-$loops*, 3rd Workshop on Differential Geometry, Olomouc, October 16–18, 2003.

15. *Exponential maps of $n-$ary differentiable loops*, Loops 03, Prague, August 11–17, 2003.

16. *On the differentiability of $n-$loops*, 4th Conference on Geometry and Topology of Manifolds, Krynica, April 29–May 4, 2002.

17. *Experiences about the mathematical and geometrical courses for the students on Faculty of Informatics*, 5th International Conference on Applied Informatics, Eger, January 28–February 3, 2001.

Chapter 8

List of publications

1. Hajdu, A., Hajdu, L., Jónás, Á., Kovács, L., Tomán, H.: Generalizing the Majority Voting Scheme to Spatially Constrained Voting, IEEE Trans. on Image Processing **22(11)** (2013), 4182–4194, IF=3.042.

2. Hajdu, A., Hajdu, L., Kovács, L., Tomán, H.: Diversity Measures for Majority Voting in the Spatial Domain, Lecture Notes in Artificial Intelligence **8073** (2013), 314–323.

3. Tomán, H., Kovács, L., Jónás, A., Hajdu, L., Hajdu, A.: Generalized Weighted Majority Voting with an Application to Algorithms Having Spatial Output, Lecture Notes in Artificial Intelligence **7209** (2012), 56–67.

   független hivatkozás:

   Wang, H., Yang, Y., Wang, H., Chen, D.: Soft-Voting Clustering Ensemble, Lecture Notes in Computer Science **7872** (2013), 307–318.

4. Tomán, H., Kovács, L., Jónás, A., Hajdu, L., Hajdu, A.: A Generalization of Majority Voting Scheme for Medical Image Detectors, Lecture Notes in Artificial Intelligence **6679/2** (2011), 189–196.

5. Tomán, H.: Canonical Coordinate Systems and Exponential Maps of $n-$loop, Note di Matematica **24/2** (2005), 1–7.
MR2223649 (2007g:20063)
Zbl 1113.22003

6. Tomán, H., Tornai, R., Zichar, M.: Complex Fiber Visualization, Ann. Math. Inform. **34** (2007), 103–109.
Zbl 1135.68604

   független hivatkozások:

   Wang, Y., Ou, Y., Qi, M., Fan, Y.: Corpus Callosum Segmentation Algorithm of Diffusion Tensor Images, Computer Science (Chinese version) **39(12)** (2012), 257–260.

   Wu, Z., Chen, C., Gao, M., Zhu, S.: Segmentation of Diffusion Tensor Image of Brain White Matter Tissues Based on Tensorial Morphological Gradient and Image Forest Transformation Watershed Algorithm, Space Medicine & Medical Engineering, Issue 2 (2011), 139–142.

   Wu, Z., Zhu, S., He, B.: Segmentation of Brain Corpus Callosum Using Graph Cuts Algorithm Based on Diffusion Tensor Imaging, Journal of Zhejiang University (Engineering Science) **45(1)** (2011), 163–167.

   Megyesi, Z. et al: GPU Based Techniques in Medical Imaging, 13th International Conference on Geometry and Graphics, Dresden, 2008 (7 pages)

7. Tomán, H., Tornai, R., Zichar, M.: New DTI Visualization Methods, 13th International Conference on Geometry and Graphics, Dresden, 2008, 1–6. ISBN 978-3-86780-042-6

8. Tomán, H.: On $n-$loops and $(n+1)-$webs, 26th Conference on Geometry and Computer Graphics, Nové Mesto na Morave, 2006, 259–264.

9. Szeghalmy, Sz., Tomán, H., Hajdu, A.: Detecting Digital Intersections Using Line Approximation, 8th International Conference on Applied Informatics, Eger, 2010, 161–171.
Zbl 1245.68236

10. Tomán, H., Hajdu, A., Szakács, J., Hornyik, D., Csutak, A., Pető, T.: Thickness-based Binary Morphological Improvement of Distorted Digital Line Intersections, 5th Hungarian Conference on Computer Graphics and Geometry, Budapest, 2010, 133–139.

11. Papp, I., Tomán, H.: Experiences about the Mathematical and Geometrical Courses for the Students on Faculty of Informatics, 5th International Conference on Applied Informatics, Eger, 2001, 225–228.

Bibliography

[1] Akivis, M. A.: The Canonical Expansions of the Equations of a Local Analytic Quasigroup, Dokl. Akad. Nauk SSSR **188** (1969), 967–970.

[2] Akivis, M. A., Goldberg, V. V.: Algebraic Aspects of Web Geometry, Commentationes Mathematicae Universitatis Carolinae **41/2** (2000), 205–236.

[3] Akivis, M. A., Shelekov, A. M.: On the Canonical Coordinates in a Local Analytic Loop, Webs and Quasigroups, Gos. Univ. Kalinin **130** (1986), 120–124.

[4] Ali, K. M., Pazzani, M. J.: Error Reduction Through Learning Multiple Descriptions, Machine Learning **24/3** (1996), 173–202.

[5] Altincay, H.: On Naive Bayesian Fusion of Dependent Classifiers. Pattern Recognition Letters **26** (2005), 2463–2473.

[6] Appel, M. J., Russo, R. P.: On the $h-$diameter of a Random Point Set (Technical report), The University of Iowa, 2008.

[7] Appel, M. J., Najim, C. A., Russo, R. P.: Limit Laws for the Diameter of a Random Point Set, Adv. in Appl. Probab. **34/1** (2002), 1–10.

[8] Basser, P. J., Mattiello, J., LeBihan, D.: MR Diffusion Tensor Spectroscopy and Imaging, Biophysical Journal **66** (1994), 259–267.

[9] Beaulieu, C.: The Basis of Anisotropic Water Diffusion in the Nervous System, NMR Biomed. **15** (2002), 435–455.

[10] Beaulieu, C., Allen, P. S.: Determinants of Anisotropic Water Diffusion in Nerves, Magn. Reson. Med. **31** (1994), 394–400.

[11] Bödi, R., Kramer, L.: Differentiability of Continuous Homomorphisms Between Smooth Loops, Result. Math. **25** (1994), 13–19.

[12] Cho, S., Kim, J.: Combining Multiple Neural Networks by Fuzzy Integral for Robust Classification, IEEE Trans. Systems, Man and Cybernetics **25/2** (1995), 380–384.

[13] Demirekler, M., Altincay, H.: Plurality Voting-based Multiple Classier Statistically Independent with Respect to Dependent Classifier Sets, Pattern Recognition **35** (2002), 2365–2379.

[14] Dufour, J-P.: Rigidity of Webs, Web Theory and Related Topics, edited by Grifone, J. and Salem, E., World Scientific (2001), 106–113.

[15] Dufour, J-P., Jean, P.: Rigidity of Webs and Family of Hypersurfaces, Singularities and Dynamical Systems, edited

by Pnevmatikos, S. N., Elsevier Science Publisher B.V., North-Holland (1985), 271–283.

[16] Fuchs, T. J., Haybaeck, J., Wild, P. J., et al.: Randomized Tree Ensembles for Object Detection in Computational Pathology, Proc. 5th International Symposium on Advances in Visual Computing (ISVC 2009): Part I, 367–378.

[17] Gilat, D.: Monotonicity of a Power Function: An Elementary Probabilistic Proof, The American Statistician **31** (1977), 91–93.

[18] Goldberg, V. V.: Theory of Multidimensional (n+1)−Webs, Kluwer Academic Publishers, 1988.

[19] Hajdu, A., Hajdu, L., Jónás, Á., Kovács, L., Tomán, H.: Generalizing the Majority Voting Scheme to Spatially Constrained Voting, IEEE Trans. on Image Processing **22(11)** (2013), 4182–4194.

[20] Hajdu, A., Hajdu, L., Kovács, L., Tomán, H.: Diversity Measures for Majority Voting in the Spatial Domain, Lecture Notes in Artificial Intelligence **8073** (2013), 314-323.

[21] Hansen, L. K., Salamon, P.: Neural Network Ensembles, IEEE Transactions on Pattern Analysis and Machine Intelligence **12** (1990), 993–1001.

[22] Harangi, B., Qureshi, R. J., Csutak, A., Pető, T., Hajdu, A.: Automatic Detection of the Optic Disc Using Majority Voting in a Collection of Optic Disc Detectors, 7th IEEE International Symposium on Biomedical Imaging (IEEE Press), Rotterdam, 2010, 1329–1332.

[23] Ho, T. K., Hull, J., Srihari, S.: Decision Combination in Multiple Classifier Systems, IEEE Trans. on Pattern Analysis and Machine Intelligence **16/1** (1994), 66–75.

[24] Huang, Y. S., Suen, C. Y.: A Method of Combining Multiple Experts for the Recognition of Unconstrained Handwritten Numerals, IEEE Trans. Pattern Analysis and Machine Intelligence **17/1** (1995), 90–94.

[25] Kolter, J. Z., Maloof, M. A.: Dynamic Weighted Majority: An Ensemble Method for Drifting Concepts, The Journal of Machine Learning Research **8** (2007), 2755–2790.

[26] Kong, E. B., Diettrich, T.: Error-Correcting Output Coding Corrects Bias and Variance, 12th International Conference on Machine Learning (ICML 1995), 313–321.

[27] Kozma, J.: On the Differentiability of Loopmultiplication in Canonical Coordinate System, Publ. Math. Debrecen **37** (1990), 313–325.

[28] Kozma, J.: Behaviour of Loops in a Canonical Coordinate, Arch. Math. **55** (1990), 498–502.

[29] Kozma, J.: Loops With and Without Subloops, Acta Sci. Math. (Szeged) **55** (1991), 21–31.

[30] Kuncheva, L. I.: Combining Pattern Classifiers, Methods and Algorithms. John Wiley & Sons, Inc., New Jersey (2004)

[31] Kuncheva, L. I., Whitaker, C. J., Shipp, C. A.: Limits on the Majority Vote Accuracy in Classifier Fusion, Pattern Analysis and Applications **6** (2003), 22–31.

[32] Kuzmin, E. N.: The Connection Between Malcev Algebras and Analytic Moufang Loops, (Russian), Algebra i Logika **10** (1971), 3–22.

[33] Lalonde, M., Beaulieu, M., Gagnon, L.: Fast and Robust Optic Disc Detection Using Pyramidal Decomposition and Hausdorff-based Template Matching, IEEE Trans. Med. Imaging **20/11** (2001), 1193–1200.

[34] Lam, L., Suen, C. Y.: Application of Majority Voting to Pattern Recognition: An Analysis of Its Behavior and Performance, IEEE Transactions on Systems, Man, and Cybernetics Part A: Systems and Humans **27** (1997), 553–568.

[35] Lam, L., Suen, C. Y.: Optimal Combinations of Pattern Classifiers, Pattern Recognition Letters **16** (1995), 945–954.

[36] Lam, L., Suen, C. Y.: Multiple Classifier Combination Methodologies for Different Output Levels, Lecture Notes in Computer Science **1857** (2000), 52–66.

[37] Lazar, M.: White Matter Tractography: An Error Analysis and Human Brain Fiber Tract Reconstruction Study (Thesis), University of Utah, 2003.

[38] Littlestone, N., Warmuth, M.: The Weighted Majority Algorithm, Proc. 30th Symposium on Foundations of Computer Science (SFCS 1989), 256–261.

[39] Mahfouz, A. E., Fahmy, A. S.: Ultrafast Localization of the Optic Disc Using Dimensionality Reduction of the Search Space, Med. Image Comput. Assist. Interv. **12** (2009), 985–992.

[40] Matan, O.: On Voting Ensembles of Classifiers, Proc. AAAI-96 Workshop on Integrating Multiple Learned Models (1996) 84–88.

[41] McReynolds, T., Blythe, D.: Advanced Graphics Programming Using OpenGL, Morgan Kaufmann, 2005.

[42] Dataset MESSIDOR provided by the Messidor program partners [Online]. Available: http://messidor.crihan.fr.

[43] Mori, S., van Zijl, P. C. M.: Fiber Tracking: Principles and Strategies - A Technical Review, NMR Biomed. **15** (2002), 468–480.

[44] Moseley, M. E., Cohen, Y., et al.: Diffusion Weighted MR Imaging of Anisotropic Water Diffusion in Cat Central Nervous System, Radiology **176** (1990), 439–445.

[45] Nagy, P. T., Strambach, K.: Loops in Group Theory and Lie Theory, Expositions in Mathematics, Vol. 35. De Gruyter, 2002.

[46] Papadakis, N. G., Xing, D., Huang, C. L., et al.: A Comparative Study of Acquisition Schemes for Diffusion Tensor Imaging Using MRI, J. Magn. Reson. **137** (1999), 67–82.

[47] Perez-Rovira, A., Trucco, E.: Robust Optic Disc Location via Combination of Weak Detectors, Proc. IEEE Engineering in Medicine and Biology Society, 30th Annual International Conference (EMBS 2008), 3542–3545.

[48] Pierpaoli, C., Jezzard, P., et al.: Diffusion Tensor MR Imaging of Human Brain, Radiology **201** (1996), 637–648.

[49] Pierpaoli, C., Basser, P. J.: Toward a Quantitative Assessment of Diffusion Anisotropy, Magn. Reson. Med. **36** (1996), 893–906.

[50] Rost, R. J.: OpenGL ® Shading Language, Addison Wesley Professional, 2006.

[51] Saranli, A., Demirekler, M.: On Output Independence and Complementariness in Rank-based Multiple Classfier Decision Systems, Pattern Recognition **34** (2001), 2319–2330.

[52] Shapley, L., Grofman, B.: Optimizing Group Judgemental Accuracy in the Presence of Interdependencies, Public Choice **43** (1984), 329–343.

[53] Sirlantzis, K., Hoque, S., Fairhurst, M. C.: Diversity in Multiple Classifier Ensembles Based on Binary Feature Quantisation with Application to Face Recognition, Appl. Soft Comput. **8/1** (2008), 437–445.

[54] Sternberg, S.: On the Behaviour of Invariant Curves near a Hyperbolic Point of a Surface Transformation, American Journal of Mathematics **75** (1955), 526–534.

[55] Sternberg, S.: Local Contractions and a Theorem of Poincaré, Am. J. Math. **79** (1957), 809–824.

[56] Stieltjes, B., Kaufmann, W. E., van Zijl, P. C. M., et al.: Diffusion Tensor Imaging and Axonal Tracking in the Human Brainstem, NeuroImage **14** (2001), 723–735.

[57] Taylor, A. J.: Diffusion Tensor Imaging: Evaluation of Tractography Algorithm Performance Using Ground Truth Phantoms (Thesis), Virginia Polytechnic Institute and State University, 2004.

[58] Tomán, H.: Canonical Coordinate Systems and Exponential Maps of $n-$loop, Note di Matematica **24/2** (2005), 1–7.

[59] Tomán, H.: On $n-$loops and $(n+1)-$webs, 26th Conference on Geometry and Computer Graphics, Nové Mesto na Morave, 2006, 259–264.

[60] Tomán, H., Kovács, L., Jónás, Á., Hajdu, L., Hajdu, A.: Generalized Weighted Majority Voting with an Application to Algorithms Having Spatial Output, Lecture Notes in Artificial Intelligence **7209** (2012), 56–67.

[61] Tomán, H., Kovács, L., Jónás, Á., Hajdu, L., Hajdu, A.: A Generalization of Majority Voting Scheme for Medical Image Detectors, Lecture Notes in Artificial Intelligence **6679/2** (2011), 189–196.

[62] Tomán, H., Tornai, R., Zichar, M.: Complex Fiber Visualization, Ann. Math. Inform. **34** (2007), 103–109.

[63] Tomán, H., Tornai, R., Zichar, M.: New DTI Visualization Methods, 13th International Conference on Geometry and Graphics, Dresden, 2008, 1–6. ISBN 978-3-86780-042-6

[64] van Gelderen, P., de Vleeschouwer, M. H., et al.: Water Diffusion and Acute Stroke, Magn. Reson. Med. **31** (1994), 154–163.

[65] Wang, X., Davidson, N. J.: The Upper and Lower Bounds of the Prediction Accuracies of Ensemble Methods for Binary Classification, Proc. Int. Conf. Mach. Learn. Appl. 2010, 373–378.

[66] Xu, L., Krzyzak, A., Suen, C. Y.: Several Methods for Combining Multiple Classifiers and Their Applications in

Handwritten Character Recognition, IEEE Trans. on System, Man and Cybernetics **22/3** (1992), 418–435.