AGH University of Science and Technology in Kraków

Faculty of Electrical Engineering, Automatics, Computer Science and Biomedical Engineering

PhD Dissertation

Pattern recognition in superpixel graphs

Author:

Mateusz Baran

Supervisor:

Zbisław Tabor, Ph.D.

Kraków 2018

OŚWIADCZENIE AUTORA PRACY

OŚWIADCZAM, ŚWIADOMY ODPOWIEDZIALNOŚCI KARNEJ ZA POŚWIADCZENIE NIEPRAWDY, ŻE NINIEJSZĄ ROZPRAWĘ DOKTORSKĄ WYKONAŁEM OSOBIŚCIE I SAMODZIELNIE, I NIE KORZYSTAŁEM ZE ŹRÓDEŁ INNYCH NIŻ WYMIENIONE W PRACY.

.....

PODPIS

Akademia Górniczo-Hutnicza im. Stanisława Staszica w Krakowie

Wydział Elektrotechniki, Automatyki, Informatyki i Inżynierii Biomedycznej

ROZPRAWA DOKTORSKA

ROZPOZNAWANIE WZORCÓW W GRAFACH SUPERPIKSELOWYCH

AUTOR:

Mateusz Baran

PROMOTOR:

dr hab. Zbisław Tabor, prof. PK

Kraków 2018

Serdecznie dziękuję moim rodzicom i promotorowi za wsparcie udzielane podczas pracy nad opisanymi w tej rozprawie badaniami.

Abstract

The primary goal of this thesis is development of an algorithm for accurate object recognition in a given image based on its statistical model of shape. A common approach to this problem is construction of a Point Distribution Model (PDM), initial placement of these points in an image (either by hand or using feature detectors) and iterative optimization of their position, minimizing a certain function combining conformity to the model on one hand and matching image features on the other. Algorithms of this type include Active Shape Model, Pictorial Structures, Constrained Local Models and Regression-Voting. These methods are very effective in certain applications but their performance is strongly affected by appropriate selection of feature descriptors and detectors, correct placement of landmarks in training images and sufficiently good initialization of landmark positions in an analysed image. In certain applications, for example in quantitative analysis of medical images, a higher accuracy than provided by state-of-the-art algorithms is needed.

In response to this demand two object segmentation methods based on statistical shape model are proposed and analysed in this thesis. These methods are based on analysing a superpixel segmentation of a given image, thus constraining the possible shapes to paths in a graph defined by this preprocessing method. In this way the number of possible shapes is greatly reduced, limiting the search space. It allows for application of global optimization methods, solving the problem of finding an initial shape match.

Moreover, a continuous approach to shape representation based on Elastic Shape Analysis (ESA) is explored in this thesis. This approach is well-suited for pattern recognition in superpixel graphs. The standard Dynamic Programming algorithm of ESA is extended in this thesis. Instead of just comparing two curves, it can efficiently find a pair of paths in two digraphs that are the closest under an elastic metric of ESA. This task was solvable only for certain simple classes of graphs using existing methods. The new algorithm is analysed in this thesis.

Finally, a new machine learning-based approach to extraction of shape features is developed and integrated into the novel ESA-based algorithm. This combined method has comparable accuracy to state-of-the-art methods, good performance and a low number of free parameters. The thesis ends with a list of possible directions for further development of the presented image segmentation algorithm.

Streszczenie

Celem rozprawy jest opracowanie algorytmu pozwalającego na dokładne rozpoznawanie wzorca będącego obiektem na obrazie w oparciu o jego model statystyczny. Typowym podejściem w tego rodzaju problemach jest zbudowanie modelu opartego o punkty charakterystyczne (Point Distribution Model, PDM), wstępne rozmieszczenie tych punktów na zadanym obrazie (ręcznie lub w oparciu o detektory cech) i iteracyjne optymalizowanie ich położeń minimalizując pewną funkcję celu wyważającą z jednej strony zgodność z modelem, a z drugiej dopasowanie do cech obrazu. Do algorytmów tego typu można zaliczyć Active Shape Model, Pictorial Structures, Constrained Local Models i Regression-Voting. Metody te w wielu zastosowaniach cechuje wysoka skuteczność, jednak jest ona zależna od właściwego doboru cech i detektorów cech, umieszczenia punktów charakterystycznych we właściwych, odpowiadających sobie miejscach obrazów oraz dostatecznie dobrej inicjalizacji rozmieszczenia tych punktów na analizowanym obrazie. W niektórych zastosowaniach, na przykład w ilościowej analizie obrazów medycznych, wymagana jest dokładność większa niż oferowana przez istniejące algorytmy.

W odpowiedzi na to zapotrzebowanie dwie metody rozpoznawania obiektów w oparciu o ich model kształtu są proponowane i analizowane w tej rozprawie. Metody te opierają się o analizę obrazu po segmentacji superpikselowej, która ogranicza możliwe kształty do ścieżek w grafie określonym wspomnianą procedurą przetwarzania wstępnego. W ten sposób redukowana jest liczba możliwych przebiegów kształtu, znacząco zmniejszając przestrzeń rozwiązań którą trzeba przeszukać. Pozwala to na zastosowanie metod optymalizacji globalnej, rozwiązując problem inicjalizacji.

Rozprawa zajmuje się również kwestią reprezentacji kształtu w oparciu o funkcje ciągłe, inspirowaną Elastyczną Analizą Kształtu (Elastic Shape Analysis, ESA). Podejście to bardzo dobrze pasuje do rozpoznawania wzorców w grafach superpikselowych. Zamiast porównywać dwie krzywe, zaproponowany algorytm w sposób wydajny znajduje parę ścieżek w dwóch grafach skierowanych które są sobie najbliższe w metryce elastycznej rozważanej w ESA. Zadanie to było rozwiązywalne z użyciem istniejących metod tylko dla prostych klas grafów. Rozprawa zawiera analizę omawianego algorytmu.

Wymieniony nowy algorytm jest ponadto zintegrowany ze specjalnie zaprojektowaną metodą ekstrakcji cech kształtu opartą o uczenie maszynowe. Uzyskana w ten sposób metoda dorównuje skuteczności najlepszych istniejących algorytmów segmentacji obrazu, działając stosunkowo szybko i posiadając niewielką

liczbę wolnych parametrów. Rozprawa podaje też dalsze możliwe kierunki rozwoju zaproponowanego algorytmu segmentacji.

Contents

List of Figures

List of Tables

List of Symbols and Abbreviations

Symbol	Meaning	Page
$\mathrm{L}^2([0,1],\mathbb{R}^n)$	The space of square-integrable functions	10
$\langle\cdot,\cdot\rangle_{\mathbb{R}^n}$	Standard inner product in $\mathbb{R}^n$	10
T_pM	Tangent space at point p to a manifold M	11
TM	Tangent bundle to a manifold M	11
$\exp_p$	An exponential map at point p	13
$\log_p$	A logarithmic map at point p	13
$\mathrm{SO}(n,\mathbb{R})$	Special orthogonal group	14
ASM	Active Shape Model	15
AAM	Active Appearance Model	15
ESA	Elastic Shape Analysis	16
SRV	Square Root Velocity transform	17
d^c_{pre}	A partially invariant elastic distance	17
Γ	Group of orientation-preserving diffeomorphisms of $[0,1]$	18
$q\cdot\gamma$	Action of $\gamma\in\Gamma$ on an SRV function q	18
d^c	A fully invariant elastic distance	19
S^n	An n-dimensional sphere	21
$\langle\cdot,\cdot\rangle_{(\phi,\theta)}$	A general elastic metric	21
PGA	Principal Geodesic Analysis	21
PCA	Principal Component Analysis	21
i.i.d.	independent and identically distributed	22
PDM	Point Distribution Model	26
ROC	Receiver Operating Characteristic	32

Symbol	**Meaning**	**Page**
AUC	Area under the ROC curve	32
T^n	An n-dimensional torus	34
$G_1 \times_\sigma G_2$	Sigma-product of digraphs G_1 and G_2	58

Chapter 1

Introduction

Motivation and scope

Segmentation is one of the most important tasks in computer vision (Haralick and Shapiro, 1992, Freixenet et al., 2002). In this work segmentation is understood in two similar but distinct meanings. First, it is a task of grouping pixels in a given image into regions that are uniform with respect to a certain property but distinct between themselves. This meaning is used when referring to, for example, superpixel segmentation. The second, more narrow meaning is grouping pixels in an image into two regions: an object of interest (usually of a specified type) and the background. This task is sometimes referred to as binarization (Sauvola and Pietikäinen, 2000). Segmentation in the second meaning is the primary goal in this thesis.

An important factor with significant impact on the outcome of image segmentation is appropriate usage of a priori knowledge about the shape to be found. This knowledge is typically extracted from other images of a similar type into a statistical model of shape. Many of the most common approaches to shape representation for segmentation are based on landmarks (Dryden and Mardia, 1998, Cootes et al., 1995). While they are very successful, the problem of selection and identification of landmarks has not been satisfactorily solved (Zhang and Golland, 2016, Gao et al., 2010). Another problem faced by landmark-based methods is adequate statistical modelling of nonlinear shape changes. Proper handling of this issue would significantly increase the computational complexity of algorithms. Additionally, many popular segmentation methods perform a local search for the best match, making them dependent on sufficiently good initialization. This initialization is typically obtained using feature detectors (Martins et al., 2016) but they are not fully reliable (Mukherjee et al., 2015, Mainali et al., 2014).

Commonly used algorithms based on local optimization, applied in standard approaches to landmark-based segmentation, are sensitive to initialization. One possible solution of this problem is provided by global optimization methods, receiving an increasing attention for different image processing and computer vision tasks (El-Zehiry and Grady, 2010, Schoenemann et al., 2009, Bruhn et al., 2014, Chen and Koltun,

2016). The interest in global methods is supported by their successful applications. Some of the most prominent examples in this category is the graph cut algorithm (Boykov and Jolly, 2001) or Conditional Random Fields methods (Lafferty et al., 2001).

A significant challenge for global optimization-based image segmentation is efficient combination with a statistical model of shape to be found in a given image. Machine learning methods are employed to efficiently construct and apply such a model of shape. Easy comprehensibility of the shape model was not pursued as a goal, however it may be a future direction of research. While certain existing approaches rely on approximate global optimization methods (Mesejo et al., 2013, Heimann and Meinzer, 2009), their performance and robustness is not sufficient for all applications (Hum et al., 2014, Wojciechowski et al., 2016, Huber and Ronchetti, 2009).

The problem of landmark selection can be solved by placing them uniformly along the boundary. The performance loss associated with selecting more landmarks than is necessary was mitigated by appropriately constructed further stages of proposed algorithms. In particular, the most promising approach presented in this thesis explains the dense landmark selection by using ideas from continuous approaches to shape representation pioneered by, among others, Grenander and Miller (1998) and Younes (1998). These approaches, and in particular Elastic Shape Analysis (Srivastava and Klassen, 2016), provide a very well motivated solution to landmark identification, extended in this thesis for a new and interesting case of graphs representing superpixel segmentations.

In this work a superpixel segmentation is used to construct a global optimization algorithms for the problem of image segmentation. The primary goals are very good accuracy, good efficiency and a small space of input parameters. Further research in this area is important as it is a promising direction for a new generation of segmentation algorithms that reduce the problems of achieving efficient and accurate solutions in particular applications to practical computer vision problems.

The study is limited to digital greyscale images represented by a uniform rectangular grid of pixels. The task is to find a simply connected area in the image that corresponds to an object of a known type, thus its border may be represented as a non-intersecting open or closed planar curve. It is assumed that the border of the searched object coincides with borders between certain superpixels of a superpixel segmentation of the input image. This is the core assumption that enables application of a certain class of methods based on global optimization that received only a limited attention in the past (Greig et al., 1989, Shi and Malik, 2000, Boussaid et al., 2014).

The algorithm may be provided a limited amount of information about the position of the searched object, for example endpoints of the curve representing its border. These points are considered to be correctly established. Exhaustive search among all pairs of vertices of the superpixel graph would lead to a fully

global and parallelizable search strategy. The endpoints may also be automatically located in the image using feature descriptors and detectors but this task is outside of the scope of this thesis.

Objectives and plan of work

The main objective of this thesis is proving that **superpixel segmentation methods are a viable preprocessing step for pattern recognition based on global optimization**. This goal is divided into three research hypotheses:

1. It is possible to achieve high segmentation accuracy when the search space is constrained by a superpixel segmentation of the input image.
2. There is an efficient, global, low-parameter method of image segmentation that respects the superpixel constraints.
3. Machine learning methods can be effectively used to augment the above-mentioned algorithm to match the accuracy of state-of-the-art algorithms with limited parameter tuning.

Chapter 2 reviews the literature relevant the this thesis, primarily the topics of image and shape representation, Principal Geodesic Analysis, image segmentation and machine learning. This discussion is accompanied by a short introduction to mathematical tools that are used to describe the ideas behind presented algorithms.

Chapter 3 describes an initial approach to the problem of pattern recognition in superpixel graphs. An iterative algorithm is designed to find a path in the superpixel graph of a given image that conforms the best to a statistical model of shape. Results described in this chapter support the first research hypothesis.

Chapter 4 presents an extension of Elastic Shape Analysis to the pairs of graph drawings. The extended algorithm is thoroughly analysed and it is shown that it can be used to solve the problem of shape matching in superpixel graphs. Under certain conditions, the exact globally optimal matching is returned by the algorithm. These results support the second research hypothesis.

Next, in Chapter 5, the extended elastic shape matching algorithm is combined with a statistical model of shape. This model is extracted from raw annotated training data using machine learning methods. A specialized feature extraction procedure is designed to convert the raw data to input vectors for classification algorithms. High accuracy of this image segmentation method combined with small space of free parameters and decent performance for a global optimisation method, as well as high potential for further development, make this algorithm a promising alternative to state-of-the-art approaches. This proves the third research hypothesis.

Finally, the thesis is summarized in Chapter 6. Conclusions based on results described in previous chapters are presented. The summary also contains a description of areas left open for future research.

Contents and original contribution

The following results are the most important original contributions of this thesis:

1. An analysis of the effects of constraining the task of object segmentation by a superpixel segmentation.
2. Extension of curve matching in Elastic Shape Analysis to pairs of planar graphs (presented in Section 4.1).
3. Application of machine learning methods to graph-based Elastic Shape Analysis for improved segmentation accuracy (presented in Section 5.1).

Partial results of the research presented in this dissertation were published or accepted for publication in international journals:

- **Mateusz Baran**, Zbisław Tabor. Principal Geodesic Analysis Boundary Delineation with Superpixel-based Constraints, *Image Analysis & Stereology*, vol. 36, pp. 223–232, 2017. Impact Factor: **1.135** (for 2016).
- **Mateusz Baran**. Closest Paths in Graph Drawings under an Elastic Metric. Accepted for publication in *International Journal of Applied Mathematics and Computer Science*. Impact Factor: **1.420** (for 2016).

Other co-authored works include:

- T. Pięciak, **M. Baran**, M. Ubrańczyk. Level-set based segmentation of carotid arteries in computer tomography angiography images, *Journal of Medical Informatics & Technologies*, vol. 17, pp. 281–286, 2011.
- **M. Baran**, L. K. Bieniasz. Experiments with an adaptive multicut-HDMR map generation for slowly varying continuous multivariate functions, *Applied Mathematics and Computation*, vol. 258, pp. 206–219, May 2015. Impact Factor: **1.345** (for 2015).
- **M. Baran**, L. K. Bieniasz. An adaptive multicut-HDMR map generation, in *AIP Conference Proceedings*, 2016, vol. 1738, p. 480055.
- **M. Baran**, K. Kluza, G. J. Nalepa, A. Ligęza. A hierarchical approach for configuring business processes, in *2013 Federated Conference on Computer Science and Information Systems (FedCSIS)*, Krakow, 2013, pp. 915–921.
- **M. Baran**, A. Ligęza. Rule-Based Knowledge Management in Social Threat Monitor, in *Multimedia Communications, Services and Security*, A. Dziech, A. Czyżewski, Eds. Springer Berlin Heidelberg, 2013, pp. 1–12.
- S. Bobek, **M. Baran**, K. Kluza, G. J. Nalepa. Application of Bayesian Networks to Recommendations in Business Process Modeling, presented at the AIBP@AI*IA, 2013, pp. 41–50.

- K. Kluza, **M. Baran**, S. Bobek. Overview of Recommendation Techniques in Business Process Modeling, presented at the KESE 2013.
- **M. Baran**, K. Kułakowski, A. Ligęza. A Note on Machine Learning Approach to Analyze the Results of Pairwise Comparison Based Parametric Evaluation of Research Units, in *Artificial Intelligence and Soft Computing*, L. Rutkowski, M. Korytkowski, R. Scherer, R. Tadeusiewicz, L. A. Zadeh, J. M. Zurada, Eds. Springer International Publishing, 2014, pp. 27–39.
- **M. Baran**, K. Kluza, G. J. Nalepa, A. Ligęza. A Multi-level Hierarchical Approach for Configuring Business Processes, in *Advances in ICT for Business, Industry and Public Sector*, M. Mach-Król, C. M. Olszak, T. Pełech-Pilichowski, Eds. Springer International Publishing, 2015, pp. 1–18.
- **M. Baran**. Multivariate function approximation using sparse grids and high Dimensional Model Representation – a comparison, *Czasopismo Techniczne*, Nauki Podstawowe Zeszyt 3 NP (17) 2014, pp. 97–107, Feb. 2015.

Chapter 2

Literature review

Mathematical preliminaries

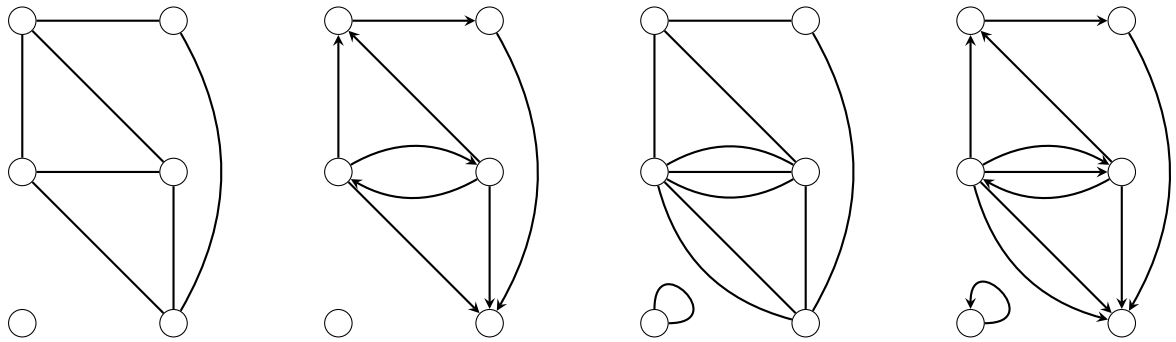
This section briefly describes a few mathematical topics that are used further in this thesis. Graph theory is used as a language for handling superpixel segmentation. Lie theory and even more importantly differential geometry are the language of Elastic Shape Analysis and geometric statistical methods like Principal Geodesic Analysis. For this reason and because of more theoretic nature of these topics they are presented separately.

Graph theory

Graph theory is a popular tool in image processing (Grady, 2004), used even in early image processing applications (Zahn, 1971). Applications include image denoising (e.g. using graph cuts (Boykov and Jolly, 2001, Bae et al., 2011)), segmentation (Greig et al., 1989, Felzenszwalb and Huttenlocher, 2004, Zahn, 1971, Shi and Malik, 2000, Wang and Siskind, 2003, Li et al., 2012), image processing on a foveal sensor using connectivity graphs (Wallace et al., 1994), detection of salient groupings (Perona and Freeman, 1998) or salient regions (Yang et al., 2013). Graphs are also used in computer vision to describe structure of a scene (Sarkar and Soundararajan, 2000). Many methods represent images as graphs where vertices correspond to pixels and edges denote adjacency, although some authors use other approaches.

Hypergraphs (Berge, 1984) gained some popularity for image representation (Bretto, 2004, Bretto and Gillibert, 2005) and processing (Rital et al., 2001, Li et al., 2013). These methods are, although, significantly different from other discussed in this dissertation and will not be described here.

Definition 1. *A **graph** (also called an **undirected graph**) $G = (V, E)$ is an ordered pair where the first element, V, is a non-empty finite set of vertices and the second element, E, is a set of edges. Each element $e \in E$ is a two-element set (an unordered pair) of vertices $e = \{v_1, v_2\}$ where $v_1, v_2 \in V$ (Wilson, 1996, Harary, 1969).*



(a) A graph (b) A digraph (c) A multigraph (d) A multidigraph

Figure 2.1: Graphical representations of a graph, a digraph, a multigraph and a multidigraph

Definition 2. *A **digraph** (also called a **directed graph**) $G = (V, E)$ is an ordered pair where the first element, V, is a non-empty finite set of vertices and the second element, E is a set of edges. Each element $e \in E$ is an ordered pair of different vertices $e = (v_1, v_2)$ where $v_1, v_2 \in V$ and $v_1 \neq v_2$ (Wilson, 1996, Harary, 1969).*

Definition 3. *A **multigraph** (called graph in (Bondy and Morty, 1982)) $G = (V, E, r)$ is an ordered triple where V is a nonempty finite set of vertices, E is a finite set of edges (disjoint from V) and $r\colon E \to \{\{v_1, v_2\}\colon v_1, v_2 \in V\}$ is an incidence function assigning to every edge a set of two not necessarily distinct vertices.*

Definition 4. *A **multidigraph** (called digraph in (Bondy and Morty, 1982)) $G = (V, E, r)$ is an ordered triple where V is a nonempty finite set of vertices, E is a finite set of edges (disjoint from V) and $r\colon E \to \{(v_1, v_2)\colon v_1, v_2 \in V\}$ is an incidence function assigning to every edge an ordered pair of two not necessarily distinct vertices.*

Although the concept of a graph had existed in science before formal introduction of graph theory (Foulds, 1992), some basic terms like graph or digraph are not consistently defined in literature. Figure 2.1 contains examples of graphical representations of a graph, a digraph, a multigraph and a multidigraph. Vertices are represented by circles, undirected edges by lines and directed edges by arrows. Vertices in all types of graphs are often labelled to enable distinguishing them easily.

Definition 5. *A graph (or a digraph) $G_1 = (V_1, E_1)$ is a **subgraph** of a graph (or, respectively digraph) $G_2 = (V_2, E_2)$ if $V_2 \subseteq V_1$ and $E_2 \subseteq E_1$ (Bondy and Morty, 1982).*

Definition 6. *A **path** in a graph or a digraph $G = (V, E)$ is a sequence of vertices $v_1, v_2, \ldots, v_n \in V$ such that they are pairwise different ($v_i \neq v_j$ if $i \neq j$ for $i, j \in \{1, 2, \ldots, n\}$) and each pair of consecutive vertices v_i, v_{i+1} for $i = 1, 2, \ldots, n-1$ is connected by an edge (Wilson, 1996, Bondy and Morty, 1982), that is $\{v_i, v_{i+1}\} \in E$ in case of a graph or $(v_i, v_{i+1}) \in E$ in case of a digraph.*

Definition 7. *A **cycle** in a graph or a digraph $G = (V, E)$ is a sequence of vertices $v_1, v_2, \ldots, v_n \in V$ such that $n \geq 2$, elements of the sequence are pairwise different except the first and the last ones which are equal ($v_i \neq v_j$ if $i \neq j$ for $i, j \in \{1, 2, \ldots, n-1\}$ and $v_1 = v_n$) and each pair of consecutive vertices v_i, v_{i+1} for $i = 1, 2, \ldots, n-1$ is connected by an edge, that is $\{v_i, v_{i+1}\} \in E$ in case of a graph or $(v_i, v_{i+1}) \in E$ in case of a digraph (Wilson, 1996).*

Definition 8. *A **routing algebra** is a triple $(W, \oplus, \preceq)$ such that $(W, \oplus)$ is a monoid with zero element and $\preceq$ is a total order on W with a maximal element equal to the zero element of $(W, \oplus)$ (Zubor et al., 2014). Equivalently, the following conditions hold for all $w_1, w_2, w_3 \in W$:*

- *$w_1 \oplus w_2 \in W$ (closure),*
- *$(w_1 \oplus w_2) \oplus w_3 = w_1 \oplus (w_2 \oplus w_3)$ (associativity),*
- *there is an identity element $e \in W$ such that for all $w \in W$ the following holds: $e \oplus w = w \oplus e = w$,*
- *there is a zero element $\phi \in W$ such that for all $w \in W$ the following holds: $\phi \oplus w = w \oplus \phi = \phi$,*
- *$w_1 \preceq w_1$ (reflexivity),*
- *if $w_1 \preceq w_2$ and $w_2 \preceq w_1$, then $w_1 = w_2$ (antisymmetry),*
- *if $w_1 \preceq w_2$ and $w_2 \preceq w_3$, then $w_1 \preceq w_3$ (transitivity),*
- *either $w_1 \preceq w_2$ or $w_2 \preceq w_1$ (totality),*
- *there is a maximal element $\infty \in W$ such that $w \preceq \infty$ for all $w \in W$,*
- *maximal element is equal to the zero element ($\infty = \phi$).*

Routing algebras, originally proposed by Sobrinho (2002), are a framework for specifying complex ways of calculating a cost of a path in a graph. In this thesis a simplified definition of a routing algebra is used (Zubor et al., 2014). More commonly used definitions of a routing algebra (Sobrinho, 2003, 2005) make the distinction between the set of edge weights and the set of path costs which would be an unnecessary complication here.

Definition 9. *A routing algebra $(W, \oplus, \preceq)$ is called **commutative** if for all $w_1, w_2 \in W$ the equality $w_1 \oplus w_2 = w_2 \oplus w_1$ holds (Zubor et al., 2014).*

Definition 10. *A routing algebra $(W, \oplus, \preceq)$ is called **monotonic** if for all $w_1, w_2 \in W$ the inequalities $w_1 \preceq w_2 \oplus w_1$ and $w_1 \preceq w_1 \oplus w_2$ hold (Yang and Wang, 2008).*

Definition 11. *A routing algebra $(W, \oplus, \preceq)$ is called **isotonic** if for all $w_1, w_2, w_3 \in W$ the property $w_1 \preceq w_2$ implies both $w_1 \oplus w_3 \preceq w_2 \oplus w_3$ and $w_3 \oplus w_1 \preceq w_3 \oplus w_2$ hold (Yang and Wang, 2008).*

Zubor et al. (2014) use the term *non-decreasing* instead of monotonic and *monotone* instead of isotonic. All routing algebras considered in this thesis are commutative. For such algebras common routing algorithms, with few exceptions, require them to be isotonic to find an optimal path. Additionally, the Dijkstra's algorithm for routing algebras requires monotonicity (Yang and Wang, 2008).

Functional analysis

Functional analysis studies spaces of functions. It has found many applications in image analysis (Siddiqi, 2003). This area of mathematics introduces a few concepts relevant to this thesis, most notably the space of square-integrable functions L^2.

Definition 12. *The space $\mathrm{L}^2([0,1],\mathbb{R}^n)$ where $n > 0$ is the space of all square-integrable functions $f\colon [0,1] \to \mathbb{R}^n$ (Titchmarsh, 1939, Rudin, 1986, 1991). Two functions are considered equal if they differ at the set of points of Lebesgue measure 0. This space is a Hilbert space with an inner product of two functions $f, g \in \mathrm{L}^2([0,1],\mathbb{R}^n)$ equal to*

$$\langle f, g\rangle = \int_0^1 \langle f(t), g(t)\rangle_{\mathbb{R}^n}\,\mathrm{d}t, \tag{2.1}$$

where $\langle\cdot,\cdot\rangle_{\mathbb{R}^n}$ is the standard inner product in $\mathbb{R}^n$.

Differential geometry

Differential geometry is a mathematical theory gaining popularity among image processing researchers. It provides a formalism for dealing with spaces with non-Euclidean geometry. Many such spaces, like spheres, arise naturally. Attempts at exploiting their intrinsic properties in image processing were made as early as in 1990 (Lenz, 1990, Grenander and Miller, 1998, Dupuis et al., 1998), although certain manifolds were used for particular purposes even earlier (Smith and Jain, 1982, Natterer, 1980). These efforts intensified during the last decade (Fletcher et al., 2003, 2009, Fletcher, 2013, Zhang and Fletcher, 2013, Zhang and Golland, 2016, Miller and Qiu, 2009).

There are numerous mathematical books introducing concepts of differential geometry (Lang, 1985, 1996, Spivak, 1999, Abraham et al., 2001). It is, although, troublesome that many important concepts, like even the fundamental concept of a differential manifold, do not have a single universally accepted definition. For this reason some important concepts of differential geometry are introduced in this section.

Definition 13. *Let X be a set and E be a Euclidean (or, more generally, Banach) space. An **atlas** of class C^p (for $p \geq 0$) on X is a family of pairs $\{(U_i, \phi_i)\}_{i\in I}$ called **charts** satisfying the following conditions: (Lang, 1983)*

1. *Each U_i is a subset of X and the family $\{U_i\}_{i\in I}$ covers X, that is $X = \bigcup_{i\in I} U_i$.*
2. *Each ϕ_i is a bijection of U_i onto an open subset of E, and for every pair i, j of indices the set $\phi_i(U_i \cap U_j)$ is open in E.*
3. *For each pair of indices i, j the map $\phi_{ij}\colon \phi_i(U_i \cap U_j) \to \phi_j(U_i \cap U_j)$ defined by $\phi_{ij} = \phi_j \circ \phi_i^{-1}$ is a bijection and a C^p function.*

A **differential manifold** M is a set E equipped with an atlas. Hereafter, all differential manifolds are considered to be equipped with an atlas of class C^∞ unless noted otherwise. In this thesis the set E is either a Euclidean space $\mathbb{R}^n$ or an $\mathrm{L}^2([0,1],\mathbb{R}^n)$ space for some $n \geq 1$. In the former case the manifold M has a finite dimension n. Sometimes additional properties are assumed, for example for topological manifolds M is a Hausdorff space and is second-countable (Tu, 2010).

In a Euclidean space $\mathbb{R}^n$ for some $n \geq 1$ one can define a tangent vector at a point $x \in \mathbb{R}^n$ in two ways. Kinematic tangent vectors at x are derivatives of smooth curves[1] $c\colon \mathbb{R} \to \mathbb{R}^n$ such that $c(0) = x$ at 0. Operational tangent vectors at x are directional derivatives at x of differentiable functions $f\colon \mathbb{R}^n \to \mathbb{R}$. Both definitions result in a vector space that is canonically isomorphic to $\mathbb{R}^n$: a kinematic tangent vector is identified with the derivative itself and an operational tangent vector is identified with the direction of derivation. Both definitions can be used as a basis to define a local linearisation of a differential manifold and, for finite-dimensional manifolds, the definitions coincide (Spivak, 1999). It is, however, not true in general. For example in case of the space $\mathrm{L}^2([0,1],\mathbb{R}^n)$ there are operational tangent vectors that do not correspond to any kinematic tangent vectors (Kriegl and Michor, 1997, 1991). In the relevant literature the kinematic definition is used (Srivastava and Klassen, 2016), although this issue has not been discussed.

Definition 14. *Let M be a differential manifold with a smooth atlas $\{(U_i,\phi_i)\}_{i\in I}$. Furthermore, let $p \in M$ be any point on the manifold M. Any two differentiable curves $\alpha,\beta\colon \mathbb{R} \to M$ such that $\alpha(0) = \beta(0) = p$ are equivalent with respect to the equivalence relation $\sim$ if and only if for any chart (U_i,ϕ_i) such that $p \in U_i$ the derivatives of $\phi_i \circ \alpha$ and $\phi_i \circ \beta$ are equal at 0:*

$$\frac{\mathrm{d}}{\mathrm{d}t}\phi_i(\alpha(t))\bigg|_{t=0} = \frac{\mathrm{d}}{\mathrm{d}t}\phi_i(\beta(t))\bigg|_{t=0}. \tag{2.2}$$

*The **tangent space** (or the **kinematic tangent space**) to a differential manifold M at $p \in M$, denoted by T_pM, is the quotient vector space of differentiable curves passing through p by the relation $\sim$ (Spivak, 1999, Srivastava and Klassen, 2016):*

$$T_pM = \{\gamma\colon \mathbb{R} \to M\colon \gamma(0) = p\}/\sim. \tag{2.3}$$

Definition 15. ***Tangent bundle*** *is the disjoint union of all spaces tangent to a manifold M, $TM = \bigcup_{p\in M}(p, T_pM)$ (Spivak, 1999).*

Definition 16. *For any curve $\alpha\colon \mathbb{R} \to M$ of class C^1 one can define a **velocity curve** $\dot{\alpha}\colon \mathbb{R} \to TM$ such that*

$$\dot{\alpha}(t) = (\alpha(t), [u \mapsto \alpha(u-t)]_\sim) \tag{2.4}$$

[1] Smoothness (when referring to functions on continuous domains) in this work is understood as existence of all derivatives (C^∞) unless noted otherwise. This should not be confused with smoothing of digital images.

where $[u \mapsto \alpha(u-t)]_\sim$ is the equivalence class of the curve α shifted by t with respect to relation $\sim$ (see Definition 14).

Definition 16 can be extended to curves defined on closed intervals in a straightforward way.

Definition 17. *Let M, N be differential manifold with atlases $\{(U_i, \phi_i)\}_{i\in I_M}$ and $\{(V_j, \xi_j)\}_{j\in J_N}$ of class C^p for some $p \in \{0, 1, \ldots, \infty\}$. A function $f\colon M \to N$ is said to be of class C^p if for all points $p \in M$ and all pairs of charts (U_i, ϕ_i), (V_j, ξ_j) such that $p \in U_i$ and $f(p) \in V_j$ the function $\xi_j \circ f \circ \phi_i^{-1}$ is of class C^p (Srivastava and Klassen, 2016, Spivak, 1999).*

Definition 18. *The **differential** (or a **pushforward**) of a smooth function $f\colon M \to N$ at point $p \in M$ where M, N are differential manifolds with C^∞ atlases is the function $\mathrm{d}f_p\colon T_pM \to T_{f(p)}N$ such that for each vector $v_p \in T_pM$*

$$(\mathrm{d}f_p(v_p))g = v_p(g \circ f) \tag{2.5}$$

for each real-valued function $g \in C^\infty(N)$ (Tu, 2010). On both sides of Equation (2.5) application of a vector from the tangent space to a function is understood as a directional derivative in the direction of that vector.

Some definitions from Riemannian geometry are necessary for understanding certain topics in shape analysis. Such mathematical tools are gaining popularity. This topic is covered in more depth in e.g. (Michor and Mumford, 2006).

Definition 19. *A **vector field** on a differential manifold M is any function $f\colon M \to TM$ such that for each point $p \in M$ the value of $f(p)$ is a pair (p, v_p) where $v_p \in T_pM$.*

Any C^∞ atlas on M can be used to define a C^∞ atlas on TM, thus equipping it with a structure of a differential manifold. If a vector field is a smooth function (see Definition 17), then it is called a **smooth vector field**.

Definition 20. ***Riemannian metric** $g_p\colon T_pM \times T_pM \to \mathbb{R}$ is an inner product defined on tangent spaces T_pM for all $p \in M$ that changes smoothly with p, that is for any two smooth vector fields $f, g\colon M \to TM$ such that $f(p) = (p, v_{f,p})$ and $g(p) = (p, v_{g,p})$ the function $e\colon M \to \mathbb{R}$ defined as*

$$e(p) = g_p(v_{f,p}, v_{g,p}) \tag{2.6}$$

is of class C^∞ (Srivastava and Klassen, 2016, Spivak, 1999).

Definition 21. ***Length** of a curve $\alpha\colon [0,1] \to M$ is defined by (Spivak, 1999)*

$$L(\alpha) = \int_0^1 \sqrt{g_{\alpha(t)}(v_t, v_t)}\,\mathrm{d}t, \tag{2.7}$$

assuming $\dot{\alpha}(t) = (\alpha(t), v_t)$.

Definition 22. ***Energy*** *of a curve* $\alpha\colon [0,1] \to M$ *is defined by (Spivak, 1999)*

$$E(\alpha) = \int_0^1 g_{\alpha(t)}(v_t, v_t)\,\mathrm{d}t, \tag{2.8}$$

assuming $\dot{\alpha}(t) = (\alpha(t), v_t)$ *for all* $t \in [0,1]$.

Definition 23. ***Geodesic*** *is a critical point of the energy functional (Definition 22) on a manifold (Spivak, 1999).*

Geodesics are not only lowest-energy curves between two points p, q on a manifold M, but also lowest-length curves between these two points. There may exist exactly one geodesic between p and q (up to a reparametrization), more than one (even infinitely many) or none. Distance between these two points is defined as an infimum over the lengths of smooth paths $\alpha\colon [0,1] \to M$ connecting them ($\alpha(0) = p$ and $\alpha(1) = q$):

$$d(p,q) = \inf\{L(\alpha)\colon \alpha \text{ is a smooth path joining } p \text{ and } q\}. \tag{2.9}$$

Note that in general the distance may be defined even for pairs of functions not connected by any geodesic.

Definition 24. *Given a point* $p \in M$ *the* ***exponential map*** $\exp_p\colon T_pM \supseteq V \to M$ *assigns to a vector* $v \in V$ *from the tangent space the point* $\gamma_v(1)$ *where* $\gamma\colon [0,1] \to M$ *is the unique geodesic satisfying* $\gamma(0) = p$ *and* $\frac{\mathrm{d}\gamma}{\mathrm{d}t}(0) = v$ *(Spivak, 1999). The domain* V *of the exponential map is a neighbourhood of the zero vector containing vectors for which such a geodesic exists.*

Definition 25. ***Logarithmic map*** $\log_p\colon M \supseteq N \to T_pM$ *is the inverse of an exponential map* $\exp_p$ *(Spivak, 1999).*

The concepts of exponential and logarithmic maps are illustrated in Figure 2.2. A vector $v \in V$ is transformed by the exponential map into a point $q = \exp_p(v) \in M$, while a point $q \in N$ corresponds to the vector $v = \log_p(q) \in T_pM$ through the logarithmic map. Intuitively, a tangent space is a local linearisation of a manifold M around a given point $p \in M$. Exponential and logarithmic maps are used to translate between the manifold M and the tangent space T_pM.

Lie theory

Lie theory is the theory of groups and algebras that are also differential manifolds.

Definition 26. *A group* G *is a* ***Lie group*** *if it is also a manifold with a* C^∞ *atlas and both the group operation and the group inverse are smooth functions (Spivak, 1999).*

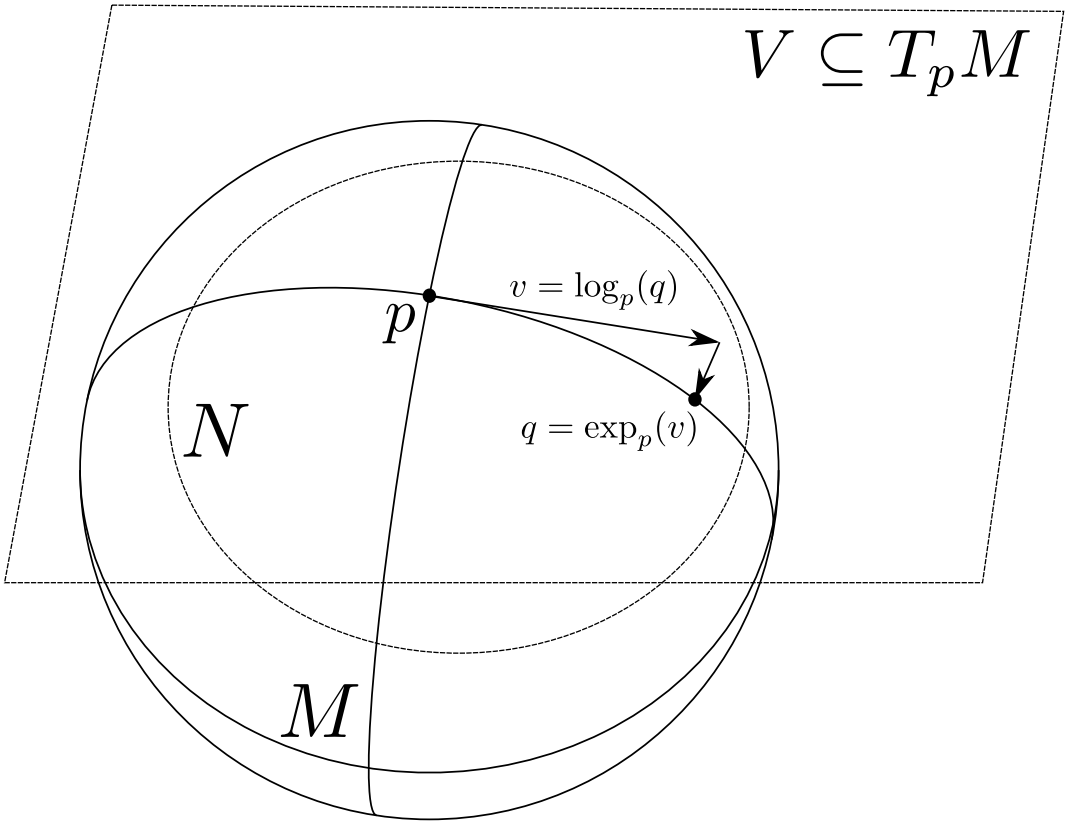


Figure 2.2: A manifold M with a selected point $p \in M$. Exponential map transforms a subset V of the tangent space T_pM to M. Logarithmic map transforms a submanifold $N \subseteq M$ to T_pM

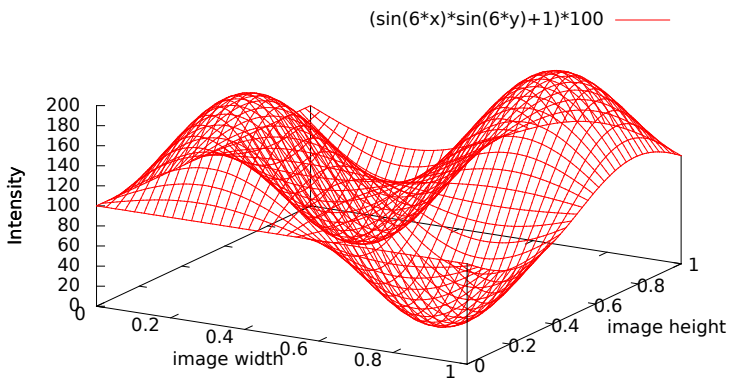
Lie theory is useful in formal analysis of invariants of metrics. For example rotational invariance in a Euclidean space $\mathbb{R}^n$ can be described as invariance under action of the special orthogonal Lie group $\mathrm{SO}(n, \mathbb{R})$, that is the group of all rotations (without reflection) of $\mathbb{R}^n$. This is also true for other operations, like translation and scaling (Srivastava and Klassen, 2016).

Image representation

Images are represented by functions of two arguments with finite domains (Gonzalez and Woods, 2007, Jähne, 2002). Some authors use more complex schemes (Pratt, 2007) but since the reasons for their introduction (for example processing of non-monochrome or three-dimensional images, processing of image sequences) are not considered in this thesis, a more simple formalism is selected. For digital images, the domain is discrete: $f\colon \{1, 2, \ldots, W\} \times \{1, 2, \ldots, H\} \to \mathbb{R}$ where W is the width of the image represented by f and H is its height. A rectangular grid is always used. Wherever a continuous image is concerned, its domain is the unit square: $f\colon [0, 1]^2 \to \mathbb{R}$. Only greyscale images are considered in this thesis, hence in both cases the codomain of f is the set of real numbers. In case of digital images, effects of quantization are not taken into account.

6	66	29	57	124	168	149
5	53	2	41	134	195	168
4	83	65	79	112	134	124
3	128	159	135	79	41	57
2	147	199	159	65	2	29
1	122	147	128	83	53	66
	1	2	3	4	5	6

(a) A digital image. Each cell contains intensity of a particular pixel. The numbers below and on the left side of the grid represent, respectively, first and second coordinates of the image



(b) A continuous image. Intensity is represented by the vertical axis

Figure 2.3: A comparison of a digital image and a continuous image

Digital and continuous image representations are compared in the Figure 2.3. Problems of coding and representation of digital image in a computer (for example (Huo, 1999, Bretto, 2004, Bretto and Gillibert, 2005)), including image compression (Shukla et al., 2010), are outside of the scope of this dissertation.

Shape representations and analysis

Shape representation is a problem in image processing of formal encoding shapes of objects and devising methods of their manipulation. D'Arcy Thompson's book (Thompson, 2014) (originally printed in 1917) is considered as one of the earliest attempts in comparing shapes invariantly with respect to certain transformations.

Shape representations fall into one of two categories: point-based and domain-based. Methods from the first category use sets or lists of points to represent shapes. These shapes can be compared using a number of methods:

- Iterative Closest Point (ICP) (Besl and McKay, 1992, Arun et al., 1987, Chen and Medioni, 1992),
- Active Shape Models (ASM) (Cootes et al., 1995),
- Active Appearance Models (AAM), an extension to ASM (Cootes et al., 2001, Elder et al., 2003, Gao et al., 2010),
- Ordinary Procrustes Analysis (OPA) and Generalized Procrustes Analysis (GPA) (Gower, 1975, Goodall, 1991), with many variants and extensions (Koschat and Swayne, 1991, Theobald and Wuttke, 2006, Turaga and Srivastava, 2016, Brignell, 2007),
- Kendall's shape spaces (Kendall, 1984, 1989, Kendall et al., 1999), which aims to provide a more precise mathematical description of shape from Procrustes Analysis,

- bicycle chain shape model (Sommer et al., 2009),
- Elastic Shape Analysis (ESA) (Younes, 1998, Mio and Srivastava, 2004, Srivastava et al., 2006, Joshi et al., 2007, Srivastava et al., 2011),
- deformation-based shape representation (Demisse et al., 2016, 2017),
- shape analysis using conformal mapping (Sharon and Mumford, 2006),
- other energy-based methods (Cohen et al., 1992, Sebastian et al., 2003, Khaneja et al., 1998),
- other PDE-based methods (Frenkel and Basri, 2003).

Remarkably, some of these methods, like ESA and conformal mapping based shape analysis, conceptually represent shapes as infinite-dimensional objects (Younes, 1998, Sharon and Mumford, 2006). Some methods like Active Contours (Kass et al., 1988), including variants like Sobolev Active Contours (Sundaramoorthi et al., 2006, 2007), have different focus than direct shape comparison but they can be used in this way as well (Schnabel and Arridge, 1995).

Domain-based methods represent shapes as functions whose domain is the entire image, i.e. $[0, 1]^2$ in its continuous form (see 2.1.5). Among shape representation methods in this category are:

- Level-set methods (Osher and Sethian, 1988, Osher and Fedkiw, 2001, 2003, Pięciak et al., 2011),
- Deformation-based shape analysis (Trouvé, 1998, Grenander and Miller, 1998, Grenander, 1994, Grenander and Keenan, 1993).

An important topic in shape analysis is the selection of invariants that are assumed to preserve shapes. Most commonly they are translations, scaling and rotations but in a smaller number of works affine and projective transformations are also considered (Ambartzumian, 1990, Aguado et al., 2002, Bryner et al., 2014), see also (Guo et al., 2013).

Elastic Shape Analysis

The most important family of methods for representing and comparing shapes for this thesis is the Elastic Shape Analysis (ESA). This method was initially described in (Younes, 1998). Many important developments were made during the last decade (Michor and Mumford, 2006, Mio et al., 2007, Sundaramoorthi et al., 2011), including introduction of the Square Root Velocity function representation (Joshi et al., 2007) (although it was actually used earlier (Younes, 1999)), the path straightening method for computation of geodesics (Srivastava et al., 2011) and a textbook summary of the field (Srivastava and Klassen, 2016). ESA has also been extended to surfaces (Kurtek et al., 2013) and quotients of Lie group by compact subgroups (Su et al., 2017).

Elastic Shape Analysis has been applied to many problems, for example palm print identification (Mokni et al., 2016) and shape morphing (Mio et al., 2007). Biological applications include comparison of axon

morphology (Mottini et al., 2015), clustering of brain fibres (Kurtek et al., 2012, Mani et al., 2010) and protein structure alignment (Liu et al., 2010). Very encouraging results were obtained in these cases.

In Elastic Shape Analysis shapes are represented as curves from certain shape spaces. The exact shape space is determined by the group of transformations that do not change the shape. The most general space for the case of no invariants is the space of absolutely continuous functions $\mathrm{AC}([0,1],\mathbb{R}^n)$, that is functions $f\colon [0,1]\to\mathbb{R}^n$ such that

1. f is differentiable almost everywhere and
2. $f(t)=f(0)+\int_0^t \dot{f}(u)\,\mathrm{d}u$ for all $t\in[0,1]$.

In the first step a Square Root Velocity (SRV) transformation is applied. A function $f\in\mathrm{AC}([0,1],\mathbb{R}^n)$ is changed into a function

$$q(t)=Q(\dot{f}(t)), \tag{2.10}$$

where the function $Q\colon \mathbb{R}^n\to\mathbb{R}^n$ is defined by

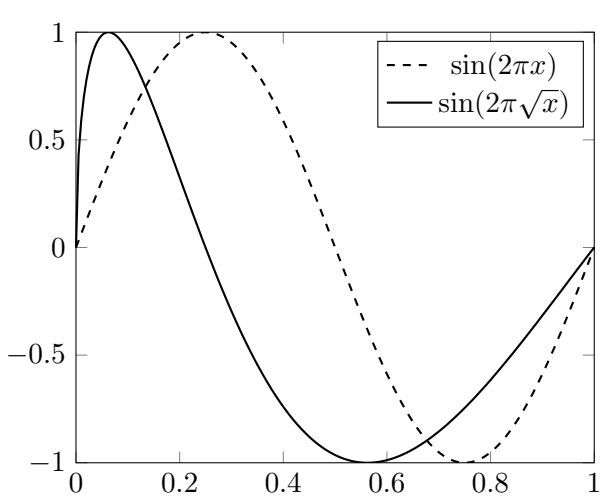
$$Q(v)=\begin{cases}\frac{v}{\sqrt{\|v\|}} & \text{if } \|v\|\neq 0\\ 0 & \text{if } \|v\|=0.\end{cases} \tag{2.11}$$

The resulting function q belongs to the space of square-integrable functions $\mathrm{L}^2([0,1],\mathbb{R}^n)$. As a result of differentiation, this representation is automatically translation-invariant: for all constants $c\in\mathbb{R}^n$ and functions $f\in\mathrm{AC}([0,1],\mathbb{R}^n)$, SRV representation of f and $f+c$ are the same. On the other hand, differentiation in Equation (2.10) makes the representation more susceptible to noise (Demisse et al., 2017). It is, although, true for most shape analysis metrics (Mennucci, 2013). Figure 2.4 compares SRVs of two functions where one is a reparametrization of the other. It shows that the magnitude of the SRV of a function is correlated with the speed of traversal of a part of its domain. For the selected reparametrization the effect is the most visible near the endpoints of the domain.

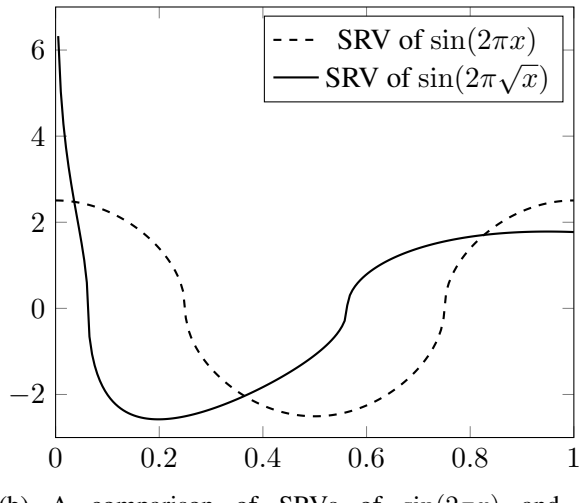
The standard L^2 metric on the space of SRV representations can be used for shape analysis. It is an elastic metric when understood as a metric on the original space $\mathrm{AC}([0,1],\mathbb{R}^n)$. This elastic metric can be seen as an extension to the nonparametric Fisher-Rao metric (Rao, 1945). This metric can be used to define a distance between shapes described by curves $f_1, f_2\colon [0,1]\to\mathbb{R}^n$ as follows:

$$d^c_{pre}(q_1,q_2)=\sqrt{\int_0^1 \langle q_1(t),q_2(t)\rangle_{\mathbb{R}^n}\,\mathrm{d}t}, \tag{2.12}$$

where q_1 and q_2 are SRV transforms of, respectively, f_1 and f_2, as defined by Equation (2.10)).



(a) A comparison of $\sin(2\pi x)$ and $\sin(2\pi\sqrt{x})$



(b) A comparison of SRVs of $\sin(2\pi x)$ and $\sin(2\pi\sqrt{x})$

Figure 2.4: A comparison of $\sin(2\pi x)$ and $\sin(2\pi\sqrt{x})$ and their SRV transforms

An important property of d^c_{pre} is partial reparametrization invariance. Let Γ be the set of orientation preserving diffeomorphisms from $[0,1]$ to itself, that is

$$\Gamma = \{\gamma\colon [0,1] \to [0,1]\colon \gamma \text{ is a diffeomorphism, } \gamma(0) = 0 \text{ and } \gamma(1) = 1\}. \tag{2.13}$$

The set Γ with function composition is a group, that is composition of two elements of Γ is an element of Γ, the identity on $[0,1]$ belongs to Γ and for every reparametrization $\gamma \in \Gamma$ its inverse γ^{-1} belongs to Γ as well. Associativity is provided as a general property of function composition.

Let us first observe that reparametrization of original curve provides an action of the group Γ on the space of SRV representations of curves. This action is defined by:

$$(q \cdot \gamma)(t) \equiv Q\left(\frac{\mathrm{d}}{\mathrm{d}t} f(\gamma(t))\right) = q(\gamma(t))\sqrt{\dot{\gamma}(t)}, \tag{2.14}$$

where q is an SRV of f and $\gamma \in \Gamma$ is an arbitrary reparametrization.

Now, let q_1 and q_2 be two arbitrary SRV functions and $\gamma \in \Gamma$ be a reparametrization. The partial invariance can be observed as follows:

$$\begin{aligned}\left(d^c_{pre}(q_1 \cdot \gamma, q_2 \cdot \gamma)\right)^2 &= \int_0^1 \|(q_1 \cdot \gamma)(t) - (q_2 \cdot \gamma)(t)\|_2^2 \,\mathrm{d}t = \\ &\int_0^1 \|(q_1 \circ \gamma)(t) - (q_2 \circ \gamma)(t)\|_2^2 \dot{\gamma}(t) \,\mathrm{d}t = \\ &\int_0^1 \|q_1(t) - q_2(t)\|_2^2 \,\mathrm{d}t = \left(d^c_{pre}(q_1, q_2)\right)^2,\end{aligned} \tag{2.15}$$

where the next to last step involved integration by substitution. Now, a fully invariant distance function can be defined by minimization over the set of reparametrizations Γ:

$$d^c(q_1, q_2) \equiv \min_{\gamma \in \Gamma} d^c_{pre}(q_1, q_2 \cdot \gamma). \tag{2.16}$$

For this definition it is essential that Γ with function composition is a group that acts by isometries. More details can be found in (Srivastava and Klassen, 2016).

The terms elastic distance and elastic metric can, in many places, be used interchangeably. The former is primarily used in places where considering the exact metric space or Riemannian manifold is not important.

The dynamic programming algorithm for calculation of elastic distance

There is no general method for determining the value of the elastic distance d^c between a given pair of curves. Usually the set Γ of reparametrizations is limited to a certain finite subset and the minimization is performed using a Dynamic Programming algorithm (Srivastava et al., 2011, Bernal et al., 2016). The algorithm described in this thesis is a minor modification of Algorithm 58 from (Srivastava and Klassen, 2016). Discretization that does not admit any admissible reparametrization is explicitly detected and the symbol N_{ij} (used in the cited book) is redefined in terms of a set σ, described in Section 4.1.

Algorithm 1 describes the procedure of calculating the elastic distance d^c between two functions with SRV representations q_1, q_2 discretized at $0 = t_{1,0} < t_{1,1} < \cdots < t_{1,M_1} = 1$ and $0 = t_{2,0} < t_{2,1} < \cdots < t_{2,M_2} = 1$. The parameter σ encodes allowed slopes of the discretized, piecewise-linear reparametrization γ. The function Q, describing the discretized d^c_{pre} distance between two reparametrized pieces of $\hat{q}_1$ and $\hat{q}_2$, is defined by

$$Q(i, j, k, l) = I(t_{1,i-k}, t_{1,i}, t_{2,j-l}, t_{2,j}, \hat{q}_1, \hat{q}_2), \tag{2.17}$$

where functions $\hat{q}_1, \hat{q}_2$ are SRV representations of first order spline interpolants to curves represented by q_1, q_2 sampled at points, respectively, $\{t_{1,i}\}_{i=0}^{M_1}$ and $\{t_{2,i}\}_{i=0}^{M_2}$. Some authors (Doğan et al., 2015) use higher-order interpolation but for reasons explained in Chapter 4 it is not used in this work. The function I is the trapezoidal quadrature of d^c_{pre}:

$$I(t_{1,i-k}, t_{1,i}, t_{2,j-l}, t_{2,j}, \hat{q}_1, \hat{q}_2) = \int_{t_{1,i-k}}^{t_{1,i}} \left\| \hat{q}_1(t) - \hat{q}_2(\gamma(t)) \sqrt{\dot{\gamma}(t)} \right\|_2^2 \mathrm{d}t, \tag{2.18}$$

where $0 \leq i - k < i \leq M_1$, $0 \leq j - l < j \leq M_2$ and $\gamma(t) = t_{2,j-l} + \frac{t_{2,j} - t_{2,j-l}}{t_{1,i} - t_{1,i-k}}(t - t_{1,i-k})$.

Algorithm 1: Dynamic Programming calculation of $d^c(q_1, q_2)$ discretized at arguments $\{t_{1,i}\}_{i=0}^{M_1}, \{t_{2,i}\}_{i=0}^{M_2}$.

Require: $\{t_{1,i}\}_{i=0}^{M_1}, \{t_{2,i}\}_{i=0}^{M_2}, \hat{q}_1, \hat{q}_2, \sigma$
$E(0,0) := 0$
$P(i,j) := (-1,-1)$ for $1 \leq i \leq M_1, 1 \leq j \leq M_2$
for $i \in \{0, 1, \ldots, M_1\}$ **do**
  for $j \in \{0, 1, \ldots, M_2\}$ **do**
    if $(i,j) \neq (0,0)$ **then**
      for $(k,l) \in \sigma$ **do**
        if $i - k \geq 0$ **and** $j - l \geq 0$ **then**
          $E(i,j,k,l) := E(i-k, j-l) + Q(i,j,k,l)$
        else
          $E(i,j,k,l) := \infty$
        end if
      end for
      $E(i,j) = \min_{(k,l)\in\sigma} E(i,j,k,l)$
      if $E(i,j) < \infty$ **then**
        $(k_{opt}, l_{opt}) = \arg\min_{(k,l)\in\sigma} E(i,j,k,l)$
        $P(i,j) = (i - k_{opt}, j - l_{opt})$
      end if
    end if
  end for
end for
$(i_1, i_2) := (M_1, M_2)$
$\gamma_d := \{(i_1, i_2)\}$ {A discrete reparametrization.}
while $(i_1, i_2) \neq (0,0)$ **do**
  $(i_1, i_2) := P(i_1, i_2)$
  if $(i_1, i_2) \neq (-1,-1)$ **then**
    PREPEND$(\gamma_d, (i_1, i_2))$
  else
    return "ERROR" {There are no reparametrizations satisfying given constraints.}
  end if
end while
return $E(M_1, M_2), \gamma_d$ {Returns cost and a reparametrization that realizes this cost.}

The general elastic metric

Historically, a different and more general transformation based on complex numbers was proposed first for the case planar curves (Younes, 1998) and later extended to curves in higher-dimensional Euclidean spaces (Mio et al., 2007). SRV representation is, though, often preferred for its relative ease of application. In this more general version, a function $f\colon [0,1] \to \mathbb{R}^n$ is first transformed into a pair of functions $\phi\colon [0,1] \to \mathbb{R}$, $\theta\colon [0,1] \to \mathrm{S}^{n-1}$ such that

$$\phi(t) = \ln(\|\dot{f}(t)\|), \qquad \theta(t) = \frac{\dot{f}(t)}{\|\dot{f}(t)\|}. \tag{2.19}$$

This pair of functions defines a map from the space of open curves to the set $\Phi \times \Theta$, where Φ and Θ are sets of functions, respectively, ϕ and θ corresponding to open curves f. The tangent space to $\Phi \times \Theta$ at (ϕ, θ) is given by

$$T_{(\phi,\theta)}(\Phi \times \Theta) = \Phi \times \{v \in \mathrm{L}^2([0,1], \mathbb{R}^2) \colon \langle v(t), \theta(t) \rangle_{\mathbb{R}^n} = 0 \text{ for all } t \in [0,1]\}, \tag{2.20}$$

where a standard embedding of the sphere S^{n-1} in $\mathbb{R}^n$ as a unit sphere is assumed and the scalar product $\langle \cdot, \cdot \rangle_{\mathbb{R}^n}$ is the standard scalar product in $\mathbb{R}^n$. The general elastic metric is provided by the following definition.

Definition 27. *A **general elastic metric** on $\Phi \times \Theta$ is given by*

$$\langle (u_1, v_1), (u_2, v_2) \rangle_{(\phi,\theta)} = a^2 \int_0^1 u_1(t) u_2(t) e^{\phi(t)} \,\mathrm{d}t + b^2 \int_0^1 \langle \phi_1(t), \phi_2(t) \rangle_{\mathbb{R}^n} e^{\phi(t)} \,\mathrm{d}t, \tag{2.21}$$

where a and b are positive real numbers (Mio et al., 2007, Srivastava et al., 2011).

It is easy to prove that the metric from Definition 27 also defines a metric on the original space of curves. When $a = \frac{1}{2}$ and $b = 1$, this general elastic metric coincides with the L^2 metric on the space of SRV representations (Srivastava et al., 2011).

Other developments in Elastic Shape Analysis

Other developments in Elastic Shape Analysis include tree registration (Mottini et al., 2015), landmark-guided elastic surface registration and morphing (Kurtek et al., 2013), graph-constrained surface registration (Zeng et al., 2016) and analysis of trajectories on manifolds through transported SRV function representation (Su et al., 2012, 2014).

There is one open source library that implements elastic registration (Huang et al., 2016). The source code is available at `http://www.math.fsu.edu/~whuang2/Indices/index_ROPTLIB.html`, see also `http://www.math.fsu.edu/~whuang2/papers/RORCESA.htm`.

The discussion of other important topics in Elastic Shape Analysis, such as handling of closed shapes (Klassen et al., 2004) and statistical modelling of shapes (Srivastava and Jermyn, 2009) is skipped because it will not be used in this thesis.

Principal Geodesic Analysis

Principal Geodesic Analysis (PGA) (Fletcher et al., 2003, 2004) is an extension of Principal Component Analysis (PCA) (Jolliffe, 2002) to differential manifolds. Taking properties of manifolds into account leads

in many cases to more accurate results (Fletcher et al., 2004, Su et al., 2008, Sommer et al., 2010). Additionally, for example many biological shapes change along geodesics in Kendall's shape spaces (Le and Kume, 2000).

Let $x_1, x_2, \ldots, x_N \in M$ be a set of points from a differential manifold M. There are two components of PCA and PGA: mean calculation and mode calculation. Both need to be changed to accommodate manifold-valued data.

There are two common ways to calculate the mean of a given set of points from a manifold. The first one in the so-called intrinsic mean, also called the Karcher mean (Karcher, 1977) or Fréchet mean (Fréchet, 1948). The mean μ is defined by the minimization of sum of squared distances over the entire manifold:

$$\mu = \arg\min_{x \in M} \sum_{i=1}^{N} d(x, x_i)^2, \tag{2.22}$$

where the distance is the intrinsic distance in the manifold (see Equation (2.9)). The formula comes from a more general definition of mean value for probability distributions on manifolds. Existence and uniqueness conditions for the intrinsic mean are studied in e.g. (Karcher, 1977, Kendall, 1990).

One of the important downsides of the intrinsic mean is that it is hard to compute for many manifolds. An alternative way of computing a mean of points on a manifold is the extrinsic mean (Grenander et al., 1998). It uses an embedding $\Phi\colon M \to \mathbb{R}^n$ to translate points into a Euclidean space, calculates the standard mean in this space and projects the result back to the manifold using a projection $\pi\colon \mathbb{R}^n \to M$:

$$\mu = \pi\left(\arg\min_{\bar{x} \in \mathbb{R}^n} \sum_{i=1}^{N} \|\bar{x} - \Phi(x_i)\|^2\right). \tag{2.23}$$

Averaging on certain classes of manifolds is sometimes considered separately (see for example (Hartley et al., 2013) for rotation averaging and (Hauberg et al., 2014) for averaging on Grassmann manifolds). For PGA, other averaging methods like median (Yang, 2010) can also be applied, which is more robust in certain cases (Fletcher et al., 2009). In particular, while the minimization of sum of squares is optimal for independent and identically distributed (i.i.d.) Gaussian noise, it is also sensitive to outliers (Torre and Black, 2003, He et al., 2011, Kwak, 2008).

The second part of PGA is mode calculation. There is, however, no direct generalization of principal components to general manifolds. Geodesics are one possible generalization but on non-geodesically complete manifolds they cannot be extended indefinitely which causes problems. Another problem is that best-fit principal geodesic may not pass through the calculated mean. This line of research is pursued under the name of Geodesic Principal Component Analysis (Huckemann et al., 2010, Huckemann and Ziezold, 2006). It should be noted that these problems are less severe in Riemannian symmetric spaces (Fletcher

et al., 2004, Cornea et al., 2016, Lazar and Lin, 2017) which commonly appear in computer vision. Other approaches to modelling principal components in manifolds are also used like nested manifolds (Jung et al., 2012), barycentric subspaces (Pennec, 2015), principal submanifolds (Yao and Pham, 2016) or horizontal component analysis (Sommer, 2013). Variants of PGA include exact PGA (Sommer et al., 2013), probabilistic PGA (Zhang and Fletcher, 2013) and kernel PGA (Awate et al., 2014).

In PGA, all points $x_1, x_2, \ldots, x_N \in M$ are translated to the tangent space $T_\mu M$ using the logarithmic map, that is

$$v_i = \log_\mu(x_i), \tag{2.24}$$

for $i = 1, 2, \ldots, N$. Next, since the tangent space is isometric to $\mathbb{R}^k$ where k is the dimension of M (it is assumed that M is a finite-dimensional manifold), standard PCA is performed in the tangent space using vectors $v_1, v_2, \ldots, v_N \in T_\mu M$.

This approach uses the fact that the logarithmic map offers a local linearisation of the manifold. The results strongly rely on the assumption that given points lie in an area that can be well approximated in this way. When it is not the case, e.g. for points distributed uniformly across a sphere, a more sophisticated generalization of PCA is necessary. On the other hand, PGA is quite fast and relatively simple (Sommer et al., 2010).

Segmentation methods

Image segmentation is one of the most prominent task in image processing. It is understood as a process of separating an object of interest from the background or dividing an image into constituent parts (Zaitoun and Aqel, 2015, Fu and Mui, 1981, Withey and Koles, 2008). There are many types of segmentation methods. One possibility is to divide them to layer-based methods and block-based methods (Zaitoun and Aqel, 2015). Algorithms in the first category use object detectors to analyse a given image (Yang et al., 2012). Block-based methods on the other hand operate directly on the image.

Segmentation evaluation

An important aspect of segmentation is its evaluation. In this thesis evaluation of single object segmentation is the most important (Zhang, 1996). The measure that was selected is the Dice coefficient (Dice, 1945):

$$QS = \frac{2\,TP}{2\,TP + FP + FN}, \tag{2.25}$$

where TP is the number of pixels correctly classified as belonging to the segmented object, FP is the number of pixels incorrectly classified as belonging to the segmented object and FN is the number of pixels incorrectly classified as background. Other evaluation measures include:

- analytical methods which derive bounds on segmentation performance (Liedtke et al., 1987, Abdou and Pratt, 1979),
- empirical methods that evaluate segmentation algorithms by performance on a certain set of test images. This group can be divided further into:
  - reference-based methods that compare obtained results with reference segmentation of test images (Jaccard index (Jaccard, 1901), maximum likelihood estimate of the fraction of correctly detected edges (Fram and Deutsch, 1975), figure of merit (Pratt, 1978)) and
  - reference-free methods that use criteria such as uniformity to assess the segmentation (busyness measure (Weszka and Rosenfeld, 1978), normalized uniformity (Sahoo et al., 1988), entropy-based methods (Pal and Pal, 1989, Pal and Bhandari, 1993), shape measure (Sahoo et al., 1988)).
- human evaluation, where segmentation results are manually evaluated by humans (Zaitoun and Aqel, 2015).

Analytical methods are only applicable to simple algorithms. Reference-based methods are commonly used for e.g. single object segmentation while reference-free methods are popular tool in evaluating the quality of superpixel segmentation. It is important that in the case of reference-based methods, the reference images may be erroneously segmented (Zhang, 1996), especially in case of non-synthetic images, as a result of human error. This issue further complicates error analysis but is not considered in this thesis.

Watershed segmentation

The watershed segmentation (or watershed transform) (Beucher, 1982, Beucher and Meyer, 1993, Meyer, 1994, 2005, Meyer and Beucher, 1990), originally proposed in (Digabel and Lantuéjoul, 1978, Beucher and Lantuéjoul, 1979), is a popular tool for dividing images into constituent parts. It uses the interpretation of an image as a graph of a function (see Figure 2.3b, it also applies to digital images after interpolation).

The watershed transform consists of several steps. Firstly, local minima in the input image are found. Secondly, the image is simultaneously flooded from each local minimum. This is depicted in Figure 2.5 where water from each flooding source is on the same level. When this step is proceeded, the water level raises uniformly as if all flooding sources were connected (all catchment basins can be understood as communicating vessels (Prokhorov, 1979)). When water from two different minima meets, a dam is built to separate them. These dams are built along watershed lines, that is borders between distinct catchment basins. The flooding is continued until all of the image is covered. Each flooded region corresponds to a single seg-

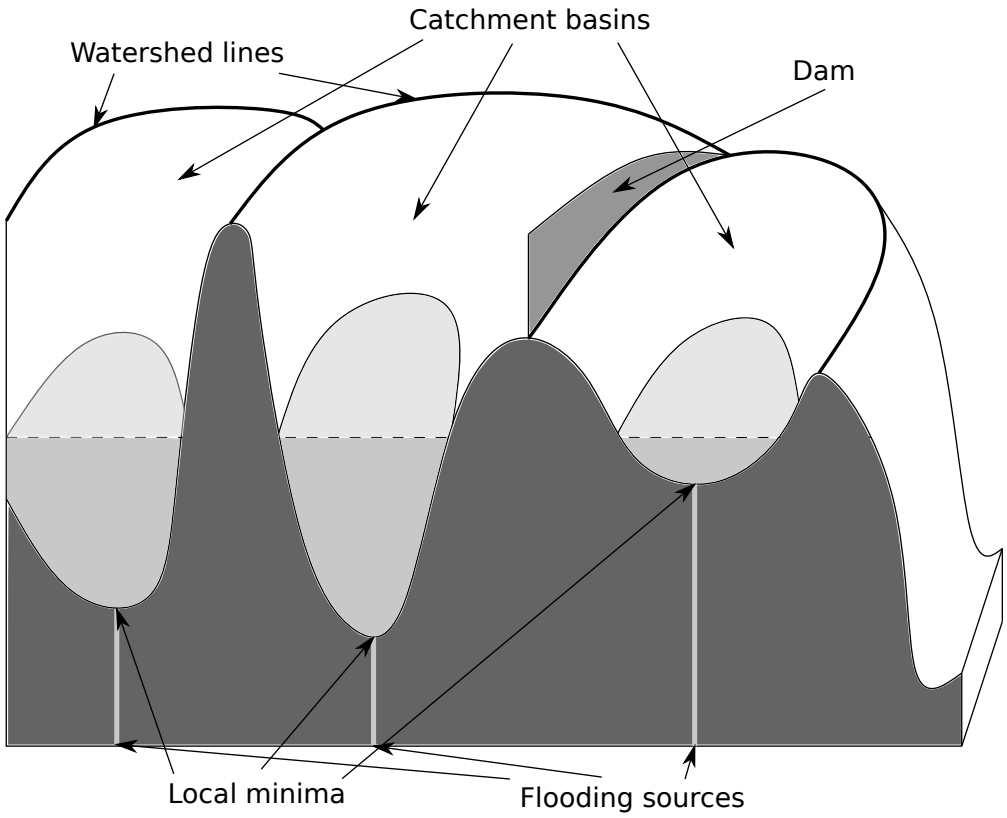


Figure 2.5: Explanation of terms related to the watershed transform

ment of the segmentation. This terminology, illustrated in Figure 2.5, is borrowed from geography. Original meanings of these words (Strahler, 1957) serve as a useful intuition.

The watershed from markers method (Meyer and Beucher, 1990, Meyer, 2005) is a variant of the watershed segmentation when instead of taking each local minimum as a flooding source, only a certain subset of these minima is selected. Different strategies have been developed for selection of flooding sources. An important difference is that in this variant flooding of neighbouring catchment basins may occur. For example, if only two leftmost minima in Figure 2.5 were selected as flooding sources, the marked dam would not be created. As a result, oversegmentation that may occur for example due to noise is reduced (Beucher and Meyer, 1993).

Other variants of watershed transform include stochastic watershed segmentation (Angulo and Jeulin, 2007), a level set-inspired variant (Tai et al., 2007) and a recent generalization of power watersheds (Couprie et al., 2011) that reveals the connection with graph cut and shortest path algorithms. Different implementations of the watershed transform are discussed in (Vincent and Soille, 1991, Verbeek and Verwer, 1990, Beucher and Meyer, 1993, Najman and Schmitt, 1994, Meyer, 1994).

Segmentation based on Point Distribution Model

Segmentation methods based on Point Distribution Model (PDM) use a statistical, deformable model of shape. It is used to ensure that the obtained segmentation represents a shape that is typical for the class of objects that is expected. PDMs are based on modelling variability in relative positions of landmarks. Methods based on PDM became popular after the success of the Active Shape Method (ASM) (Cootes et al., 1995).

PDM based methods need to be trained by supplying a number of sets of landmark positions extracted from different images. These landmarks need to be identified in the image and corresponding landmarks have to be matched across images. Significant manual effort is typically needed in this step.

The ASM has been enhanced in many ways, including adaptation to segmentation of neuroanatomic structures (Duta and Sonka, 1998), Active Appearance Model (Cootes and Taylor, 1999, Cootes et al., 2001), optimal selection of appearance features (van Ginneken et al., 2002), using a hierarchical processing and the wavelet transform (Davatzikos et al., 2003, Nain et al., 2007) and an ASM-driven graph cuts boundary delineation (Chen et al., 2013). As ASM converges to the final shape through local optimization, selection of initial shape is important (Cosío, 2008). An initial shape localization method based on the Hough transform has been proposed by Brejl and Sonka (2000). Use of application-driven constraints on boundary location has been applied by Chen et al. (2011) to three-dimensional image segmentation (Heimann and Meinzer, 2009). In this case constraints represent prior knowledge about relative positions of segmented objects.

Superpixel segmentation

The idea of superpixel segmentation is to divide an image into its natural elementary pieces for segmentation (Ren and Malik, 2003, Moore et al., 2008). There are two main reasons to do this. First, pixels are not natural elementary pieces of an image but rather a consequence of discretization. Second, global optimization at the level of pixels is intractable even for moderately sized images. Thus dividing an image into relatively uniform areas, known as superpixels, is a viable solution to both of these problems. Such superpixels can be used for further processing of an image (Zhang and Ji, 2011).

It is worth noting that the watershed segmentation (described in Section 2.4.2), as it typically leads to an oversegmentation, can be considered a superpixel algorithm (Levinshtein et al., 2009). It is, though, criticized for irregular borders between superpixels and overall lack of superpixel compactness (Levinshtein et al., 2009, Achanta et al., 2010). These issues were addressed to some extent in (Neubert and Protzel, 2014).

Superpixel segmentation has gain a lot of popularity in recent years. This led to introduction of many superpixel algorithms, including normalized cuts (Shi and Malik, 2000), turbopixels (Levinshtein et al., 2009), simple linear iterative clustering (SLIC) (Achanta et al., 2010, 2012) and other (Veksler et al., 2010,

Li et al., 2012). Some other approaches like mean shift (Comaniciu and Meer, 2002), quick shift (Vedaldi and Soatto, 2008) or compact watershed (Neubert and Protzel, 2014) are sometimes also considered as superpixels algorithms.

Figure 2.6 compares three superpixel algorithms (preemptive SLIC, watershed from markers and compact watershed). The compactness parameter is set to 0.5 for compact watershed and 22 for preemptive SLIC. The influence of this parameter on these two methods is depicted in Figures **??** and 2.8. The image used for comparison is an X-ray image of a knee smoothed using an anisotropic diffusion filter (Perona and Malik, 1990). Superpixel borders were obtained for gradient magnitude images.

Dividing an image into more superpixels increases the chance that the border of a selected object corresponds to boundaries between superpixels. On the other hand, processing of finer oversegmentation is more costly. For preemptive SLIC and compact watershed algorithms, the compactness parameter represents a tradeoff between obtaining more compact superpixels (and smoother boundaries between them) and better adherence to edges in the image. This is important because jagged boundaries have a detrimental influence on the performance of certain algorithms (see Section 4.2).

Superpixel algorithms have been used for image analysis in conjunction with graph algorithms previously (Achanta et al., 2010), for example in (Li et al., 2004, Fulkerson et al., 2009). One of the approaches was to use Conditional Random Fields (Lafferty et al., 2001) on superpixels instead of pixels.

Machine learning

Machine learning is a broad discipline encompassing many different problems, such as classification, regression, clustering, structured learning and other (Russell and Norvig, 2009, Cios et al., 2007, Flach, 2012,

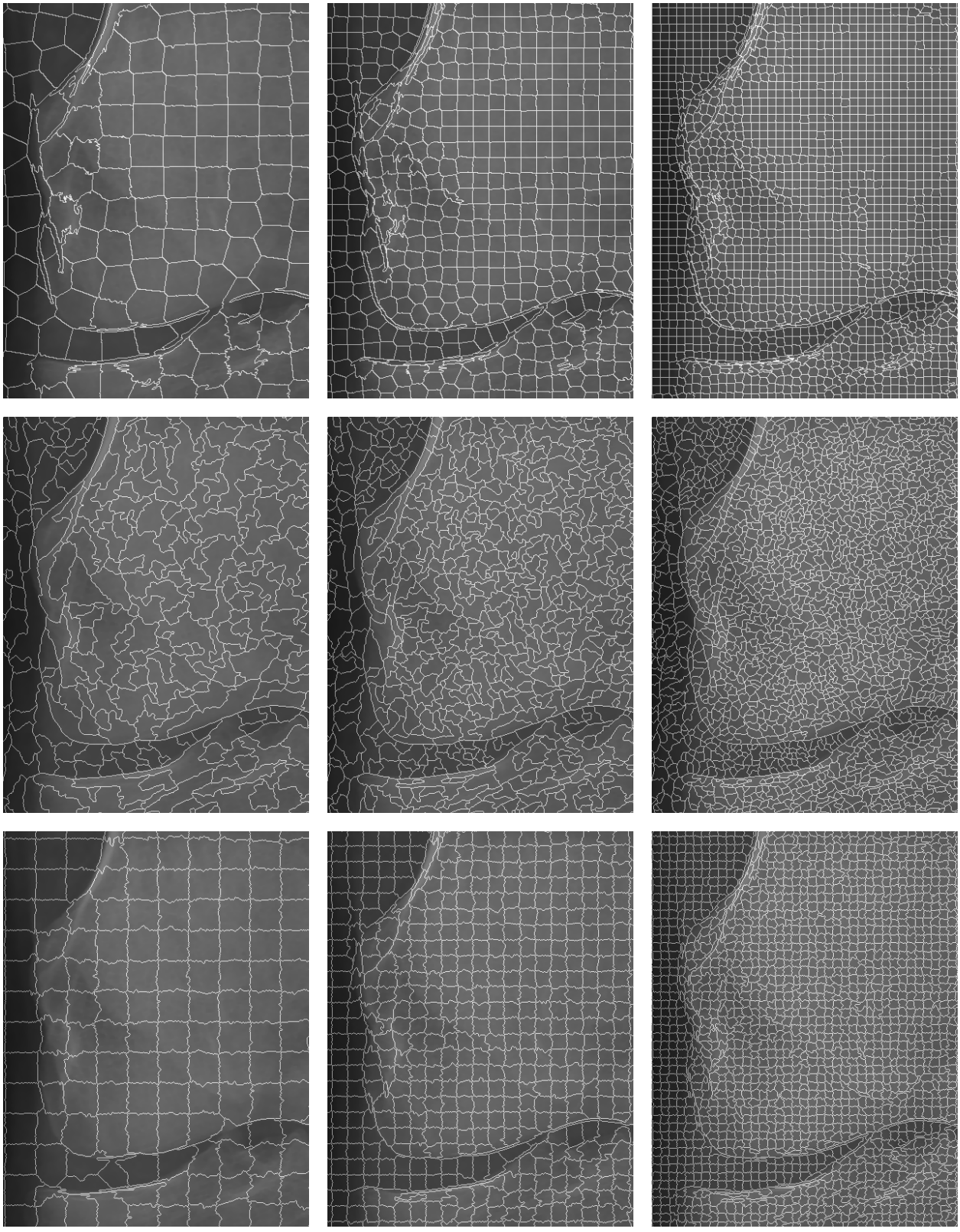


Figure 2.6: A comparison of preemptive SLIC (first row), watershed from markers (second row) and compact watershed (third tow) algorithms. Images are segmented into approximately 900 superpixels in the first column, 3400 in the second one and 12000 in the third one. Only a selected part of the image is shown

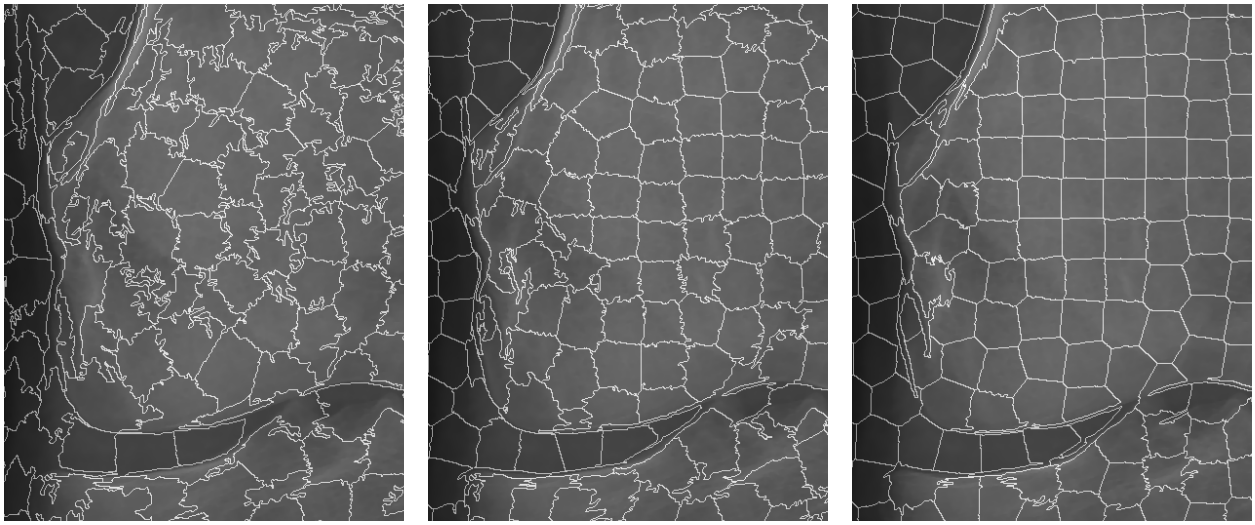


Figure 2.8: A comparison of preemptive SLIC segmentation for approximately 900 superpixels and compactness value equal to 5 for the left image, 10 for the middle image and 22 for the right image

Mohri et al., 2012, Murphy, 2012). The first of these tasks, classification, is relevant for this thesis. This Section is divided into two parts. In Section 2.5.1 the classification problem is first described. Well known algorithms are mentioned and briefly discussed. This is followed by a short introduction to feature extraction. Section 2.5.2 describes various methods of comparing performance of classification algorithms.

Classification and feature extraction

To formulate the classification problem we need to start with a training set of n pairs, $D = \{(x_i, y_i)\}_{i=1}^{n}$ where each x_i belongs to a certain set X and each y_i belongs to a finite set $Y = \{c_1, c_2, \ldots, c_l\}$ for some $l > 1$. The goal is to construct a function $h\colon X \to Y$ that appropriately returns labels $h(x) \in Y$ for each element x of the set X. To precisely specify this problem, a hypothesis space H is also needed, that is the set of functions h that are taken into consideration. A measure of accuracy A is also necessary to decide which functions fit the data better. In this way, classification can be expressed as an optimization problem:

$$h^* = \arg\max_{h \in H} A(h), \tag{2.26}$$

where the most accurate function h^* is searched. Different methods of measuring the accuracy are described in Section 2.5.2.

Sometimes, as used in Chapter 5, instead of simply ascribing a single label $c \in Y$ to each element of the set X, a probabilistic approach is taken. We assume that there exist a joint probability distribution[2]

[2]This definition assumes, for simplicity, a discrete set X. In general, a more complicated measure-theoretic definition should be used (Çınlar, 2011).

$P\colon X \times Y \to [0, 1]$. In this setting we expect the function h to return a probability distribution

$$h(x)_j = P(\boldsymbol{Y} = c_j | \boldsymbol{X} = x) \tag{2.27}$$

where $j = 1, 2, \ldots, l$ and bold symbols $\boldsymbol{X}$ and $\boldsymbol{Y}$ denote random variables with values in respective sets X and Y. Methods that solve this problem can be broadly categorized as belonging to one of the following classes (Murphy, 2012):

- Generative methods, which are based on modelling class-conditional probabilities $P(\boldsymbol{X}|\boldsymbol{Y} = c_j)$ and class priors $P(\boldsymbol{Y} = c_j)$, and thus have the capacity to generate new synthetic data points from the set X (Bishop, 2006). Among generative methods there are naive Bayes classifier (Russell and Norvig, 2009), linear discriminant analysis (Murphy, 2012) and the k-nearest neighbour classifier (Murphy, 2012).
- Discriminative methods, which model $h(x) = P(\boldsymbol{Y}|\boldsymbol{X} = x)$ directly. Most classification algorithms belong to this category (Murphy, 2012), for example decision tree-based methods (Cios et al., 2007), logistic regression or neural network-based classification algorithms (Dreiseitl and Ohno-Machado, 2002) and Support Vector Machine (Cortes and Vapnik, 1995, Russell and Norvig, 2009). Methods based on regression-voting (Gall and Lempitsky, 2009) are an important subcategory, although some authors consider them separately (Lindner et al., 2015, Bromiley et al., 2016).

Regression-voting is popular in image analysis, for methods based on variants of Constrained Local Models (Cristinacce and Cootes, 2008, 2006).

In many applications learning multiple classifiers and then aggregating their results is beneficial. This is the principle of the so-called ensemble learning, comprising techniques such as boosting (Schapire et al., 1998) or bagging (Breiman, 1996). Random forests (Breiman, 2001) constitute a prominent example of an ensemble classification method.

Feature extraction (Guyon et al., 2006) is an important tool in applying classification algorithms to practical problems. In many cases objects that need to be classified can be described by a vast amount of features or in different but equivalent ways. Appropriate and relevant features need to be selected and described in a way that facilitates effective object classification. Some authors even use a different formalism for the classification problem where classified objects and their features are considered separately (Cheng and Hüllermeier, 2009, Tong et al., 2004, Tomczyk and Szczepaniak, 2005). This highlights the issue of describing features of an instance but complicates the equations. It is worth noting that the problem of feature extraction can be automatically solved by very successful deep learning methods (LeCun et al., 2015, Krizhevsky et al., 2012), although they have limited applicability when the set of training images is small, as considered in this thesis.

Measuring performance of classification algorithms

There are many performance measures for classification algorithms. Firstly, one can be interested in time necessary to construct a classifier based on the training dataset or the time it takes to classify a single instance. Obviously, the fastest classifier would just return a single selected class. This is done by the ZeroR algorithm (Devasena et al., 2011). This method has, however, poor accuracy in most datasets and is thus used only as a baseline for other classifiers.

Typically, the accuracy is measured using a second set of pairs $\{(\hat{x}_i, \hat{y}_i)\}_{i=1}^{n_{test}}$, called a test set, by comparing $h(\hat{x}_i)$ and $\hat{y}_i$ for a given function $h \in H$ and $i \in \{1, 2, \ldots, n_{test}\}$. This approach, when applied directly, has several problems. In practice we have only a finite set of instances. When a fraction of them are used in the testing set and the rest forms the training set, the holdout cross-validation is performed (Kohavi, 1995). The drawback of this method is decreasing the potential of obtaining a more accurate classifier by not using a part of the data in training (Russell and Norvig, 2009).

The most popular tool for reliable assessment of classifier accuracy is cross-validation (Stone, 1974). The k-fold cross-validation divides the set of all learning instances into k sets D_j for $j = 1, 2, \ldots, k$ of approximately equal size. The analysed classifier is learned k times. In the jth iteration, $\bigcup_{l=1,2,\ldots,j-1,j+1,\ldots,k} D_l$ is used as the training set and D_j is used as the test set. Obtained accuracies are averaged to get the final accuracy estimate. The cross-validation is called stratified when, additionally, in each set D_j for $j = 1, 2, \ldots, k$ the distribution of classes is similar to their distribution in the entire set D. Kohavi (1995) suggests using stratified ten-fold cross-validation for model selection.

The accuracy estimation using cross-validation has an inherent statistical uncertainty. One method of dealing with this problem is construction of confidence intervals, that is we have to select two numbers a, b such that $P(a \leq A \leq b) \approx \gamma$, where A is a random variable describing accuracy of a classifier and $\gamma \in [0, 1]$ is a desired probability. The issue of formal definition of A is discussed by Vanwinckelen and Blockeel (2012).

This interval can be constructed in many ways, depending on assumptions about the performed cross-validation. In (Kohavi, 1995), the confidence interval is constructed based on the De Moivre-Laplace limit theorem. Another option is considered in (Moore and Lee, 1994, Maron and Moore, 1994) using Hoeffding's inequality, which has weaker assumptions about the random variable A. Different bounds are considered and compared in (Langford, 2005, Kääriäinen and Langford, 2005). Sometimes a box plot of the empirical distribution across cross-validation folds is given to illustrate the uncertainty (Sing et al., 2005).

Other statistical tools can also be used for measuring the accuracy, for example the bootstrap (Efron and Gong, 1983) and jackknife (Quenouille, 1949, Tukey, 1958, Miller, 1974) methods. They are, although, less popular than cross-validation.

There are also other tools for measuring performance of a classifier. Among the most popular ones are, for example:

- Receiver Operating Characteristic (ROC) curve, which is sometimes reduced to a single number, area under the ROC curve (AUC) (see (Pollack and Decker, 1958) for an early usage and (Bradley, 1997) for a more recent analysis),
- multi-class focused methods like Cohen's kappa (Cohen, 1960) and Fleiss kappa (Fleiss, 1971),
- single class focused methods, like true positive rate, false positive rate, precision, recall, specificity, sensitivity or F-measure (Japkowicz and Shah, 2011).

Statistical tests for measuring performance have been proposed (Japkowicz and Shah, 2011, Golland and Fischl, 2003). Their utility, however, should not be overvalued (Drummond, 2006, Demšar, 2008, Rosenthal, 1979). As machine learning is not the main focus of this thesis and applying advanced statistical methods is not a standard practice in image segmentation (Zhang, 1996, Zhang et al., 2008, Noble and Boukerroui, 2006), statistical tests are not performed.

Chapter 3

PGA-based boundary delineation

In this chapter a novel paradigm for fitting a statistical shape models to image data is described. This approach is influenced by achievements in the field of superpixels algorithms (Neubert and Protzel, 2014). A given image is divided into superpixels. Their borders can be represented as a multigraph in which edges are candidate boundary patches.

The PDM-based search for an object boundaries is generally not constrained (except constraints on the PCA modes within which a reasonable shape model is fitted). In contrast, the method presented in this chapter performs the search for the boundary with an additional constraint that it must lie within the boundaries generated by a superpixel algorithm. As far as I know models considering edges as hard constraints for combinatorial fitting of a trainable statistical model of shape were not previously considered. As a side effect, the problem of detecting boundaries of an object can be formulated as a global search. This results in reduced dependency on good initialization. In contrast, in PDM-based methods the final success of the boundary detection does depend on both landmark selection and initialization.

Methods

The PGA-based binarization algorithm consists of two phases of operation, training of the statistical shape model and matching against the trained model. In the first stage the algorithm is given a number of curves $\gamma^i \colon [0,1] \to \mathbb{C}$, $i = 1, 2, \ldots, M$ representing boundaries of object instances. A continuous image description is assumed, and the curve lies in $\mathbb{C}$ instead of $\mathbb{R}^2$ for more simple expression of certain operations. The standard vector space isomorphism between these two spaces is used where necessary. The curves γ_i are either extracted from hand-annotated training images or automatically generated for computer-generated images.

During the second stage the constructed shape model is used to guide boundary delineation in given images. An algorithm iteratively elongates partially matched shapes and prunes the set of curve fragments

using a combination of parameters related to either fitness to the model or certain cues extracted from the analysed image.

Training

Training begins with sampling each training curve $\gamma^i\colon [0,1] \to \mathbb{C}$, $i = 1,2,\ldots,M$ at n points $\{p_1^i, p_2^i, \ldots, p_n^i\}$, where n is a selected fixed number. The points are equidistant with respect to the Euclidean metric in $\mathbb{C}$, that is

$$|p_2^i - p_1^i| = |p_3^i - p_2^i| = \cdots = |p_n^i - p_{n-1}^i|. \tag{3.1}$$

In practice, curves γ^i are represented by sequences of neighbouring pixels and the points $\{p_1^i, p_2^i, \ldots, p_n^i\}$ for $i = 1,2,\ldots,M$ are obtained using first-order interpolation.

In the next step the Ordinary Procrustean Matching (OPA) (Kendall, 1984, Gower, 1975, Goodall, 1991) of sampled curves is performed with $\{p_j^1\}_{j=1}^n$ selected as the reference. Points adjusted using the OPA are denoted by $\overline{p}_j^i$ for $i = 1,2,\ldots,M$ and $j = 1,2,\ldots,n$. Two sets of features are extracted from the adjusted points $\overline{p}_j^i$:

1. The angles $\alpha_{j,0}^i = \arg(\overline{p}_{j+1}^i - \overline{p}_j^i)$ where $j = 1,2,\ldots,n$ and $\overline{p}_{n+1}^i = \overline{p}_1^i$ is assumed.
2. The angles $\alpha_{j,k}^i = \arg(\overline{p}_{(j+1)2^k}^i - \overline{p}_{j\cdot 2^k}^i)$ where $j = 1,2,\ldots,\lfloor n/2^k \rfloor$, $k = 1,2,\ldots,\lceil \log_2(n) - 1 \rceil$ and $\overline{p}_{(j+1)2^k}^i = \overline{p}_{(j+1)2^k - n}^i$ is assumed where necessary.

The second set of features, however redundant, increases the robustness of shape representation in a way similar to a multi-scale approach.

A similar point distribution model based on angles has been discussed in e.g. (Sommer et al., 2009). The difference is that in that paper the second set of angles is not considered and a different approach to scale and rotation-invariance is taken, resulting in a different Riemannian metric in the feature space.

The last step of training involves performing of the PGA (see Section 2.3), in the feature space. This is necessary as the extracted features belong to a torus $\mathrm{T}^{dim(n)}$ (that is, a Cartesian product of $dim(n)$ circles) where

$$dim(n) = \sum_{b=0}^{\lceil \log_2(n)-1 \rceil} \left\lfloor n/2^b \right\rfloor. \tag{3.2}$$

The mean shape, denoted μ,

$$\begin{aligned} \mu = (&\mu_{1,0}, \mu_{2,0}, \ldots, \mu_{n,0}, \\ &\mu_{1,1}, \ldots, \mu_{\lfloor n/2 \rfloor,1}, \\ &\ldots, \\ &\mu_{1,\lceil \log_2(n)-1 \rceil}, \ldots, \mu_{\lfloor n/2^{\lceil \log_2(n)-1 \rceil} \rfloor, \lceil \log_2(n)-1 \rceil}\Big) \end{aligned} \tag{3.3}$$

is calculated using the extrinsic mean algorithm (see Equation (2.23)) (Fletcher et al., 2004) in $\mathbb{C}^{dim(n)}$. The torus is embedded as a product of unit circles and the projection is based on normalization in each complex subspace.

Next, the PGA modes $w_i \in T_\mu \mathrm{T}^{dim(n)}$ (where $T_\mu \mathrm{T}^{dim(n)}$ is the space tangent to $\mathrm{T}^{dim(n)}$ at point μ, see Definition 14) and variances $\lambda_i \in \mathbb{R}$ for $i = 1, 2, \ldots, dim(n)$ are computed. To ensure the rotation invariance of the shape model, the mean shape is rotated by a torus diffeomorphism $\varphi\colon \mathrm{T}^{dim(n)} \to \mathrm{T}^{dim(n)}$ given by equation

$$\begin{aligned} \varphi\left(\gamma_{1,0}, \ldots, \gamma_{\lfloor n/2^{\lceil \log_2(n)-1 \rceil} \rfloor, \lceil \log_2(n)-1 \rceil}\right) = \\ \left(\gamma_{1,0} - \mu_{1,0}, \ldots, \gamma_{\lfloor n/2^{\lceil \log_2(n)-1 \rceil} \rfloor, \lceil \log_2(n)-1 \rceil} - \mu_{1,0}\right). \end{aligned} \tag{3.4}$$

The rotated mean (note that rotation corresponds to translation in the angular representation) is hereafter called $\overline{\mu}$:

$$\overline{\mu} = \varphi(\mu) \tag{3.5}$$

and rotated PGA modes are called $\overline{w}_i$:

$$\overline{w}_i = \mathrm{d}\varphi_\mu(w_i) \tag{3.6}$$

for $i = 1, 2, \ldots, dim(n)$ where $\mathrm{d}\varphi_\mu$ is the differential of φ at point μ (see Definition 18 on page 12). Finally, the highest-energy modes $\overline{w}_1, \overline{w}_2, \ldots, \overline{w}_L$ are selected. The number $1 \leq L \leq M$ is the smallest number such that tangent vectors $\overline{w}_1, \overline{w}_2, \ldots, \overline{w}_L$ correspond to at least E_{min} percent of total energy ($0 \leq E_{min} \leq 100$), as expressed by variances.

This approach is partially similar to Kendall's shape manifolds (Kendall, 1989). Due to the existence of additional constraints on the positions of points, defined by Equation (3.1), a lower-dimensional manifold is obtained.

Matching

Boundary delineation of a given discrete image $I_{mat}\colon \{1, 2, \ldots, W\} \times \{1, 2, \ldots, H\} \to \mathbb{R}$ (see Section 2.1.5) starts with edge detection. A number of existing approaches can be used (Torre and Poggio, 1986). For images with low level of noise the gradient magnitude $I_{gm}\colon \{1, 2, \ldots, W\} \times \{1, 2, \ldots, H\} \to \mathbb{R}$ is sufficient and fast to compute (Jähne, 2002). Central first-order finite difference formulas are used for all pixels except borders where one-sided first order formulas need to be applied. The image I_{gm} is oversegmented to obtain constraints for matching the statistical model of shape. Any superpixel algorithm (see Section 2.4.4) can be used, including watershed segmentation (see Section 2.4.2). The watershed from markers (see Section 2.4.2) algorithm was used since it exhibits very good boundary adherence on test datasets.

The oversegmentation results in a set of pixels $V \subset \{1, 2, \ldots, W\} \times \{1, 2, \ldots, H\}$, represented by their indices in the image, that form the borders between segments. This set can be considered as the set of vertices of an undirected graph $G = (V, E)$ where the set of edges E is the set of unordered pairs of Moore-neighbouring pixels in V, that is pixels whose column and row indices are different by at most 1. Now, delineation of boundaries of an object is reduced to finding a path in the graph G.

The graph G is then reduced to a multigraph $G_{red} = (V_{red}, E_{red}, r_{red})$. Let $G' = (V', E')$ be a subgraph (Definition 5) of G that contains only vertices v' of degree 3 or more and vertices on the edges of image ($v' \in V$ such that $v' = (1, j)$, $v' = (W, j)$, $v' = (i, 1)$ or $v' = (i, H)$ for some $i \in \{1, 2, \ldots, W\}$ and $j \in \{1, 2, \ldots, H\}$). The set of edges E' contains all edges from G that connect vertices from the set V'. Then V_{red} is the set of all connected components of the graph G' and the function $\kappa_V \colon V' \to V_{red}$ assigns to vertex $v \in V'$ the connected component in V' it belongs to. Vertices of G_{red} correspond to places in the segmented image where either three or more superpixels have a common boundary or two superpixels meet at the image boundary. The set of edges E_{red} contains all paths and cycles $v_1 v_2 \ldots v_k, k \in \mathbb{N}, k > 1$ such that v_1 and v_k belong to connected components of G', vertices v_i for $i = 2, 3, \ldots, k-1$ belong to V and are of degree 2 and $\{v_i, v_{i+1}\} \in E$ for $i = 1, 2, \ldots, k-1$. The partial function $\kappa_E \colon E \to E_{red}$ assigns to the edge $e \in E$ the edge $\hat{e} \in E_{red}$ that contains e (two consecutive vertices in the path or cycle $\hat{e}$ are connected by e in G). Finally, $r_{red} \colon E_{red} \to \{\{\hat{v}_1, \hat{v}_2\} \colon \hat{v}_1, \hat{v}_2 \in V_{red}\}$ is a function that transforms an edge $v_1 v_2 \ldots v_k$ in G_{red} to the set of connected components its endpoints belong to, that is $\{\kappa_V(v_1), \kappa_V(v_k)\}$. Figure 3.1 contains an example of graph reduction.

In the next step a path or cycle in the multigraph G_{red} that matches the model best is searched. This process is similar to the Generalized Hough Transform (Ballard, 1981) pattern matching. The key component of the parametric family of curves is expressed by the exponential map (see Definition 24) at point $\overline{\mu}$ of a linear combination of PGA modes $\exp_{\overline{\mu}}\left(\sum_{i=1}^{L} c_i \overline{w}_i\right)$ where $c_i \in \mathbb{R}$ for $i = 1, 2, \ldots, L$. Additionally, four parameters are needed to account for translation, rotation and uniform scaling. This provides the full description of a parametrized curve that can be matched using Generalized Hough Transform (GHT) (Ballard, 1981). The key difference between GHT and the approach proposed in this chapter is a new way of searching the parameter space for optimal matching and a different method of selecting the best match. A standard exhaustive search of GHT is slow and not robust enough for curves with more than a few parameters (Mukhopadhyay and Chaudhuri, 2015).

To facilitate the description of the process of pattern matching, the following definition is necessary.

Definition 28. *A **multitrail** t in a multigraph $G_{red} = (V_{red}, E_{red}, r_{red})$ is one of the following:*

1. *an empty set $\{\}$, which is called an **empty multitrail**,*
2. *a pair $(\hat{v}, (\hat{e}_1, \ldots, \hat{e}_f))$ where $\hat{v} \in V_{red}$ is the first vertex and $(\hat{e}_1, \ldots, \hat{e}_f)$ is a (possibly empty) sequence of edges in G_{red} such that $f \in \mathbb{N}$, $\hat{e}_i \in E_{red}$ for $i = 1, 2, \ldots, f$, $r_{red}(\hat{e}_1) = \{\hat{v}, \hat{v}_1\}$,*

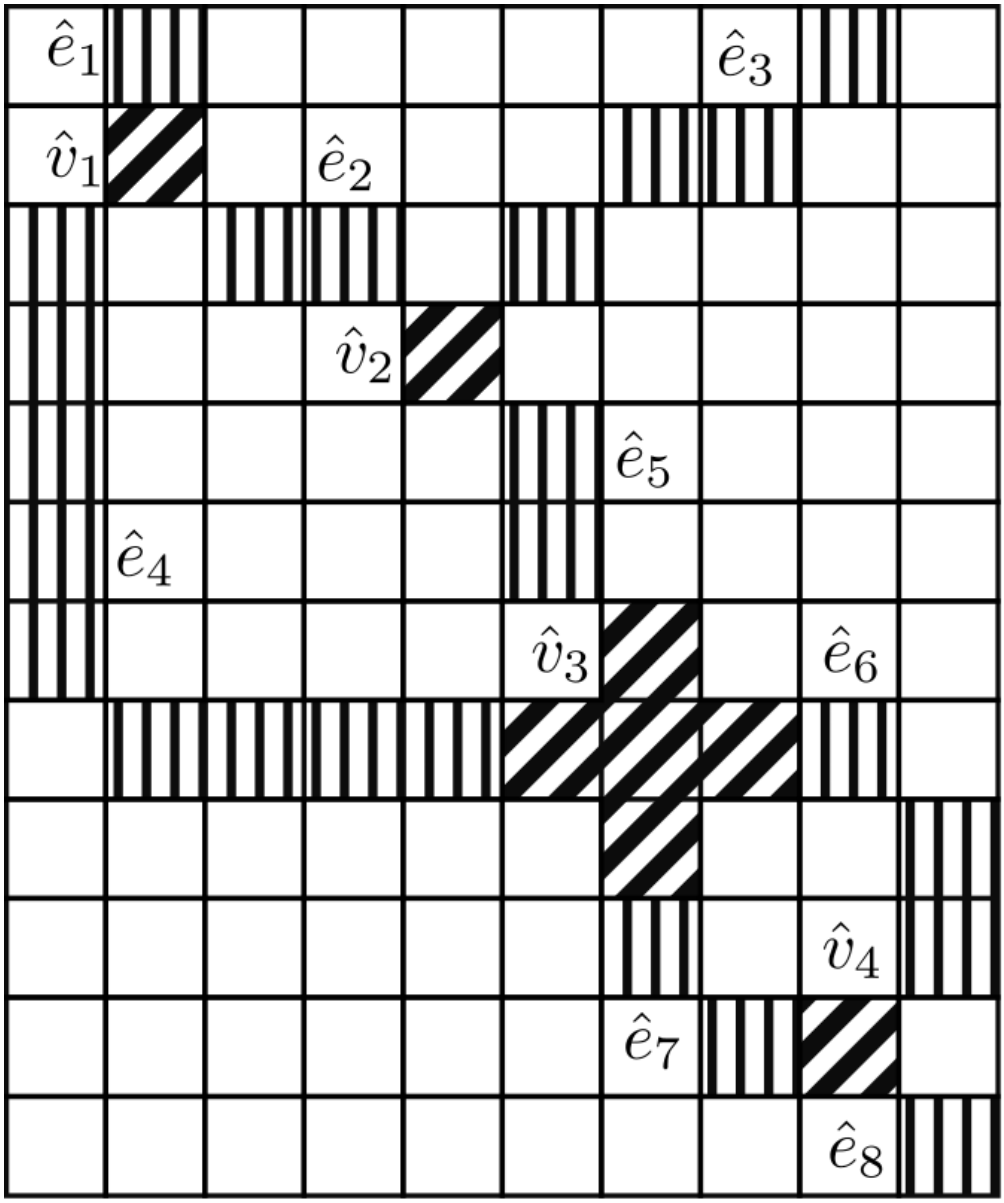


Figure 3.1: A sample fragment of a graph G. Each square corresponds to a single pixel. This part of the graph G contains four vertices of its reduced graph ($\hat{v}_1$ to $\hat{v}_4$, marked by thick slanted stripes) and eight of its edges ($\hat{e}_1$ to $\hat{e}_8$, marked by thin vertical stripes). Two of the vertices ($\hat{v}_3$ and $\hat{v}_4$) are connected by two edges ($\hat{e}_6$ and $\hat{e}_7$).

$r_{red}(\hat{e}_2) = \{\hat{v}_1, \hat{v}_2\}$, ..., $r_{red}(\hat{e}_f) = \{\hat{v}_{f-1}, \hat{v}_f\}$ and $\hat{v}, \hat{v}_1, \ldots, \hat{v}_f$ are pairwise different, except $\hat{v}$ might be equal to $\hat{v}_f$. The sequence $(\hat{v}, \hat{v}_1, \ldots, \hat{v}_f)$ is called the ***vertex sequence*** *of t and is denoted by $\hat{V}(t)$.*

The algorithm takes an initial multitrail t_0 as an input. The multitrail t_0 may be empty. In such case a global search will be performed. Otherwise, the given multitrail will be a part of all considered paths.

Pattern matching is performed iteratively. There are $n - |t_0|$ iterations, where n is the number of points the training curves were sampled at and $|t_0|$ is the number of vertices in the multitrail t_0. In each of them a set of elongated multitrails P_i is constructed as follows:

$$\begin{aligned} P_i = \{(\hat{v}, (\hat{e}_1, \ldots, \hat{e}_{k-1}, \hat{e}_k)) \colon (\hat{v}, (\hat{e}_1, \ldots, \hat{e}_{k-1})) \in \overline{P}_{i-1}, \\ \hat{e}_k \in E_{red} \text{ and } (\hat{v}, (\hat{e}_1, \ldots, \hat{e}_{k-1}, \hat{e}_k)) \text{ is a multitrail}\} \end{aligned} \tag{3.7}$$

when $i > 0$. The set P_0 is equal to $\{(t_0, ())\}$ for non-empty t_0 and $\{(\hat{v}, ()) \colon \hat{v} \in V_{red}\}$ otherwise. In practice the set P_i is constructed by considering each multitrail t from the set P_{i-1}, selecting edges $\hat{e}_k$ incident to the last vertex of the vertex sequence of t and checking if t elongated with $\hat{e}_k$ is a multitrail.

Let $\overline{P}_{i-1}$ be a subset of P_{i-1} with n_{sel} multitrails of lowest error $\hat{E}_{total}(t, center(\kappa_V^{-1}(\hat{v})), \alpha_0)$ (see Equation (3.24)), where $center \colon 2^V \to \mathbb{C}$ is a function that calculates the arithmetic mean of positions of nodes from a given set, $\hat{v}$ is the first vertex of t and α_0 is the shape rotation:

$$\alpha_0 = \arg \min_{\alpha \in (-\pi, \pi]} \hat{E}_{total}(t, x_0, \alpha). \tag{3.8}$$

Optimal rotation α_0 is recalculated when the step number i is a power of 2. The computed value is retained for subsequent steps. For best results α_0 should be recalculated after each step but this procedure is computationally expensive. This optimization follows from an observation that curves change much more significantly after elongation when they are short. The final match is the multitrail $t \in \bigcup_{i \in \{0,1,\ldots,n-|p_0|\}} P_i$ with the lowest total error $E_{total}(t, n, x_0, \alpha_0)$ (see Equation 3.21). Pseudocode for this algorithm is presented in Algorithm 2 and demonstration of its operation is shown in Figure 3.2.

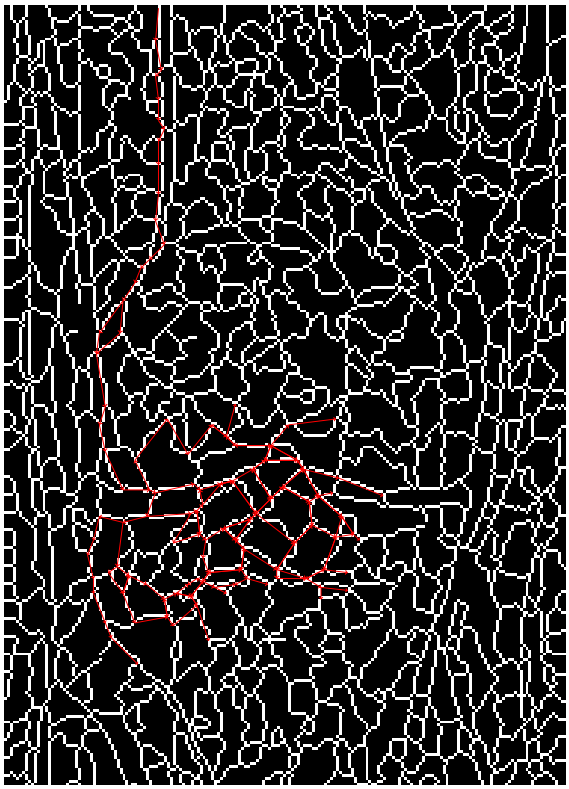
Estimation of the multitrail error

A number of different methods of multitrail error estimation has been developed. Each of them either measures a different kind of discrepancy from the model or evaluates fitness to the given image. In this section they are described in detail. A way to combine them into a single error function is also proposed.

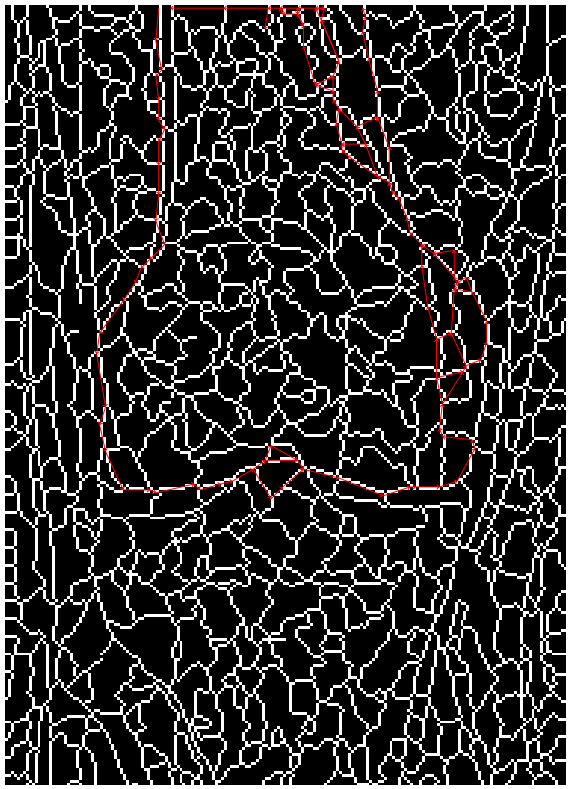
A submodel extraction procedure is used when a multitrail is compared against a part of the entire modelled curve. This method describes how a statistical model of initial fragments of curves can be extracted from the model based on entire curves.

(a) An input image

(b) Gradient magnitude of the input image 3.2a



(c) Superpixel segmentation with partial paths after 30 iterations



(d) Superpixel segmentation with partial paths after 64 iterations

Figure 3.2: Demonstration of operation of PGA-based boundary delineation algorithm. Small red arrows indicate edges of the reduced graph G_{red} belonging to paths in the set *multitrails* of Algorithm 2.

Algorithm 2: Searching for the multitrail with the lowest error in the reduced graph G_{red}.

Require: $(\overline{\mu}, \{\overline{w}_1, \overline{w}_2, \ldots, \overline{w}_L\})$ is the trained model, t_0 is the initial multitrail

```
multitrails ← {t_0} {a working set of multitrails}
fullPaths ← {} {set of best multitrails matched to full model}
for i ← 1 to n − |t_0| do
    if i is a power of 2 then
        for all t ∈ multitrails do
            RecalculateInitialAngle(t)
        end for
    end if
    multitrails ← SelectBestPaths(multitrails)
    fullPaths ← PickBestFullPaths(multitrails, fullPaths)
    multitrails ← ElongatePaths(multitrails)
end for
return GetLowestErrorPath(fullPaths)
```

Submodel extraction

Let $(\overline{\mu}, \{w_1, w_2, \ldots, w_L\})$ be the model corresponding to the set of shape samples $\{p_1^i, p_2^i, \ldots, p_n^i\}$, $i = 1, 2, \ldots, M$. During the construction of the full multitrail it is necessary to estimate the error for multitrails corresponding to just a part of the full model. Let n_{part} be the number of points the full multitrail needs to be truncated to. Now, we need to know what model corresponds to samples $\{p_1^i, p_2^i, \ldots, p_{n_{part}}^i\}$. In general this model has to be independently constructed but this can be quite costly, especially when n is large. Instead, an approximation to the exact truncated model can be extracted from the full model. This results in reduction of memory complexity of model representation from $O(n^3)$ to $O(n^2)$.

Let $Sub(\mathrm{T}^{dim(n)}, n_{part})$ be the submanifold of $\mathrm{T}^{dim(n)}$ such that

$$
\begin{aligned}
&Sub(\mathrm{T}^{dim(n)}, n_{part}) \cong \\
&\Big\{\Big(\beta_{1,0}, \beta_{2,0}, \ldots \beta_{n_{part},0}, \overline{\mu}_{n_{part}+1,0}, \ldots, \overline{\mu}_{n,0}, \\
&\beta_{1,1}, \beta_{2,1}, \ldots \beta_{n_{part},1}, \overline{\mu}_{n_{part}+1,1}, \ldots, \overline{\mu}_{\lfloor n/2 \rfloor,1}, \\
&\ldots \\
&\beta_{1,\lceil \log_2(n)-1 \rceil}, \ldots, \beta_{n_{part},\lceil \log_2(n)-1 \rceil}, \overline{\mu}_{n_{part}+1,\lceil \log_2(n)-1 \rceil}, \ldots, \\
&\qquad \overline{\mu}_{\lfloor n/2^{\lceil \log_2(n)-1 \rceil} \rfloor,\lceil \log_2(n)-1 \rceil}\Big) : \beta_{j,k} \in (-\pi, \pi], \\
&j = 1, 2, \ldots, \left\lfloor n/2^k \right\rfloor; k = 0, 1, 2, \ldots, \lceil \log_2(n) - 1 \rceil \Big\} \\
&\subseteq \mathrm{T}^{dim(n)}.
\end{aligned}
\tag{3.9}
$$

The truncated model is then given by

$$(\overline{\mu}, \{P_{T_{\overline{\mu}}\mathit{Sub}(\mathrm{T}^{\mathit{dim}(n)}, n_{part})}\overline{w}_1, P_{T_{\overline{\mu}}\mathit{Sub}(\mathrm{T}^{\mathit{dim}(n)}, n_{part})}\overline{w}_2, \ldots, P_{T_{\overline{\mu}}\mathit{Sub}(\mathrm{T}^{\mathit{dim}(n)}, n_{part})}\overline{w}_L\}) \tag{3.10}$$

where $P_{T_{\overline{\mu}}\mathit{Sub}(\mathrm{T}^{\mathit{dim}(n)}, n_{part})}$ denotes the orthogonal projection onto space tangent to submanifold $\mathit{Sub}(\mathrm{T}^{\mathit{dim}(n)}, n_{part})$ at point $\overline{\mu}$.

Eigenface-like error

The first error estimator is inspired by the eigenface method (Sirovich and Kirby, 1987). The idea of pattern matching based on shape alone is also present in the Active Shape Model approach (Cootes et al., 1995). The description below is given for the case of matching the given multitrail t in multigraph G_{red} to the full model. The procedure for matching against a submodel is analogous.

The multitrail $t = (\hat{v}, (\hat{e}_1\hat{e}_2 \ldots \hat{e}_F))$, where F is the number of edges in t, is first converted into a curve $\Xi(t)$ – a polygonal chain defined by the following points:

$$(\mathit{center}(\kappa_V^{-1}(\hat{v})), v_{1,1}, v_{1,2}, \ldots, v_{1,J_1}, \mathit{center}(\kappa_V^{-1}(\hat{v}_1)), \ldots, v_{F,1}, v_{F,2}, \ldots, v_{f,J_F}, \mathit{center}(\kappa_V^{-1}(\hat{v}_F))) \tag{3.11}$$

where $(\hat{v}, \hat{v}_1, \ldots, \hat{v}_F) = \hat{V}(t)$ is the vertex sequence of t and $\mathit{center}(\kappa_V^{-1}(\hat{v}_i))$ is the coordinate-wise arithmetic mean of coordinates of nodes in the set $\kappa_V^{-1}(\hat{v}_i)$ for $i = 1, 2, \ldots, F$. Furthermore, the sequence $v_{i,1}, v_{i,2}, \ldots, v_{i,J_i}$ is equal to $\kappa_E^{-1}(\hat{e}_i)$ for $i = 1, 2, \ldots, F$. The order of vertices in this sequence is determined by their order on the edge $\hat{e}_i$, that is $v_{i,1}$ is the vertex closest to its predecessor in the sequence defined in Equation (3.11) and v_{i,j_i-1} and v_{i,j_i+1} are the only neighbours of v_{i,j_i} in the eight pixel neighbourhood for $j_i = 2, 3, \ldots, J_i - 1$, $i = 1, 2, \ldots, F$.

The curve $\Xi(t)$ is then sampled at n equidistant points

$$z(\Xi(t)) = (z_1, z_2, \ldots, z_n) \tag{3.12}$$

where $z\colon C \to \mathbb{C}^n$ is the sampling function and C is the set of polygonal chains in $\mathbb{C}$. The points are converted into a collection of angles like in Section 3.1.1 and the first angle in subtracted like in Equation (3.5). As a result, a point in $\omega(z(\Xi(t))) \in \mathrm{T}^{dim(n)}$ is obtained, where $\omega\colon \mathbb{C}^n \to \mathrm{T}^{dim(n)}$ performs the described conversion.

The point $\omega(z(\Xi(t)))$ is transferred into the tangent space $T_{\overline{\mu}}\mathrm{T}^{\mathit{dim(n)}}$ using the logarithmic map (the inverse of the exponential map) $\log_{\overline{\mu}}\colon \mathrm{T}^{\mathit{dim(n)}} \to T_{\overline{\mu}}\mathrm{T}^{\mathit{dim(n)}}$. The result of this operation is then compared to

the model and the error is calculated:

$$E_{eig}(t, n_{part}) = \|(I - UU^T)(\log_{\overline{\mu}}(\omega(z(\Xi(t))))\|_2 \tag{3.13}$$

where n_{part} is the number of points in the employed submodel $Sub(\mathrm{T}^{dim(n)}, n_{part})$ (used in the general case) and $U = [w_1, w_2, \ldots, w_L]$ is the matrix composed of coordinates of selected orthonormal eigenvectors and $(\log_{\overline{\mu}}(\omega(z(\Xi(t)))))$ is the column matrix of coordinates of that vector. Both matrices of coordinates are computed in the same orthonornal basis $B_{\overline{\mu}}$ composed from basis vectors from spaces tangent to the submanifold

$$\begin{aligned} S_{\overline{\mu};j,k} = \big\{ \big(&\overline{\mu}_{1,0}, \ldots, \overline{\mu}_{n,0}, \ldots, \overline{\mu}_{j-1,k}, \beta_{j,k}, \overline{\mu}_{j+1,k}, \ldots, \\ &\overline{\mu}_{1,\lceil \log_2(n)-1 \rceil}, \ldots, \overline{\mu}_{\lfloor n/2^{\lceil \log_2(n)-1 \rceil} \rfloor, \lceil \log_2(n)-1 \rceil} \big) \\ &: \beta_{j,k} \in (-\pi, \pi] \big\} \subseteq \mathrm{T}^{dim(n)} \end{aligned} \tag{3.14}$$

at $\overline{\mu}$ for $j = 1, 2, \ldots, \lfloor n/2^k \rfloor$, $k = 0, 1, 2, \ldots, \lceil \log_2(n) - 1 \rceil$.

In the general case of matching against a submodel, a few changes are necessary. First, the curve is sampled at a lower number of points (n_{part}). Second, next steps after sampling are performed in a submanifold $Sub(\mathrm{T}^{dim(n)}, n_{part})$. Finally, the martix U is constructed from vectors w_i, $i = 1, 2, \ldots, L$ truncated to the tangent space of the submanifold $Sub(\mathrm{T}^{dim(n)}, n_{part})$.

High deviation error

In the PCA (and therefore PGA) method the eigenvalues λ are the empirical variances. It is possible that a certain vector lies in the PCA subspace but much further from the mean than the training vectors. Such a vector does not fit the data well despite its low eigenface error. Therefore the high deviation error term is introduced to reject such vectors.

The high deviation error is defined as:

$$E_{hde}(t) = \sqrt{\frac{\sum_{i=1}^{L} \lambda_i}{L}} \sum_{i=1}^{L} \frac{c_i}{\sqrt{\lambda_i}} \tag{3.15}$$

where $c_i = \langle \log_{\overline{\mu}}(\omega(z(\Xi(t)))), \overline{w}_i \rangle$ and L is the number of selected eigenvectors. The factor appearing in the Equation (3.15) before the sum is introduced to normalize the high deviation error.

Gradient error

The information carried by the image I_{mat} can be used in one more way, apart from analysis of superpixel segmentation of gradient magnitude image. It can be observed that the training shapes cover areas of training

images where the magnitude of gradient is relatively high. This property is exploited to define the gradient error term:

$$E_{grad}(t) = \exp\left(\frac{-1}{a_{grad}|\Xi(t)|}\sum_{i=1}^{|\Xi(t)|}\left\|\nabla I_{mat}\left(\lfloor q_{i,1}\rfloor, \lfloor q_{i,2}\rfloor\right)\right\|_2\right) \tag{3.16}$$

where $\Xi(t)$, defined by Equation (3.11), is equal to $(q_1, q_2, \ldots, q_{|\Xi(t)|})$ and a_{grad} is a gradient magnitude rescaling constant. Terms $q_{i,1}$ and $q_{i,2}$ are coordinates of q_i for $i = 1, 2, \ldots, |\Xi(t)|$. First order central finite difference formulas are used to compute the gradient.

The value of a_{grad} has been determined experimentally and fitted all of the tested images. In the future a more robust way of determining its value should be explored, such as using parameter optimization frameworks (Hutter et al., 2011).

Scale error

In certain cases it has been observed that the described algorithm has a tendency to overestimate the scale of the curve. It is especially likely when the initial part of the curve is approximately a straight line. The algorithm sometimes fails to correct this overestimation later while matching against longer submodels. To overcome this tendency a scale error term is introduced:

$$E_{scale}(t) = \frac{1}{n-1}\sum_{i=1}^{n-1}|z(\Xi(t))_{i+1} - z(\Xi(t))_i|. \tag{3.17}$$

In an ideal case, the expression in Equation (3.17) is equivalent to $E_{scale}(t) = |z(\Xi(t))_2 - z(\Xi(t))_1|$. In practice, though, the points $z(\Xi(t))$ are only approximately equidistant and Equation (3.17) gives a result that is more closely related to the actual scale of the shape represented by t.

Curve distance error

Let us assume for the moment that a multitrail t is matched to the full model at an initial angle α_0, initial position x_0 and scale d. The curve distance error is the distance between curves $\Xi(t)$ and the polygonal chain defined by points $\overline{\omega}(\beta_{1,0}, \ldots, \beta_{n-1,0}, x_0, d, \alpha_0)$ where $\beta_{i,0}$ for $i = 1, 2, \ldots, n-1$ is defined by

$$\begin{gathered}\left(\beta_{1,0}, \ldots, \beta_{\lfloor n/2^{\lceil \log_2(n)-1\rceil}\rfloor, \lceil \log_2(n)-1\rceil}\right) = \\ \exp_{\overline{\mu}}(P_{\mathrm{Span}\{\overline{w}_1, \overline{w}_2, \ldots, \overline{w}_L\}} \log_{\overline{\mu}}(\omega(z(\Xi(t)))))\end{gathered} \tag{3.18}$$

and the function $\overline{\omega}$ is given by the following formula:

$$\overline{\omega}(\beta_{1,0},\ldots,\beta_{n-1,0},x_0,d,\alpha_0) = \left(x_0, x_0 + de^{i(\alpha_0+\beta_{1,0})},\ldots,x_0 + d\left(\sum_{i=1}^{n-1} e^{i(\alpha_0+\beta_{i,0})}\right)\right). \tag{3.19}$$

Using these symbols the curve distance error can be defined by

$$E_{cde}(t,n_{part},x_0,\alpha_0) = 1/n \left\| z(\Xi(t)) - \overline{\omega}(\beta_{1,0},\ldots,\beta_{n-1,0},x_0,d,\alpha_0)\right\|_1. \tag{3.20}$$

Curve distance model for partial matches (that is, while matching against a submodel) can be obtained by truncating the full model and using the formulas for the extracted submodel (see Section 3.1.3, Submodel extraction subsection). Necessary changes are analogous to adaptations described for eigenface-like error (see Section 3.1.3, Eigeface-like error subsection).

Total error

The total error is a linear combination of eigenface error, high deviation error, gradient error, scale error and curve distance error multiplied by an occlusion factor $Occl(t)$:

$$\begin{aligned} E_{total}(t,n_{part},x_0,\alpha_0) = Occl(t)\,(e_{eig}E_{eig}(t,n_{part}) + \\ e_{cde}E_{cde}(t,n_{part},x_0,\alpha_0) + e_{grad}E_{grad}(t) + \\ e_{scale}\,E_{scale}(t) + e_{hde}E_{hde}(t)) \end{aligned} \tag{3.21}$$

where e_{eig}, e_{hde}, e_{grad}, e_{scale} and e_{cde} are nonnegative coefficients of the linear combination.

The occlusion factor is introduced to help reject model matchings where the model curve is not fully contained within the image I_{mat}. An expected continuation of the model fitted to the multitrail t is calculated by taking the dot products $c_i = \langle P_{T_{\overline{\mu}}Sub(\mathrm{T}^{dim(n)},n_{part})}\overline{w}_i, \log_{\overline{\mu}}(\omega(p))\rangle$, $i = 1,2,\ldots,L$ in the appropriate submodel $Sub(\mathrm{T}^{dim(n)},n_{part})$ and using them in the full model:

$$\left(\beta_{1,0},\ldots,\beta_{\lfloor n/2^{\lceil\log_2(n)-1\rceil}\rfloor,\lceil\log_2(n)-1\rceil}\right) = \exp_{\overline{\mu}}\left(\sum_{i=1}^{L} c_i\overline{w}_i\right) \tag{3.22}$$

(see Equation (3.18)). The occlusion factor is then defined as an exponential of fraction of points $W(t)$ in the multitrail given by $\overline{\omega}(\beta_{1,0},\ldots,\beta_{n-1,0},x_0,d,\alpha_0)$ that lie outside of the image I_{mat}:

$$Occl(t) = e^{a_{occl}W(t)} \tag{3.23}$$

where a_{occl} is an occlusion error coefficient.

The minimization of total error over n_{part} between $n_{min}(t)$ and $n_{max}(t)$ results in the submanifold-minimized total error:

$$\begin{aligned} &\hat{E}_{total}(t, x_0, \alpha_0) = \\ \arg &\min_{i=n_{min}(t), n_{min}(t)+1, \ldots, n_{max}(t)} E_{total}(t, i, x_0, \alpha_0). \end{aligned} \tag{3.24}$$

The following formulas are used for the functions n_{min} and n_{max}:

$$\begin{aligned} n_{min}(t) &= \max\left\{2, \frac{l(t)n}{(W+H)s_{min}}, n_{last}(t) - c_{min}n\right\} \\ n_{max}(t) &= \min\left\{n, \frac{l(t)n}{(W+H)s_{max}}, n_{last}(t) + c_{max}n\right\} \end{aligned} \tag{3.25}$$

where $l(t)$ is the length of the polygonal chain described by Equation (3.11), $n_{last}(t)$ is the number i that minimized the error in Equation (3.24) before the last elongation of t (or zero if $t \in P_0$) and s_{min}, s_{max}, c_{min} and c_{max} are certain constants that depend on the dataset.

Experiments

The algorithm was tested on four datasets: two medical sets (50 standing frontal X-ray images of the knee and 34 lateral X-ray images of the foot, see Figure 3.4) and two generated sets (40 stars obscured by noise and a mixed set of 20 images of triangles and 20 images of squares obscured by noise, see Figure 3.5). Additionally, all artificially-generated shapes are distorted (i.e. corners are rounded and sides are uneven). All images have been rescaled to contain approximately fifty thousand pixels. In case of the foot dataset the heel bone was detected, the femur was matched in the knee datasets and respective shapes were detected on stars and triangles and squares datasets. In the foot dataset it was necessary to introduce a hand-drawn line separating the heel bone from the talus bone. These two bones significantly overlap on the X-ray images which results in poorly selected boundaries of superpixels.

The number n of sampling points is chosen as the smallest power of two that accurately represents features of detected objects. The choice of n is limited to powers of two because the multiscale approach is applied. The value of n cannot be too large as it significantly increases computation time and introduces additional errors due to equidistance assumption (see Eq. (3.1)) being only approximately satisfied. Figure 3.3 presents the mean squared error from this source by comparing angles $\{\alpha^i_{j,0}\}_{i=1,\ldots,M, j=1,\ldots,n}$ obtained from reduced-scale images used for testing to angles $\{\hat{\alpha}^i_{j,0}\}_{i=1,\ldots,M, j=1,\ldots,n}$ obtained from higher-resolution im-

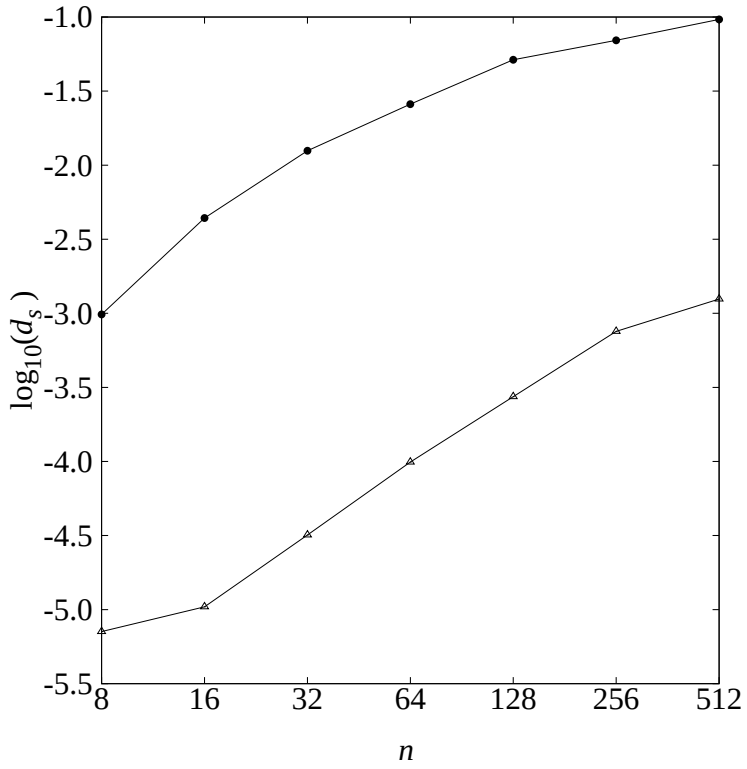


Figure 3.3: Relation between error d_s (defined in Equation (3.26)) introduced by unequal distances between successive points. Error for knee dataset is marked by circles while error for triangles and squares dataset is marked by triangles

ages having 1 to 1.5 megapixels. The formula for mean squared error is

$$d_s = \frac{1}{(n-1)M} \sum_{i=1}^{M} \sum_{j=2}^{n} (\alpha_{j,0}^i - \alpha_{1,0}^i - (\hat{\alpha}_{j,0}^i - \hat{\alpha}_{1,0}^i))^2, \tag{3.26}$$

where all angles are assumed to be in $(-\pi, \pi]$.

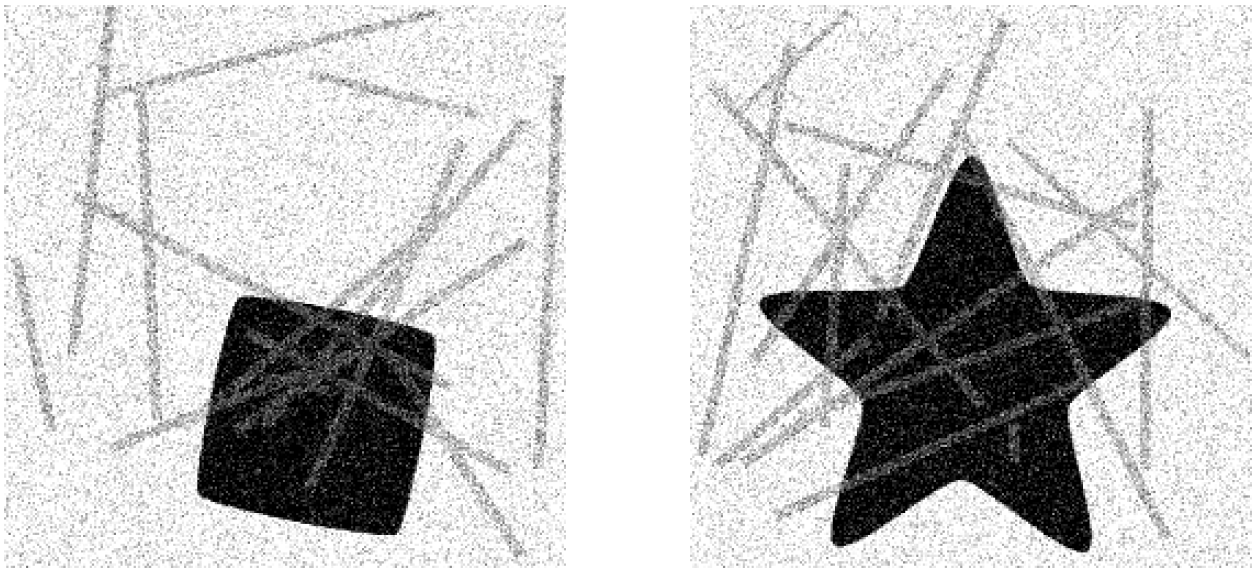
The chosen oversegmentation algorithm is the watershed from markers method (Meyer and Beucher, 1990), see Section 2.4.2. The selected markers are local minima of the gradient magnitude image in a square $2n_m + 1 \times 2n_m + 1$ neighbourhood centred on the pixel where n_m is a certain constant. The compact watershed and SLIC algorithms (Neubert and Protzel, 2014) were also tested but found inferior due to worse adherence to the real boundaries in the images.

Depending on the selected oversegmentation method the graph G may or may not be biconnected. Typically there is a single largest biconnected component of the graph G containing the shape to be found and other, much smaller ones, can be removed to speed up the computations.

The minimization in Equation (3.8) is calculated using single-variable optimization algorithm from the dlib library (King, 2009). The accuracy parameter is set to 0.01 and at most thirty iterations are allowed.

(a) (b)

Figure 3.4: Example images from the foot (Figure 3.4a) and knee (Figure 3.4b) datasets.



(a) (b)

Figure 3.5: Example images from the squares and triangles (Figure 3.5a) and the star (Figure 3.5b) datasets.

For each dataset a separate model was created. The given shape was matched in each image of the respective dataset given initial paths with three edges. For each dataset the mean Dice coefficient (Dice, 1945) d_{mean}, standard deviation d_{sd} of Dice coefficients, minimum d_{min} and maximum d_{max} of Dice coefficients were calculated. The sets of pixels inside the true object boundary and inside the curve found by the proposed algorithm were used to calculate the Dice coefficient.

The algorithm parameters were hand-selected for each dataset. Relatively high dimensionality of the parameter space and long matching time (a few minutes to half an hour for a single image analysed on a typical desktop computer) make use of general purpose optimization algorithms infeasible. Application of specialized parameter-fitting algorithms may be considered in the future as a possible improvement of the method (Hutter et al., 2011, Birattari et al., 2010, Hutter et al., 2009, Ansótegui et al., 2009, Burke et al., 2013).

Results

The results (the mean Dice coefficients, standard deviation of Dice coefficients, minimum and maximum Dice coefficient in a dataset obtained using the leave-one-out cross-validation) and selected parameters are summarized in Table 3.1. The obtained mean Dice coefficients are comparable with state-of-the-art pattern matching algorithms. Very high mean Dice coefficient in the dataset with triangles and squares indicates that the algorithm can learn more than one shape at the same time. This can also be an indication of robustness for families of shapes that exhibit significant variation.

Figure 3.6 displays the results of pattern matching in two cases from different datasets. The accuracy of matching is very good in areas of high gradient magnitude. In areas of low gradient magnitude the oversegmentation does not follow the edges of the object and, as a result, the accuracy decreases. Application of a better oversegmentation algorithm or matching the parts of shape in low gradient magnitude areas using a different approach might improve the accuracy of the match.

In a few cases, the object was not found. This may happen when there is a large area of low gradient magnitude instead of a sharp boundary. In many such cases increasing the parameter n_{sel} leads to better results at the cost of significantly increased matching time. Dynamic adjustment of n_{sel} could reduce this issue.

One example of incorrect object segmentation is depicted in Figure 3.7. An image from the stars dataset (see Subfigure 3.7a) has been segmented using parameters from Table 3.1. Black lines in Subfigures 3.7b to 3.7j correspond to the underlying superpixel segmentation. Red lines in Subfigures 3.7b to 3.7i represent edges of paths in the set *multitrails* (see Algorithm 2), whereas the red line in Subfigure 3.7j corresponds to edges of the best fit. For comparison, Figure 3.8 depicts a successful process of segmentation of a different image.

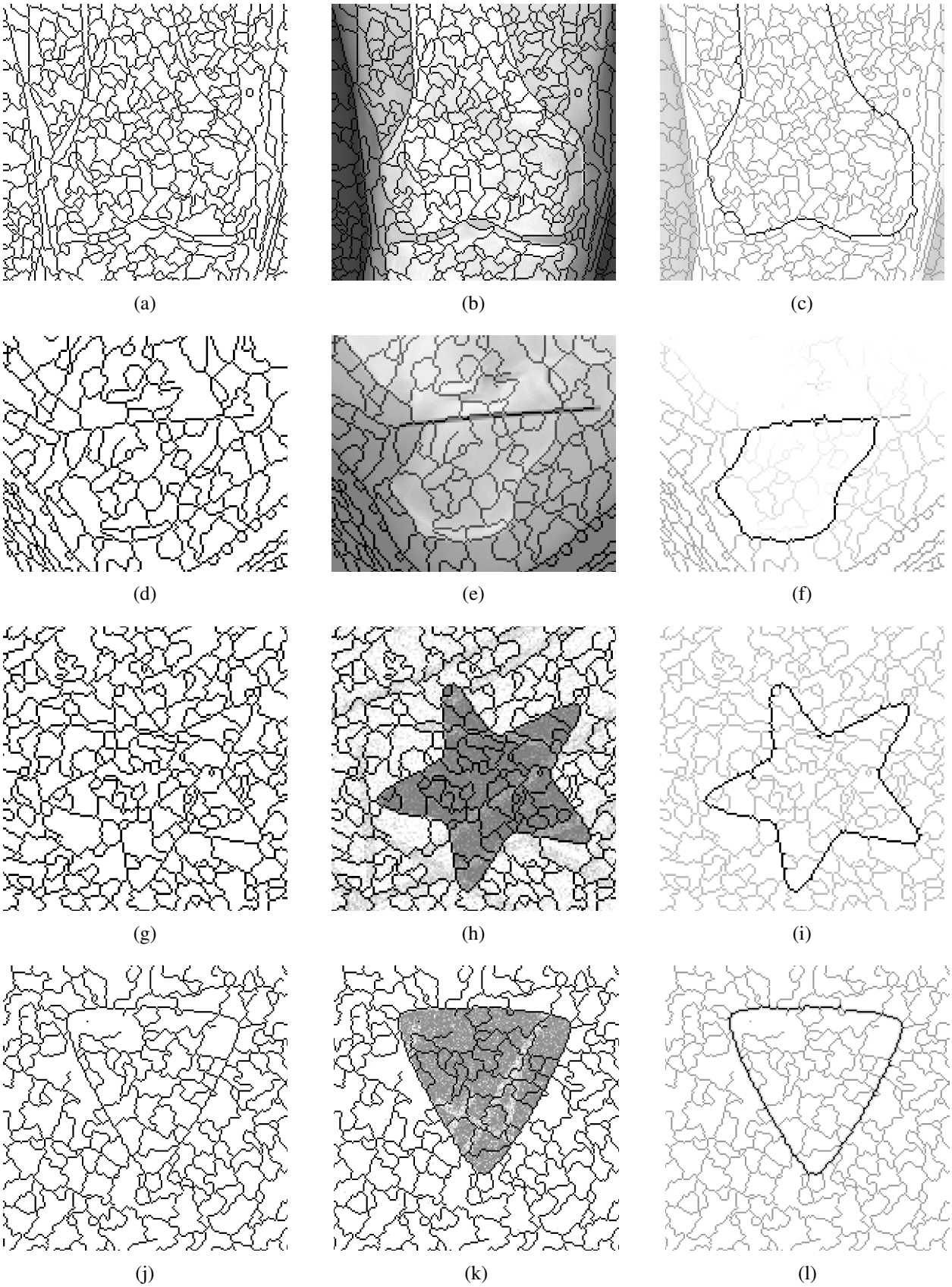


Figure 3.6: An example of pattern matching on an image from the knee dataset ((a)–(c)), foot dataset ((d)–(f)), star dataset ((g)–(i)) and triangles and squares ((j)–(l)) dataset. From the left column to the right column: the watershed segmentation of the selected image ((a), (d), (g) and (j)), watershed segmentation overlaid on top of the original image ((b), (e), (h) and (k)), the shape (nodes in the graph G) matched using the proposed approach, coloured black, on top of the watershed segmentation ((c), (f), (i) and (l)).

Table 3.1: Algorithm parameters selected for the used datasets and matching results (mean Dice coefficients d_{mean}, minimum d_{min} and maximum d_{max} of Dice coefficients in a set, standard deviation d_{sd} of Dice coefficients) and results obtained for the knee dataset using the AAM method

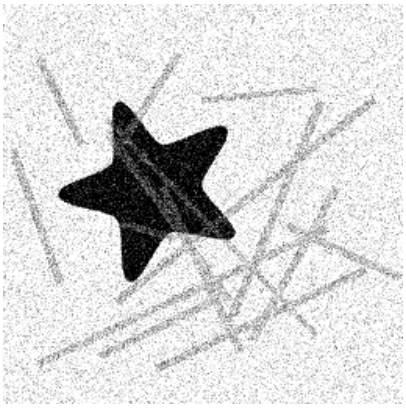
dataset	**knee (AAM)**	**knee**	**foot**	**star**	**triangle, square**
n	-	128	64	128	64
n_{sel}	-	300	500	300	500
n_m	-	5	6	5	5
e_{eig}	-	1	1	1	1
e_{hde}	-	0.005	0	0.005	0
e_{grad}	-	10	10	0.02	70
e_{scale}	-	1.0	0.6	2.0	1.0
e_{cde}	-	0.01	0.1	0.001	0.01
a_{occl}	-	10	10	10	10
s_{min}	-	2.0	1.5	2.0	3.0
s_{max}	-	1.0	0.3	0.5	0.5
c_{min}	-	0.1	0.1	0.1	0.1
c_{max}	-	0.1	0.3	0.3	0.3
E_{min}	-	95	95	95	95
a_{grad}	-	70	70	70	70
d_{mean}	0.94	0.95	0.91	0.94	0.97
d_{min}	0.77	0.47	0.66	0.21	0.74
d_{max}	0.985	0.996	1	1	1
d_{sd}	0.06	0.07	0.1	0.16	0.05

In both depicted cases initially broad search is narrowed in subsequent iterations. It can be observed that in Figure 3.7 the correct shape has been removed from consideration between steps 48 and 64 due to the tendency of the algorithm to overestimate the size of searched shape. A separate problem can be spotted in Subfigures 3.7f and 3.7g. A number of very similar paths exhaust the maximum capacity n_{sel} of the *multitrails* set. Elimination of very similar paths from this set would improve the performance of presented algorithm.

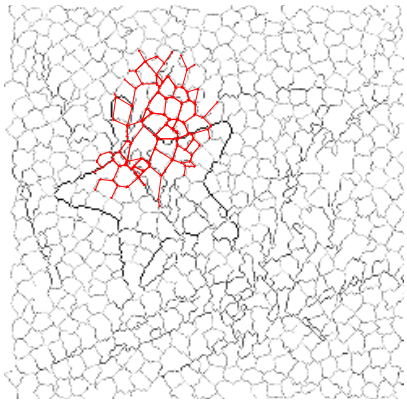
Discussion

The proposed method is a very promising pattern recognition algorithm. Good accuracy has been achieved on multiple test sets. The algorithm is a successful proof of concept of a new approach to pattern matching.

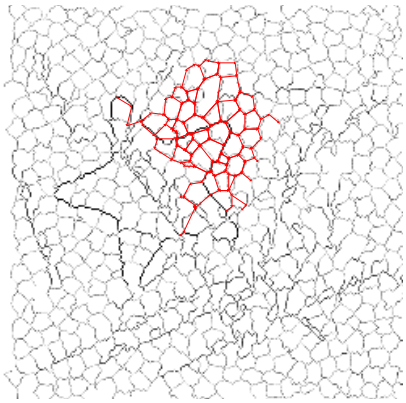
The most significant issue is lack of clear method for parameter selection. Machine learning methods, and in particular parameter optimization algorithms (Hutter et al., 2011, Birattari et al., 2010, Hutter et al.,



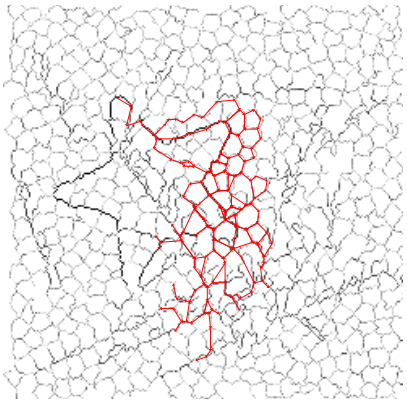
(a) Input image



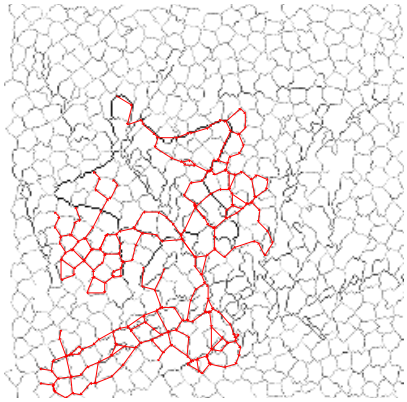
(b) After 8 steps



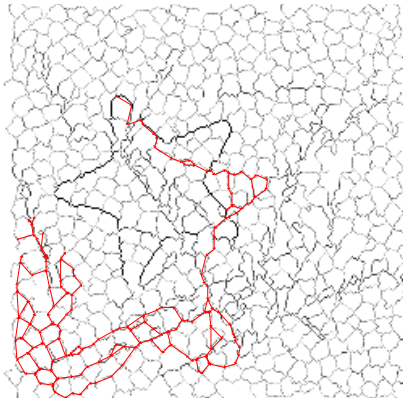
(c) After 16 steps



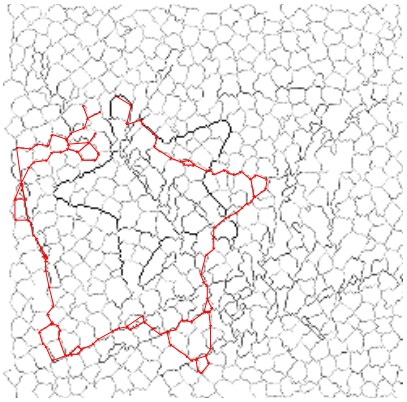
(d) After 32 steps



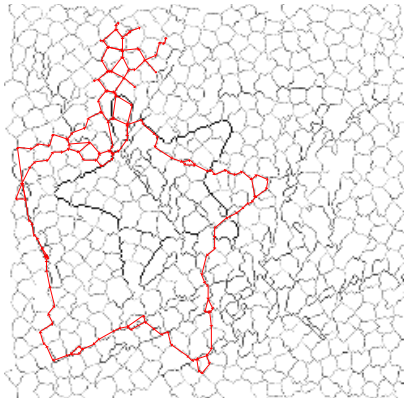
(e) After 48 steps



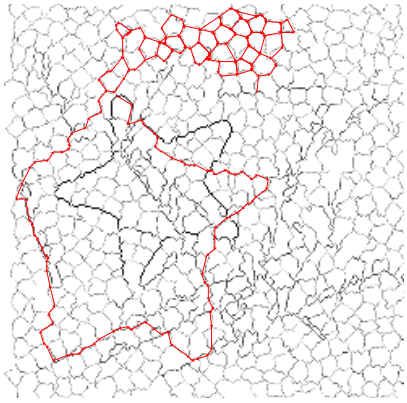
(f) After 64 steps



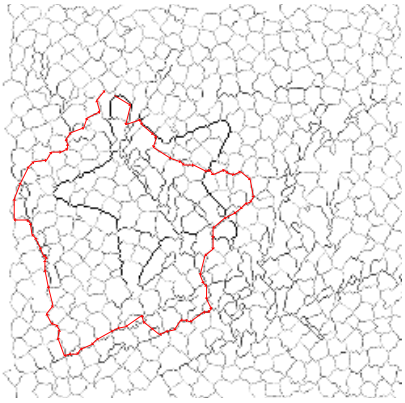
(g) After 80 steps



(h) After 96 steps

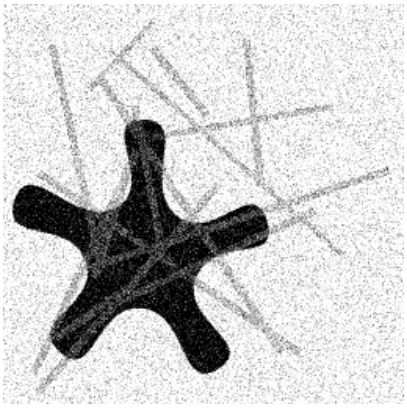


(i) After 112 steps

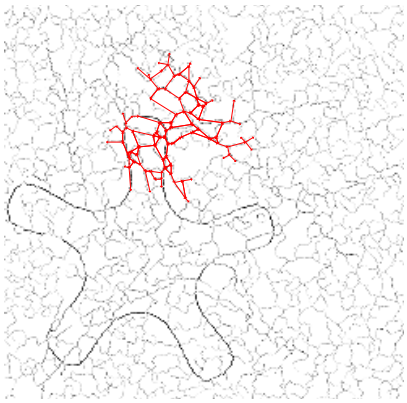


(j) Final path

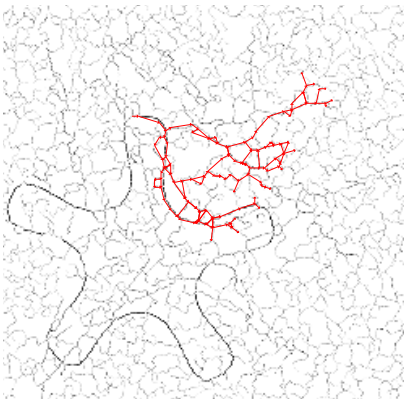
Figure 3.7: An example of incorrect pattern matching



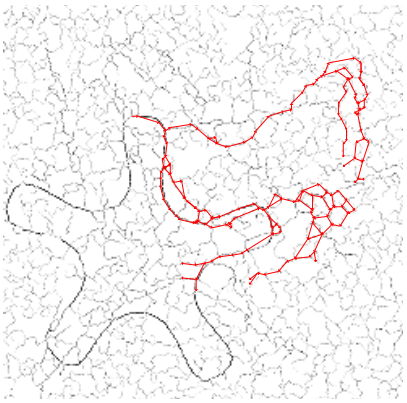
(a) Input image



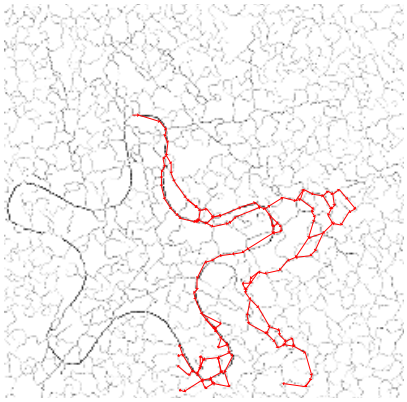
(b) After 8 steps



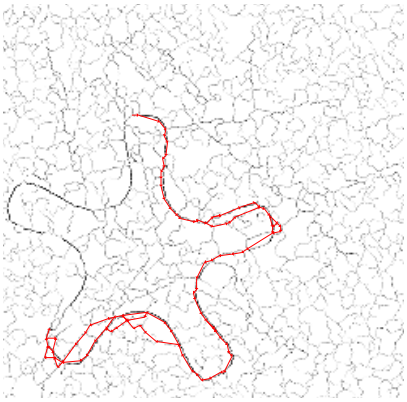
(c) After 16 steps



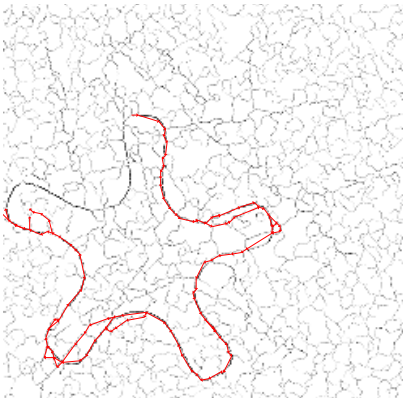
(d) After 32 steps



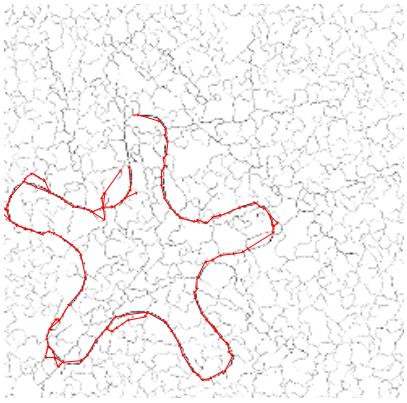
(e) After 48 steps



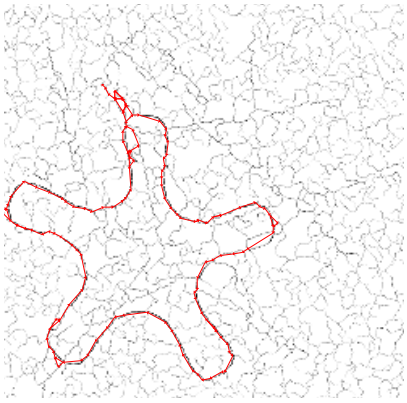
(f) After 64 steps



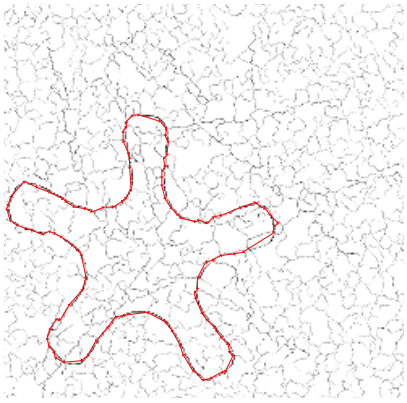
(g) After 80 steps



(h) After 96 steps



(i) After 112 steps



(j) Final path

Figure 3.8: An example of successful pattern matching

2009, Ansótegui et al., 2009, Burke et al., 2013), are expected to provide a solution to this problem. This area of research is left open to be pursued as a future work.

The possibility of recognition of multiple shapes, as seen in results for the triangles and squares dataset, is a remarkable quality of the proposed algorithm since it was not designed with this goal in mind. A more sophisticated shape model, for example a geodesic variant of Generalized PCA (Vidal et al., 2005), can be expected to give even better results. Additionally, a strategy similar to the diverse M-best solutions (Batra et al., 2012) can be used for more efficient management of the set of considered paths.

Chapter 4

Elastic shape analysis for superpixel graphs

Shape in Computer Vision is commonly represented using landmarks (Dryden and Mardia, 1998, Cootes et al., 1995) by Point Distribution Models (PDMs). Despite the unquestionable success of these methods selection and identification of landmarks remains a challenging problem and very often requires manual assistance (Zhang and Golland, 2016). The developing area of Elastic Shape Analysis (ESA), pioneered by Younes (1998), offers a solution. In this approach shape representation is parametrization-invariant. The problem of landmark matching is automatically solved and a continuous object boundary is provided. Recent developments in ESA (Michor and Mumford, 2006, Mio et al., 2007, Sundaramoorthi et al., 2011, Younes, 2012, Srivastava et al., 2012, Turaga and Srivastava, 2016), including the Square Root Velocity (SRV) function representation (Joshi et al., 2007) provide a method for effective computation of elastic distance between shapes and shape modelling, as well as variability modelling (Xie, 2017, Xie et al., 2017).

In this chapter the well-known Dynamic Programming method for computing elastic distance between curves (Bernal et al., 2016) is extended. The new algorithm selects two paths from two planar directed graphs whose drawings are closest under the elastic metric among all paths between selected nodes in these digraphs. A theoretical analysis of the problem is performed and conditions for the existence of an efficient algorithm are provided. A positive result for a class of digraphs of practical importance is reported.

The extension is primarily motivated by an application to pattern recognition using a superpixel segmentation of an image (see Section 2.4.4). Such an oversegmentation constitutes a constraint on shapes that may be present in an image (Mori et al., 2004). This division into superpixels can be interpreted as a drawing of a certain planar digraph, as described in Section 4.1.1. A curve, obtained by averaging over training shape boundaries, corresponds to the drawing of the second directed graph. Finding a path in the superpixel segmentation digraph whose drawing is the closest to the mentioned curve in the elastic metric, defines a segmentation of the given image.

The elastic metric is described in Section 2.2.1. Section 4.1 describes the extension of elastic curve registration to planar digraphs. Experiments in which the new algorithm is applied to boundary delineation are described in Section 4.2. Finally, the conclusions are presented in Section 4.3.

Methods

The new algorithm described in this chapter is an extension of the ESA framework (Srivastava et al., 2011) to drawings of planar digraphs. Figure 4.1 shows a sample pair of digraph drawings that constitute the input to the algorithm. The output, consisting of a pair of paths (one in each digraph) together with a matching between their nodes is also depicted. Drawings of these paths are the closest among drawings of all paths between given nodes and the matching realizes the optimal registration represented by the reparametrization γ from Equation (2.16). The set σ constrains the number of edges in one path that may correspond to a fragment of the other path. Each pair (i, j) in σ means that i consecutive edges in a path in G_1 may correspond to j consecutive edges in a path in G_2.

Notation

Let $G_1 = (V_1, E_1), G_2 = (V_2, E_2)$ be two given digraphs. Their planar drawings are defined by vertex position functions $\phi_k \colon V_k \to \mathbb{R}^2$ and edge functions $\xi_k \colon E_k \to \mathrm{L}^2([0,1], \mathbb{R}^2)$ for $k = 1, 2$. It is assumed that the edges cross only at vertices, that is $\xi_k(e_{m_k})(t) = \xi_k(e_{l_k})(u)$ for any two different edges $e_{l_k}, e_{m_k} \in E_k$, $e_{l_k} \neq e_{m_k}$ implies that $t, u \in \{0, 1\}$, except for edges $(v, \bar{v})$, $(\bar{v}, v)$ connecting the same pair of vertices, where it is assumed that $\xi_k((v, \bar{v}))(t) = \xi_k((\bar{v}, v))(1 - t)$ for all $t \in [0, 1]$ and all $v, \bar{v} \in V_k$, $v \neq \bar{v}$, for $k = 1, 2$. It is also assumed that there is a vertex $v_k \in V_k$ incident to both edges e_{l_k} and e_{m_k} such that

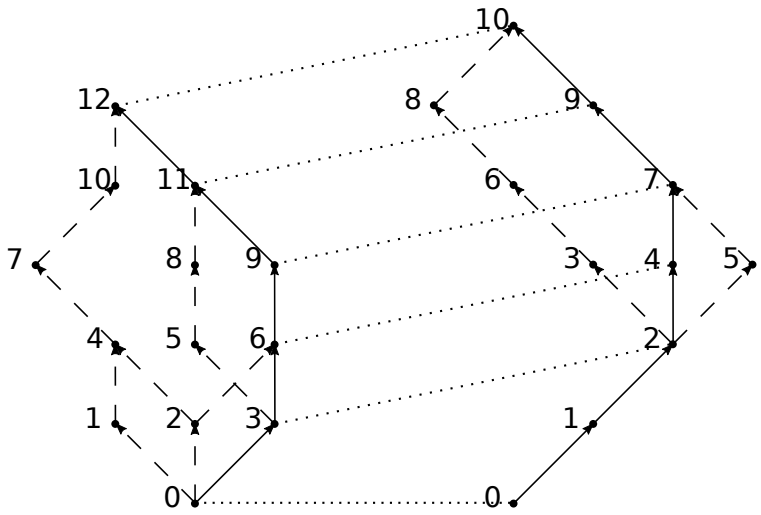


Figure 4.1: A pair of directed graphs G_1 (left) and G_2 (right) with a pair of most similar paths between nodes 0 and 12 in G_1 and 0 and 10 in G_2, drawn in solid line. Other edges are drawn as dashed lines while matched nodes are connected with dotted lines. In this example $\sigma = \{(1, 1), (1, 2), (2, 1)\}$.

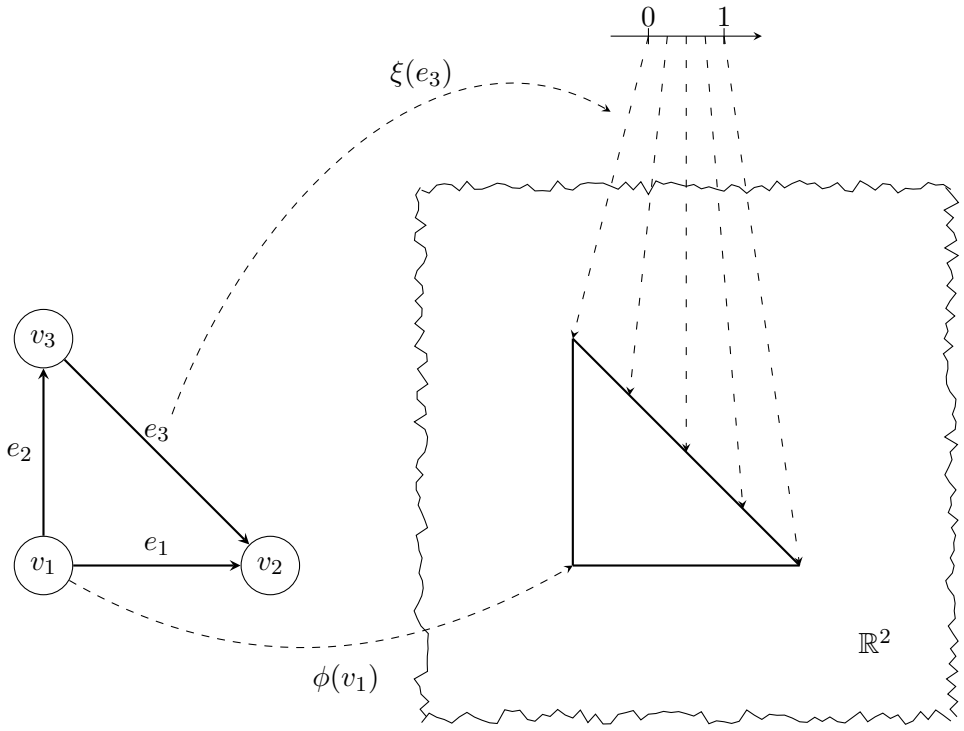


Figure 4.2: The correspondence between a digraph and its drawing

$\phi_k(v_k) = \xi_k(e_{l_k})(t) = \xi_k(e_{m_k})(u)$ for $k = 1, 2$ and some $u, t \in \{0, 1\}$. It is known that all planar digraphs can be drawn in this way, even when all edges are required to correspond to straight line segments (Fáry, 1948, Tutte, 1960, 1963).

The correspondence between a discrete digraph and its drawing is depicted in Figure 4.2. On the left hand side, there is a digraph $G = (V, E)$ with three vertices ($V = \{v_1, v_2, v_3\}$) and three edges ($E = \{e_1, e_2, e_3\}$). On the right hand side there is a drawing of G defined by functions ϕ, which maps vertices to their positions, and ξ, which maps edges to parametric descriptions of line segments.

The new algorithm described in this chapter translates the problem of finding a pair of paths in digraphs G_1, G_2 with the smallest elastic distance to a minimum weighted average path problem in a digraph called the σ-product of digraphs G_1 and G_2. This fact is formally stated in Theorem 1 and proven in Appendix B. The construction of the σ-product of two digraphs is described by Definitions 29, 30 and 31.

As stated in Section 2.2.1, some authors use higher-order interpolation for calculating elastic distance between sampled curves. This work assumes first order spline interpolation between sampling points, or drawings of nodes of graphs. These nodes naturally impose continuity of path drawings but there is no clear way to ensure the continuity of their first derivatives.

Definition 29. *Let σ be a set of pairs of positive integers. Then a pair of paths (sequences of edges) p_1 in a digraph $G_1 = (V_1, E_2)$, p_2 in a digraph $G_2 = (V_2, E_2)$ is called a* ***σ-pair of paths*** *between $(v_{b,1}, v_{b,2})$ and $(v_{e,1}, v_{e,2})$ if and only if:*

- *p_1 starts in $v_{b,1}$ and ends in $v_{e,1}$,*
- *p_2 starts in $v_{b,2}$ and ends in $v_{e,2}$,*
- *if p_1 has l_1 edges and p_2 has l_2 edges, the pair (l_1, l_2) belongs to σ.*

Definition 30. *If for each $i = 1, 2, \ldots, N$ a pair $(p_{1,i}, p_{2,i})$ is a σ-pair of paths between $(v_{1,i-1}, v_{2,i-1})$ and $(v_{1,i}, v_{2,i})$ in digraphs G_1, G_2, then $(p_{1,i}, p_{2,i})_{i=1}^{N}$ is a **sequence of σ-pairs of paths** in G_1, G_2.*

A fixed parametrization is needed to consider drawings of paths as parametrized curves. This parametrization is defined by functions $\Delta t_k \colon E_k \to \mathbb{R}$ for $k = 1, 2$. One could for simplicity assume that Δt_k is equal to 1 for all edges and $k = 1, 2$. On the other hand parametrization by arc length is more natural. In this case

$$\Delta t_k(e_k) = \int_0^1 \sqrt{\sum_{d=1}^{2} \left(\frac{\mathrm{d}\xi_{k,d}(e_k)}{\mathrm{d}t}(t) \right)^2} \, \mathrm{d}t, \tag{4.1}$$

where $\xi_{k,1}(e_k)(t)$ and $\xi_{k,2}(e_k)(t)$ for $t \in [0, 1]$ are the coordinates of $\xi_k(e_k)(t)$ in the standard basis of $\mathbb{R}^2$ and e_k is an edge from the set E_k for $k = 1, 2$.

Definition 31. *The **σ-product of two digraphs** $G_1 = (V_1, E_1)$, $G_2 = (V_2, E_2)$ with drawings ϕ_1, ξ_1, ϕ_2, ξ_2 is a weighted digraph $G_1 \times_\sigma G_2 = (V, E, w)$ such that:*

- *the set of vertices of the σ-product is the Cartesian product of the sets of vertices of digraphs G_1, G_2, that is $V = V_1 \times V_2$,*
- *an edge $((v_{b,1}, v_{b,2}), (v_{e,1}, v_{e,2}))$ belongs to the set E of edges of the σ-product if and only if there is a pair of σ-paths p_1, p_2 between $(v_{b,1}, v_{b,2})$ and $(v_{e,2}, v_{e,2})$,*
- *assuming that $w_I(e)$ for an edge $e = (v_{b,1}, v_{b,2}), (v_{e,2}, v_{e,2})$ is the minimum value of*

$$I\left(0, 1, 0, 1, \eta_1^q(p_1), \eta_2^q(p_2) \sqrt{\frac{\sum_{j=1}^{n_{2,0}} \Delta t_2(e_{2,0,j})}{\sum_{j=1}^{n_{1,0}} \Delta t_1(e_{1,0,j})}}\right), \tag{4.2}$$

where I is the function defined by Equation (2.18) and $\eta_1^q(p_1)$ and $\eta_2^q(p_2)$ are SRV representations of functions defined by Equation (4.7), over all σ-pairs $(p_1, p_2) = ((e_{1,0,j})_{j=1}^{n_{1,0}}, (e_{2,0,j})_{j=1}^{n_{2,0}})$ of paths between $(v_{b,1}, v_{b,2})$ and $(v_{e,2}, v_{e,2})$, the weight $w(e)$ of an edge e is the pair

$$w(e) = \left(w_I(e), \sum_{j=1}^{n_{1,0}} \Delta t_1(e_{1,0,j}) \right), \tag{4.3}$$

where $((e_{1,0,j})_{j=1}^{n_{1,0}}, (e_{2,0,j})_{j=1}^{n_{2,0}})$ is the σ-pair of paths that minimizes the value of Equation (4.2) for the edge e.

It is worth noting that the choice of Δt_1 over Δt_2 in Equation (4.3) follows from the same choice in Equation (2.18). The consequences of this asymmetry are more profound in the graphical setting.

A few examples of digraphs and their σ-products are given in Figures 4.3 and 4.4. The non-weighted part of the σ-product of digraphs G_1, G_2 generalizes a few standard digraph products (although they are more commonly applied to non-directed graphs):

- When $\sigma = \{(0,1),(1,0)\}$, then the σ-product is equivalent to the Cartesian product (Harary, 1969).
- When $\sigma = \{(1,1)\}$, then the σ-product is equivalent to the tensor product (McAndrew, 1963, Harary and Trauth Jr., 1966), also called the direct product (Hellmuth and Marc, 2015).
- When $\sigma = \{(0,1),(1,0),(1,1)\}$, then the σ-product is equivalent to the strong product (Hellmuth and Marc, 2015).

Additionally, when G_2 is strongly connected then $G_1 \times_{\{(1,i)\colon i\in\mathbb{N}\}\cup\{0,1\}} G_2$ is equivalent to the lexicographical product of G_1 and G_2 (Ng, 1998). If both G_1 and G_2 are strongly connected, then $G_1 \times_{\{(1,i)\colon i\in\mathbb{N}\}\cup\{(i,1)\colon i\in\mathbb{N}\}} G_2$ is equivalent to the co-normal product of G_1 and G_2 (Galluccio et al., 1994).

The functions Δt_k can be used to define drawings of paths in a σ-product and, subsequently, SRV representations of these drawings. Let $p = ((e_{1,i,j})_{j=1}^{n_{1,i}}, (e_{2,i,j})_{j=1}^{n_{2,i}})_{i=1}^{N}$ be a sequence of σ-pairs in digraphs $G_1 = (V_1, E_1)$, $G_2 = (V_2, E_2)$. Considered functions of p are piecewise linear. Their domains are subdivided at $T_{k,i}(p) \in [0,1]$ defined by

$$T_{k,i}(p) = \begin{cases} 0 & \text{if } i = 0 \\ \frac{\sum_{z=1}^{i}\sum_{j=1}^{n_{k,z}} \Delta t_k(e_{k,z,j})}{\sum_{z=1}^{N}\sum_{j=1}^{n_{k,z}} \Delta t_k(e_{k,z,j})} & \text{if } i > 0 \end{cases} \tag{4.4}$$

for $k = 1, 2$ and $i = 0, 1, \ldots, N$ and

$$T_{k,i,j}(p) = \begin{cases} 0 & \text{if } j = 0 \\ \frac{\sum_{w=1}^{j} \Delta t_k(e_{k,i,w})}{\sum_{w=1}^{n_{k,i}} \Delta t_k(e_{k,i,w})} & \text{if } j > 0 \end{cases} \tag{4.5}$$

for $k = 1, 2$, $i = 1, \ldots, N$ and $j = 0, 1, 2, \ldots, n_{k,i}$.

The drawings of a sequence of σ-pairs of paths $p = ((e_{1,i,j})_{j=1}^{n_{1,i}}, (e_{2,i,j})_{j=1}^{n_{2,i}})_{i=1}^{N}$ in digraphs $G_1 = (V_1, E_1)$, $G_2 = (V_2, E_2)$ in $\mathbb{R}^2$, denoted by $\rho_1(p)$ and $\rho_2(p)$, are defined by

$$\rho_k(p)(t) = \begin{cases} \eta_k((e_{k,1,j})_{j=1}^{n_{k,1}}) \left(\frac{t - T_{k,0}(p)}{T_{k,1}(p) - T_{k,0}(p)}\right) & \text{if } T_{k,0}(p) \le t < T_{k,1}(p) \\ \eta_k((e_{k,2,j})_{j=1}^{n_{k,2}}) \left(\frac{t - T_{k,1}(p)}{T_{k,2}(p) - T_{k,1}(p)}\right) & \text{if } T_{k,1}(p) \le t < T_{k,2}(p) \\ \vdots & \\ \eta_k((e_{k,N,j})_{j=1}^{n_{k,N}}) \left(\frac{t - T_{k,N-1}(p)}{T_{k,N}(p) - T_{k,N-1}(p)}\right) & \text{if } T_{k,N-1}(p) \le t \le T_{k,N}(p) \end{cases} \tag{4.6}$$

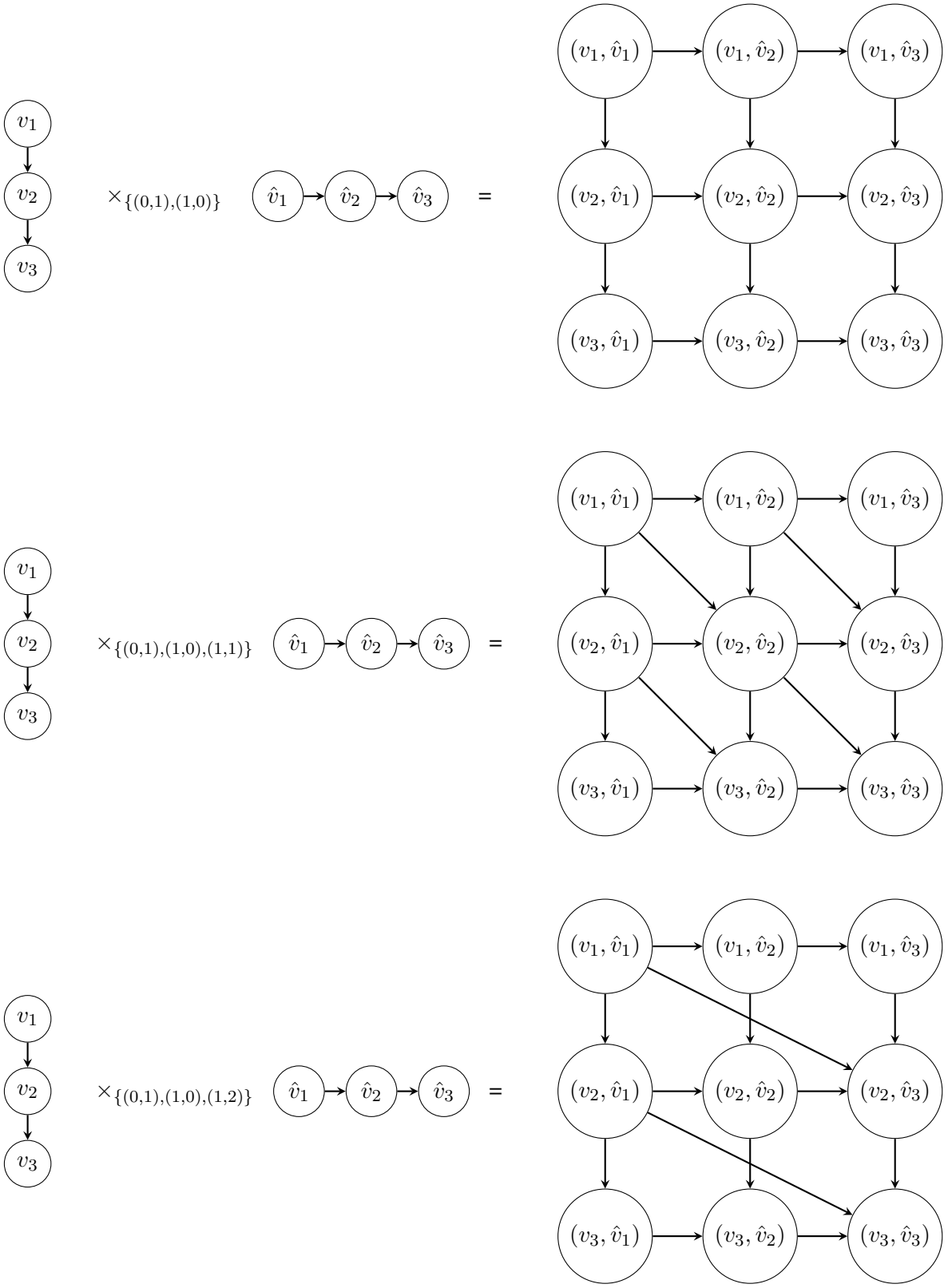


Figure 4.3: Paths and their σ-products

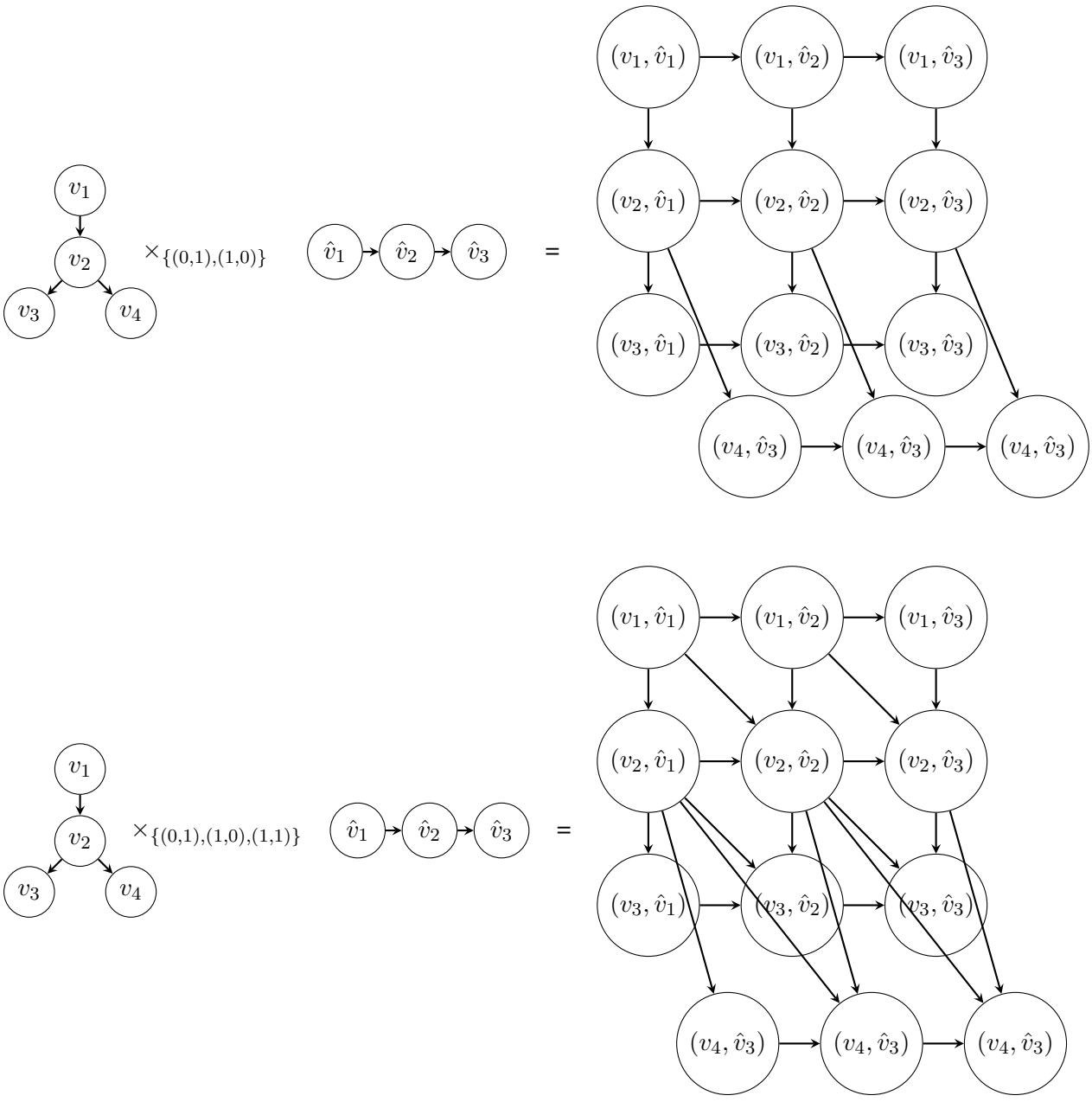


Figure 4.4: Digraphs and their σ-products

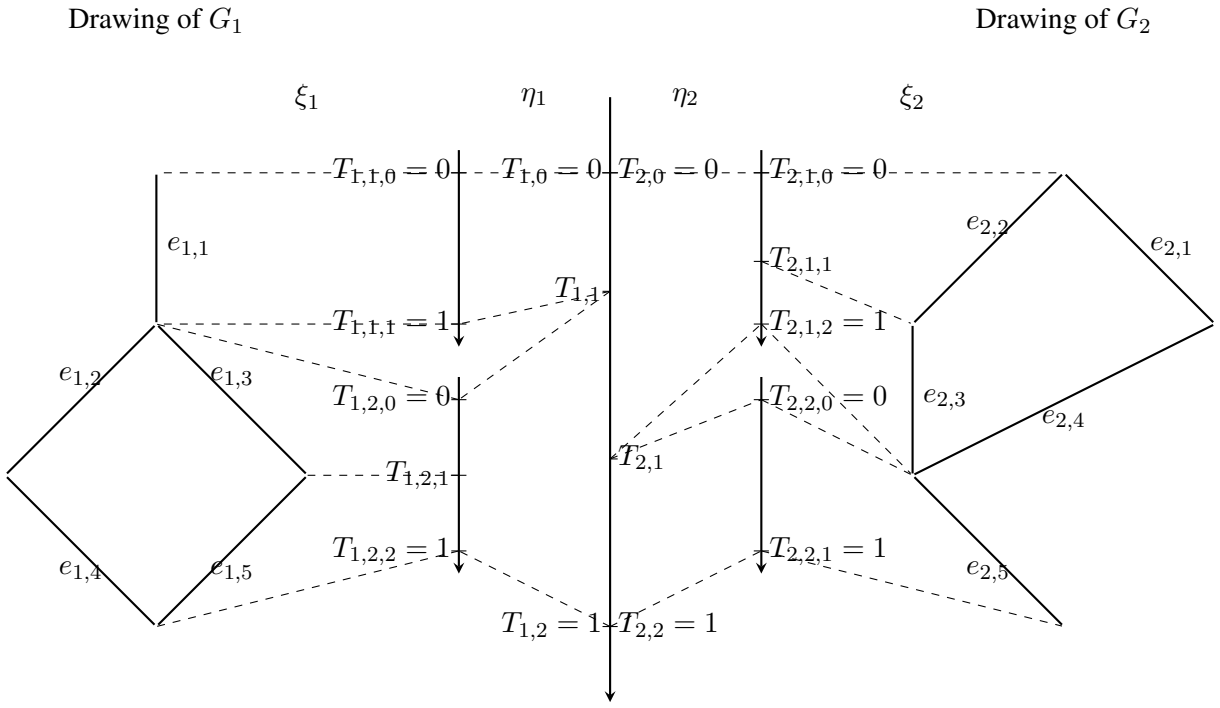


Figure 4.5: A graphical interpretation of $\rho_1(p)$ and $\rho_2(p)$ for two digraphs G_1, G_2 and $p = (((e_{1,1}), (e_{2,2}, e_{2,3})), ((e_{1,3}, e_{1,5}), (e_{2,5})))$

for $k = 1, 2$ where η_k is defined by

$$\eta_k((e_{k,i,j})_{j=1}^{n_{k,i}})(t) = \begin{cases} \xi_k(e_{k,i,1})\left(\frac{t-T_{k,i,0}(p)}{T_{k,i,1}(p)-T_{k,i,0}(p)}\right) & \text{if } T_{k,i,0}(p) \leq t < T_{k,i,1}(p) \\ \xi_k(e_{k,i,2})\left(\frac{t-T_{k,i,1}(p)}{T_{k,i,2}(p)-T_{k,i,1}(p)}\right) & \text{if } T_{k,i,1}(p) \leq t < T_{k,i,2}(p) \\ \vdots & \\ \xi_k(e_{k,i,n_{k,i}})\left(\frac{t-T_{k,i,n_{k,i}-1}(p)}{T_{k,i,n_{k,i}}(p)-T_{k,i,n_{k,i}-1}(p)}\right) & \text{if } T_{k,i,n_{k,i}-1}(p) \leq t \leq T_{k,i,n_{k,i}}(p) \end{cases} \tag{4.7}$$

for $k = 1, 2$, $i = 1, 2, \ldots, N$. Figure 4.5 graphically describes the construction of $\rho_1(p)$ and $\rho_2(p)$ for two digraphs and a sequence of σ-paths p. Natural parametrization is assumed.

The SRV representations of $\rho_k(p)$, $\eta_k((e_{k,i,j})_{j=1}^{n_{k,i}})$ and $\xi_k(e_{k,i,n_{k,i}})$ for $k = 1, 2$, $i = 1, 2, \ldots, N$, $j = 1, 2, \ldots, n_{k,i}$ are denoted by, respectively, $\rho_k^q(p)$, $\eta_k^q((e_{k,i,j})_{j=1}^{n_{k,i}})$ and $\xi_k^q(e_{k,i,n_{k,i}})$. Additionally, p defines a

reparametrization $\gamma(p)\colon [0,1] \to [0,1]$ as follows

$$\gamma(p)(t) = \begin{cases} \frac{T_{2,1}(p)-T_{2,0}(p)}{T_{1,1}(p)-T_{1,0}(p)}(t - T_{1,0}(p)) + T_{2,0}(p) & \text{if } T_{1,0}(p) \leq t < T_{1,1}(p) \\ \frac{T_{2,2}(p)-T_{2,1}(p)}{T_{1,2}(p)-T_{1,1}(p)}(t - T_{1,1}(p)) + T_{2,1}(p) & \text{if } T_{1,1}(p) \leq t < T_{1,2}(p) \\ \vdots & \\ \frac{T_{2,N}(p)-T_{2,N-1}(p)}{T_{1,N}(p)-T_{1,N-1}(p)}(t - T_{1,N-1}(p)) + T_{2,N-1}(p) & \text{if } T_{1,N-1}(p) \leq t \leq T_{1,N}(p). \end{cases} \tag{4.8}$$

Note that two different sequences of sigma paths may pass the same edges in respective digraphs but define a different reparametrization. For example, for digraphs in Figure 4.5, $\hat{p} = (((e_{1,1}, e_{1,3}), (e_{2,2}, e_{2,3})), ((e_{1,5}), (e_{2,5})))$ corresponds to the same paths in G_1 and G_2 as the p defined in caption but the reparametrization it defines is different.

The elastic distance d^c_{pre} between $\rho^q_1(p)$ and $\rho^q_2(p) \cdot \gamma(p)$, where the dot is the action defined by Equation (2.14), is equal to

$$\begin{aligned} d^c_{pre}(\rho^q_1(p), \rho^q_2(p) \cdot \gamma(p)) = \int_0^1 \left\| \rho^q_1(p)(t) - \rho^q_2(p)(\gamma(p)(t)) \sqrt{\dot{\gamma(p)}(t)} \right\|_2^2 \mathrm{d}t = \\ \sum_{i=1}^{N} (T_{1,i}(p) - T_{1,i-1}(p)) \int_0^1 \left\| \eta^q_1((e_{1,i,j})_{j=1}^{n_{1,i}})(\tau_i) - \eta^q_2((e_{2,i,j})_{j=1}^{n_{2,i}})(\tau_i) \sqrt{\frac{\sum_{j=1}^{n_{2,i}} \Delta t_2(e_{2,i,j})}{\sum_{j=1}^{n_{1,i}} \Delta t_1(e_{1,i,j})}} \right\|_2^2 \mathrm{d}\tau_i = \\ \sum_{i=1}^{N} (T_{1,i}(p) - T_{1,i-1}(p)) I\left(0, 1, 0, 1, \eta^q_1((e_{1,i,j})_{j=1}^{n_{1,i}}), \eta^q_2((e_{2,i,j})_{j=1}^{n_{2,i}}) \sqrt{\frac{\sum_{j=1}^{n_{2,i}} \Delta t_2(e_{2,i,j})}{\sum_{j=1}^{n_{1,i}} \Delta t_1(e_{1,i,j})}}\right), \end{aligned} \tag{4.9}$$

where $\tau_i = \frac{t - T_{1,i-1}(p)}{T_{1,i}(p) - T_{1,i-1}(p)}$. Each integral is equivalent to the value specified by Equation (2.18). Note that if p corresponds to a path $\bar{p}$ in a σ-product of digraphs, the value obtained in Equation (4.9) is a weighted average of weights of edges in $\bar{p}$.

The algorithm, given two pairs of vertices $v_{b,1}, v_{e,2} \in V_1$, $v_{b,2}, v_{e,2} \in V_2$ from digraphs $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$, finds a sequence of σ-pairs of paths $p = ((e_{1,i,j})_{j=1}^{n_{1,i}}, (e_{2,i,j})_{j=1}^{n_{2,i}})_{i=1}^{N}$ in, respectively, G_1 and G_2 such that the distance between their drawings in $\mathbb{R}^2$, $\rho_1(p)$ and $\rho_2(p) \circ \gamma(p)$, defined by Equation (4.6), is minimal with respect to a discretized version of d^c. This simultaneously solves the problems of finding the optimal reparametrization and finding the most similar paths in G_1 and G_2.

Table 4.1: Table of the binary operation $\oplus$. Symbols a and c represent nonnegative real numbers while b and d represent positive real numbers

$\oplus$	$(0,0)$	(a,b)	(∞,∞)
$(0,0)$	$(0,0)$	(a,b)	(∞,∞)
(c,d)	(c,d)	$\left(\frac{ab+cd}{b+d}, b+d\right)$	(∞,∞)
(∞,∞)	(∞,∞)	(∞,∞)	(∞,∞)

Algorithm

Algorithm 3 describes the procedure of finding a pair of the closest paths in digraphs G_1, G_2 between $(v_{b,1}, v_{b,2})$ and $(v_{e,2}, v_{e,2})$. The method is scale- and rotation-independent by assuming that drawings of vertices lie in fixed points, namely $\phi_k(v_{b,k}) = (0,0)$ and $\phi_k(v_{e,k}) = (0,1)$ for $k = 1, 2$. Any drawings can be made to satisfy these conditions by a pair of similarity transform, one for each digraph. The minimum weighted average path search is performed using the generalized Dijkstra's algorithm (Dijkstra, 1959) for routing algebras (see Definition 8 in Section 2.1.1). A similar technique is described in (Wang et al., 2005), where a minimum (non-weighted) average path algorithm is utilized for an analysis of wireless sensor networks. Boundary delineation using a shortest path algorithm is also described in (Martelli, 1976), although only pixel intensity levels are used to construct the cost function. A model-free segmentation algorithm based on local edge detection operators and iterative boundary elongation is described in (Lineberry, 1983). There are, however, no other similarities between Lineberry's approach and the algorithm described in this chapter.

Let $\mathcal{A} = (W, \oplus, \preceq)$ be a routing algebra. The set $W = \{(x,y)\colon x \in \mathbb{R}_{\geq 0}, y \in \mathbb{R}_{+}\} \cup \{(0,0), (\infty,\infty)\}$ is a set of pairs of numbers where the first one is nonnegative and the other is strictly positive, together with an identity element $(0,0)$ and a zero element (∞,∞). Table 4.1 contains the definition of the binary operation $\oplus$. The total order $\preceq$ is the lexicographical order, that is $(a,b) \preceq (c,d)$ if $a < c$ or $a = c$ and $b \leq d$, assuming ∞ is greater than any real number and $\infty = \infty$. The properties of $\mathcal{A}$ are discussed in Appendix A. Elements of the set W are used as weights in the product digraph $G_1 \times_\sigma G_2 = (V, E, w)$, according to Equation (4.3). A path $e_1, e_2, \ldots, e_z$ in the product digraph for some $z \geq 1$ is considered more optimal than a path $\hat{e}_1, \hat{e}_2, \ldots, \hat{e}_{\hat{z}}$ for some $\hat{z} \geq 1$ if and only if

$$w(e_1) \oplus w(e_2) \oplus \cdots \oplus w(e_z) \preceq w(\hat{e}_1) \oplus w(\hat{e}_2) \oplus \cdots \oplus w(\hat{e}_{\hat{z}}), \tag{4.10}$$

where $w\colon E \to \mathcal{A}$ is the weight function given by Equation (4.3).

Finding the optimal path with respect to weight function w corresponds to the optimal path problem where the path cost is defined by Equation (4.9). This observation can be shown by comparing the total weights of paths $p = e_1, e_2, \ldots, e_z$ and $\hat{p} = \hat{e}_1, \hat{e}_2, \ldots, \hat{e}_{\hat{z}}$ in a σ-product digraph for some $z \geq 1$ and

$\hat{z} \geq 1$:

$$w(e_1) \oplus w(e_2) \oplus \cdots \oplus w(e_z) \preceq w(\hat{e}_1) \oplus w(\hat{e}_2) \oplus \cdots \oplus w(\hat{e}_{\hat{z}}) \tag{4.11}$$

which is equivalent to

$$\begin{aligned} &\left(\frac{\sum_{i=1}^{z} w(e_i) \sum_{j=1}^{n_{1,i}} \Delta t_1(e_{1,i,j})}{\sum_{i=1}^{z} \sum_{j=1}^{n_{1,i}} \Delta t_1(e_{1,i,j})}, \sum_{i=1}^{z} \sum_{j=1}^{n_{1,i}} \Delta t_1(e_{1,i,j}) \right) \preceq \\ &\left(\frac{\sum_{i=1}^{\hat{z}} w(\hat{e}_i) \sum_{j=1}^{\hat{n}_{1,i}} \Delta t_1(\hat{e}_{1,i,j})}{\sum_{i=1}^{\hat{z}} \sum_{j=1}^{\hat{n}_{1,i}} \Delta t_1(\hat{e}_{1,i,j})}, \sum_{i=1}^{\hat{z}} \sum_{j=1}^{\hat{n}_{1,i}} \Delta t_1(\hat{e}_{1,i,j}) \right). \end{aligned} \tag{4.12}$$

Since

$$T_{1,i}(p) - T_{1,i-1}(p) = \frac{\sum_{j=1}^{n_{1,i}} \Delta t_1(e_{1,i,j})}{\sum_{i=1}^{z} \sum_{j=1}^{n_{1,i}} \Delta t_1(e_{1,i,j})} \tag{4.13}$$

for $i = 1, 2, \ldots, z$ and, analogically,

$$T_{1,i}(\hat{p}) - T_{1,i-1}(\hat{p}) = \frac{\sum_{j=1}^{\hat{n}_{1,i}} \Delta t_1(\hat{e}_{1,i,j})}{\sum_{i=1}^{\hat{z}} \sum_{j=1}^{\hat{n}_{1,i}} \Delta t_1(\hat{e}_{1,i,j})} \tag{4.14}$$

for $i = 1, 2, \ldots, \hat{z}$, the Equation (4.12) can be rewritten as

$$\begin{aligned} &\left(\sum_{i=1}^{z} (T_{1,i}(p) - T_{1,i-1}(p)) w(e_i), \sum_{i=1}^{z} \sum_{j=1}^{n_{1,i}} \Delta t_1(e_{1,i,j}) \right) \preceq \\ &\left(\sum_{i=1}^{\hat{z}} (T_{1,i}(\hat{p}) - T_{1,i-1}(\hat{p})) w(\hat{e}_i), \sum_{i=1}^{\hat{z}} \sum_{j=1}^{\hat{n}_{1,i}} \Delta t_1(\hat{e}_{1,i,j}) \right). \end{aligned} \tag{4.15}$$

Using Equations (4.2) and (4.9), Equation (4.15) is equivalent to

$$\begin{aligned} &\left(d_{pre}^{c}(\rho_1^q(p), \rho_2^q(p) \cdot \gamma(p)), \sum_{i=1}^{z} \sum_{j=1}^{n_{1,i}} \Delta t_1(e_{1,i,j}) \right) \preceq \\ &\left(d_{pre}^{c}(\rho_1^q(\hat{p}), \rho_2^q(\hat{p}) \cdot \gamma(\hat{p})), \sum_{i=1}^{\hat{z}} \sum_{j=1}^{\hat{n}_{1,i}} \Delta t_1(\hat{e}_{1,i,j}) \right). \end{aligned} \tag{4.16}$$

As the relation $\preceq$ represents the lexicographical order, this inequality holds if either the elastic distance defined by Equation (4.9) is smaller for p than for $\hat{p}$, or the distances are equal and the length of $\rho_1(p)$ is shorter than the length of $\rho_1(\hat{p})$.

The (commutative) routing algebra $\mathcal{A}$ is, in general, neither monotonic nor isotonic (see Definitions 8, 9, 11 and 10 and Appendix A). Commonly used optimal path algorithms do not guarantee optimality for such routing algebras (Yang and Wang, 2008). This problem does not occur in Algorithm 1 because de-

nominators in Equations (4.4) and (4.5) are known in advance for all possible paths from the start node to the end node. The numerators, on the other hand, can be easily factored into the edge weight. It is true because the paths and their lengths are determined and the problem is reduced to finding an optimal reparametrization. Observe that in the simplified case each element of the sum in Equation (4.9) depends only on a single edge of a product digraph.

Under certain conditions the classical Dijkstra's algorithm can be used to find the optimal path in a digraph weighted by elements of a routing algebra $\mathcal{A}$. Let $\mathcal{B} = (\mathbb{R}_{\geq 0} \cup \{\infty\}, +, \leq)$ be the routing algebra of the classical Dijkstra's algorithm. A function $h\colon \mathcal{A} \to \mathcal{B}$ defined by

$$h((a, b)) = ab \tag{4.17}$$

can be used to translate weights between two routing algebras $\mathcal{A}$ and $\mathcal{B}$. One can verify that h is a homomorphism of monoidal parts of respective routing algebras, that is monoids that are obtained by removing the order relation. Let (a, b) and (c, d) be any elements of $\mathcal{A}$, then $h((a, b) \oplus (c, d))$ is equal to $h((a, b)) + h((c, d))$:

$$h((a, b) \oplus (c, d)) = h\left(\left(\frac{ab + cd}{b + d}, b + d\right)\right) = ab + cd = h((a, b)) + h((c, d)), \tag{4.18}$$

thus h can be seen as a homomorphism of monoids.

In general, h is not a homomorphism of routing algebras since it does not preserve the order. As a counterexample, $(1, 4) \preceq (2, 1)$ while $h((1, 4)) = 4 \not\leq 2 = h((2, 1))$. When the homomorphism h preserves the order for all paths between two given nodes, then by replacing weights of edges $e \in E$, $w(e) \in \mathcal{A}$, by $h(w(e)) \in \mathcal{B}$ the problem of finding an optimal path between these nodes is reduced to the standard shortest path problem solvable by the Dijkstra's algorithm. As an example, h preserves the order when the digraph G_1 is a path. This approach can be extended to more complicated cases using the idea behind Algorithm 1 from (Zubor et al., 2014).

The optimal path problem with the weight of an edge $e \in E$ given by $h(w(e))$, disregarding the order preservation, can be interpreted as minimization of the product of the length of the drawing of a path in G_1 and the elastic distance between its drawing and the drawing of a path in G_2. This interpretation can be potentially applicable to simultaneous path planning for two robots (Švestka and Overmars, 1998) where the paths are required to be both as similar and as short as possible.

The Algorithm 3 builds the σ-product of input digraphs, translates the weights to the routing algebra $\mathcal{A}$ and invokes a procedure for finding the optimal path. This procedure can either use a standard algorithm for algebra $\mathcal{A}$ and obtain an approximate solution or, when applicable, employ the homomorphism h and perform a standard shortest path search in $\mathcal{B}$. Correctness of the Algorithm 3 in the second case is asserted by

Algorithm 3: Finding the closest sequence of σ-pairs of paths between $(v_{b,1}, v_{b,2})$ and $(v_{,2}, v_{e,2})$ in digraphs $G_1 = (V_1, E_1)$, $G_2 = (V_2, E_2)$.

Require: graphs $G_1 = (V_1, E_1)$, $G_2 = (V_2, E_2)$, drawings $\phi_1, \xi_1, \phi_2, \xi_2$, vertices $v_{b,1}, v_{e,1} \in V_1, v_{b,2}, v_{e,2} \in V_2$, discretization σ.

$V := V_1 \times V_2$ {Set of vertices V is the cartesian product of V_1 and V_2.}
$E := \{\}$ {List of edges.}
$w := \{\}$ {Weights of edges.}
for $(u_1, u_2) \in V$ **do**
  for each σ-pair of paths (p_1, p_2) starting in u_1, u_2 **do**
    $s := \sqrt{\frac{\sum_{e_2 \in p_2} \Delta t_2(e_2)}{\sum_{e_1 \in p_1} \Delta t_1(e_1)}}$
    $W := I\left(0, 1, 0, 1, \eta_1^q(p_1), \eta_2^q(p_2)s\right)$
    $v_1 :=$ end of p_1
    $v_2 :=$ end of p_2
    $e := ((u_1, u_2), (v_1, v_2))$
    if e not in E **or** $w[e] > W$ **then**
      add e to E
      $w[e] := \left(W, \sum_{e_1 \in p_1} \Delta t_1(e_1)\right)$
    end if
  end for
end for
$(p_1, p_2), c :=$ FIND-OPTIMAL-PATH $(V, E, w, (v_{b,1}, v_{b,2}), (v_{e,1}, v_{e,2}))$ {Compute optimal path and its cost.}
return p_1, p_2, c

Theorem 1. In particular it follows from this theorem that when both graphs G_1, G_2 are paths Algorithms 1 and 3 return the same results.

Theorem 1. *If the optimal path finding procedure is exact, then among all sequences of σ-pairs of paths in digraphs G_1, G_2 between $(v_{b,1}, v_{b,2})$ and $(v_{e,1}, v_{e,2})$ the one returned by Algorithm 3 has the minimum elastic distance.*

This theorem is proved in Appendix B.

Computational complexity

In this analysis of computational complexity it is assumed that the optimal path finding is performed using the (generalized) Dijkstra's algorithm. The Algorithm 3 is given two planar digraphs $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$ as well as the set σ. The Euler's formula states that the number of edges in such digraphs is linearly bound by the number of vertices. The time complexity of the first step, that is building the σ-product digraph, is $O(|V_1||V_2| + k(\sigma, G_1, G_2))$ where the number of elements in a set V is denoted by $|V|$ and $k(\sigma, G_1, G_2)$ is the number of σ-pairs of paths in digraphs G_1 and G_2. To see how fast does $k(\sigma, G_1, G_2)$

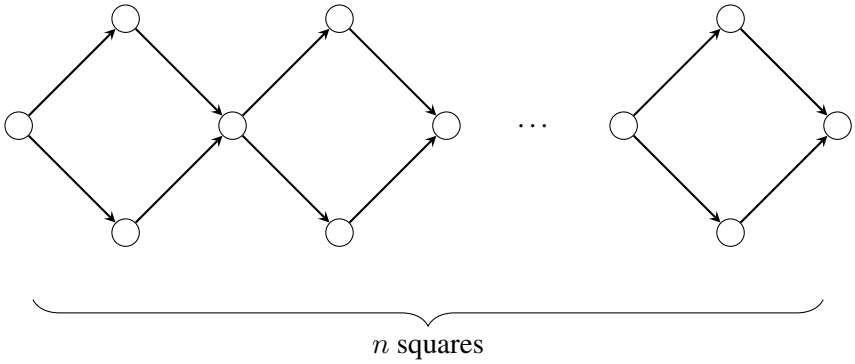


Figure 4.6: Graphical representation of a digraph $G_R(n)$

grow, let us first observe that

$$k(\sigma, G_1, G_2) = k\left(\bigcup_{i=1}^{|\sigma|}\{(k_{1,i}, k_{2,i})\}, G_1, G_2\right) = \sum_{i=1}^{|\sigma|} k\left(\{(k_{1,i}, k_{2,i})\}, G_1, G_2\right), \tag{4.19}$$

because each pair of paths has a unique pair of lengths.

It is difficult to characterize the asymptotic behaviour of $k(\{(k_1, k_2)\}, G_1, G_2)$. Firstly, for fixed G_1 and G_2 the value of $k(\{(k_1, k_2)\}, G_1, G_2)$ is bounded as there are no paths with more than $|V_i| - 1$ edges in G_i for $i = 1, 2$. One possible solution is to replace fixed digraphs by certain families of digraphs parametrized by an integer, namely $k(\{(k_1, k_2)\}, G_1(n_1), G_2(n_2))$ where $n_1, n_2 \in \mathbb{N}$. Now an upper bound can be found by taking $G_i(n_i)$ to be the complete digraph with n_i nodes, $K_d(n_i) = (\{1, 2, \ldots, n_i\}, \{(u, v)\colon u, v \in \{1, 2, \ldots, n_i\}, u \neq v\})$, for $i = 1, 2$. By counting paths one can obtain

$$k(\{(k_1, k_2)\}, K_d(n_1), K_d(n_2)) = \frac{n_1! n_2!}{(n_1 - k_1 - 1)!(n_2 - k_2 - 1)!}, \tag{4.20}$$

which in general grows faster than exponentially, although full digraphs with more than four vertices are not planar (Kuratowski, 1930)[1]. A useful lower bound can be obtained using digraphs G_R depicted in Figure 4.6. After substituting these digraphs and selecting $k_1 = 2n_1$ and $k_2 = 2n_2$ the value of k is

$$k(\{(2n_1, 2n_2)\}, G_R(n_1), G_R(n_2)) = 2^{n_1 + n_2}. \tag{4.21}$$

As shown, k can grow at least exponentially with values in pairs in σ. This strongly constraints σ for practical applications. However, the results described in Section 4.2 indicate that in practice it is not a significant limitation.

The second step of Algorithm 3 performs shortest path search in the digraph $G = G_1 \times_\sigma G_2 = (V, E)$. Using the Fibonacci heap-based Dijkstra's algorithm (Fredman and Tarjan, 1987) the time complexity of

[1] Planarity of a digraph is understood as planarity of the underlying graph.

this step is $O(|E| + |V|\log(|V|))$. Since the digraph G has no more than $|E| \leq \min(|V|^2, k(\sigma, G_1, G_2))$ edges, the total time complexity of Algorithm 3 is $O(|V|\log(|V|) + k(\sigma, G_1, G_2))$.

This is a significant improvement over state-of-the-art approaches, as they would require separate comparison of each pair of paths in G_1, G_2. A well-known result in graph theory states that just counting the number of paths between two vertices is a $\#\mathrm{P}$-complete problem (Valiant, 1979), which makes the approach feasible only for very simple graphs.

Experiments

Pattern recognition is an important task of image analysis. There are many different approaches to detecting shapes but it is sometimes very challenging to achieve satisfactory accuracy in localization of shape boundary (Hum et al., 2014, Wojciechowski et al., 2016). This results in prevalence of application-oriented algorithms. Algorithm 3, in conjunction with superpixel segmentation (see Section 2.4.4), offers a new, well-motivated solution.

In this application, the drawing of a digraph G_1 is a representative shape of the object that is to be found in a given image. The nodes of G_1 are points obtained by sampling the representative shape, while the edges connect neighbouring points ordering them in a preselected direction. The digraph G_2 is the result of orienting edges both ways in the graph of the superpixel segmentation applied to the gradient magnitude of the analysed image. The boundary is required to be an open curve, so the algorithm is best suited to partially occluded objects where the occlusion occurs at a known place.

The superpixel segmentation is obtained using the watershed from markers algorithm (Meyer and Beucher, 1990). It was selected for its good adherence to the edges of the original image and the possibility to constrain the size of a superpixel. Markers are placed at pixels that are the local minima of the gradient magnitude image in a square $2n_m + 1 \times 2n_m + 1$ neighbourhood centred on the considered pixel where n_m is a certain constant.

Conversion of a superpixel segmentation to a digraph is performed by constructing a mask where each pixel of the original image is marked either black or white. White pixels constitute borders between superpixels while black pixels are their interior. Groups of white pixels with at least three white neighbours in the Moore neighbourhood (the 8-neighbourhood) are nodes of the superpixels digraph. White pixels with two white neighbours are the other nodes of that digraph. Neighbouring nodes are connected with edges in both directions. Drawing of such a digraph is naturally defined by the image. The drawing of each edge is assumed to be a straight line.

The following types of images were used in the experiments: standing frontal X-ray images of the knee, standing frontal X-ray images of the hip bone area, images of distorted 20-corner stars with a single corner removed and images of a clothes iron. Each image has approximately 1.0 to 1.5 megapixels. Figure 4.7

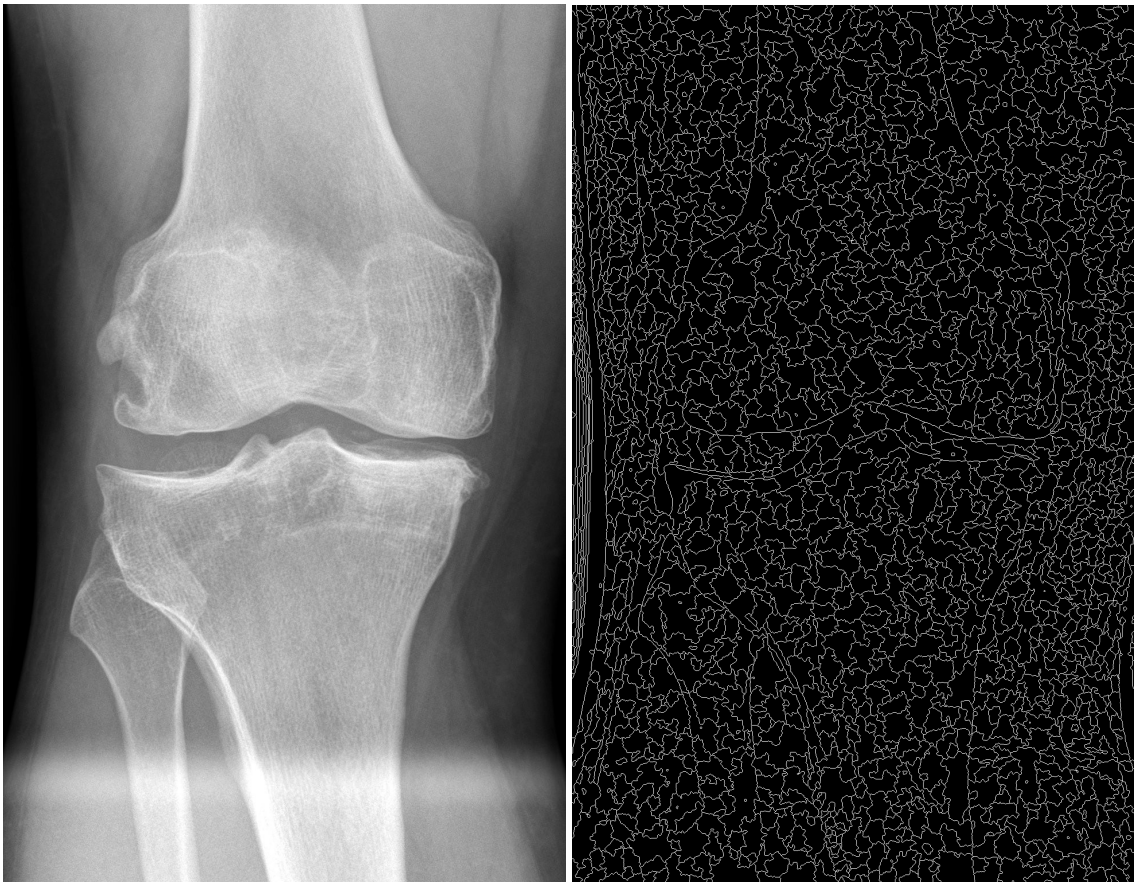


Figure 4.7: A sample image and its superpixel segmentation using the watershed from markers algorithm

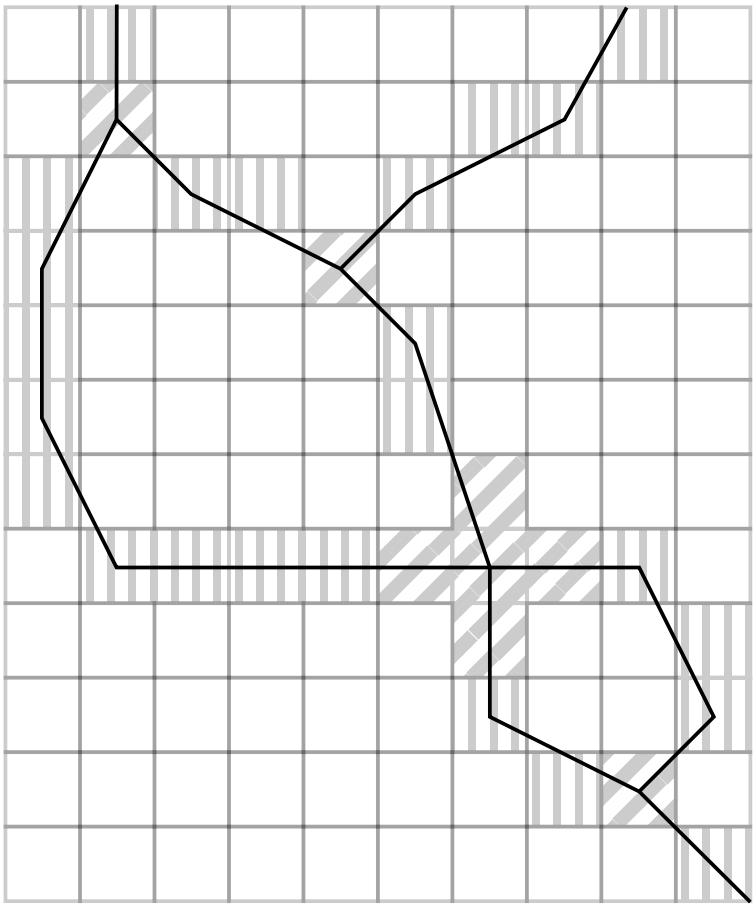


Figure 4.8: A fragment of an image where each pixel corresponds to a square. Grey vertically striped squares are the pixels with two white neighbours and squares with slanted grey stripes have at least three white neighbours. The black lines correspond to the drawing of a superpixel digraph with every second vertically striped pixel removed

shows a sample image from the set and its watershed from markers segmentation with $n_m = 15$. To reduce resource utilization every second pixel was removed from the target pattern and every second node with two white neighbours was removed from the superpixel digraph (see Figure 4.8). The values of Δt follow a constant-speed parametrization of the target boundary. Since the digraph G_1 is a path, the generalized Dijkstra's algorithm returns an exact solution to the optimization problem. In the experiments the set σ was equal to $\{(1,1),(1,2),(2,1),(1,3),(2,3),(3,1),(3,2),(1,4),(3,4),(4,1),(4,3)\}$. The resulting σ-product digraphs have about 10^8 nodes and 10^9 edges.

Figures 4.9 and 4.10 depict a few samples of segmentation results in different datasets. High accuracy of detection has been achieved in most cases. The algorithm needs about 20 to 30 seconds on an Intel Core i7 CPU and about 7 to 12 GB of RAM for processing a single image. When part of the object boundary has poor contrast, as in Figure 4.10a, the most accurate path in the superpixel digraph is noisy. Due to the differentiation in the definition of SRV transform (see Equation (2.10)), such noise significantly increases the elastic distance. Appropriate preprocessing of the input image reduces this issue. In all datasets except the knee dataset the original image and gradient magnitude image were smoothed using anisotropic diffusion (Perona and Malik, 1990). Using a different superpixel algorithm with compactness control could decrease the problem at the cost of worse boundary adherence. Additionally, the proposed algorithm works well even for objects with complex boundary (Figure 4.10d) and partially occluded objects (Figure 4.10f). Figure 4.10h shows a case where the algorithm did not find the correct shape because the boundary of the edges in the original image was not pronounced enough to result in smooth superpixel boundaries.

The proposed approach has been compared with the Active Appearance Model (AMM) algorithm (Cootes et al., 2001) using the set of X-ray knee images. The AAM algorithm has been trained on a subset of 30 images and tested on 20 different images. The described elastic metric-based algorithm was supplied with a reference knee shape from the set of training images and tested on the same set of images as the AAM algorithm. The mean Dice coefficient (Dice, 1945), used to compare automatic and reference segmentations, is equal to 0.986 for the AAM algorithm and 0.881 for the proposed method.

Conclusions

In this chapter an extension of the framework of Elastic Shape Analysis to pairs of digraph drawings is described. An algorithm for finding pairs of paths in digraphs with closest drawings is described and analysed. It was shown to be efficient for a class of digraphs important to an application in pattern recognition, whereas the state-of-the-art approach of considering each pair of paths separately is feasible only for very simple graphs.

Experimental verification confirmed that the described algorithm is applicable to boundary delineation. In this version the only information about the image that the method uses is its superpixel segmentation.

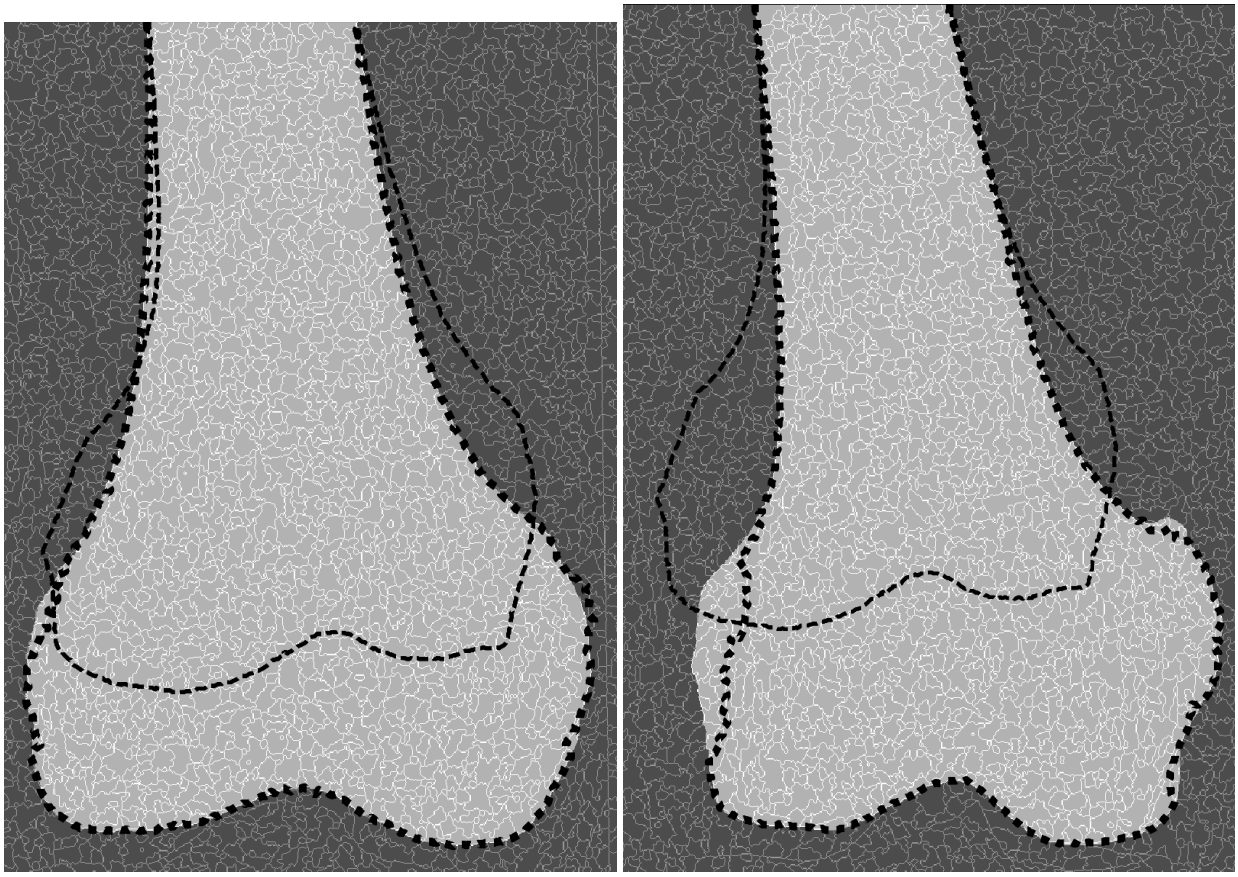
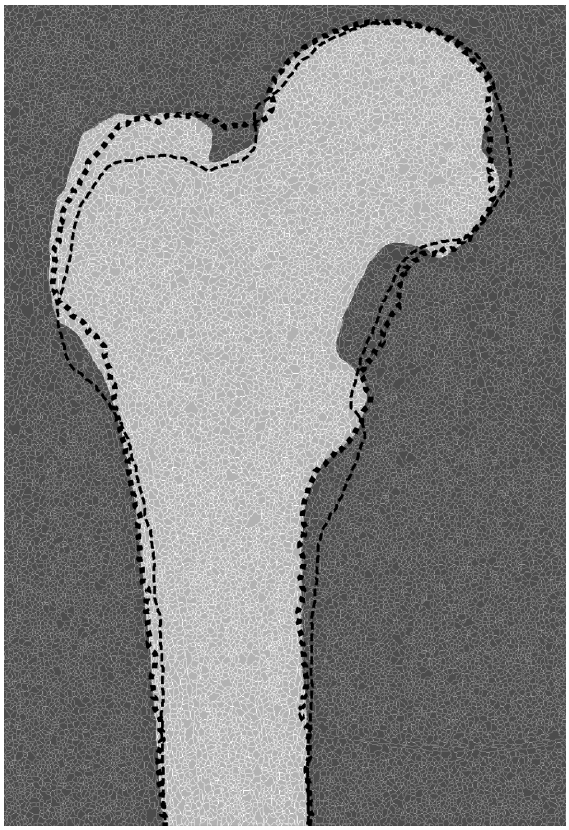
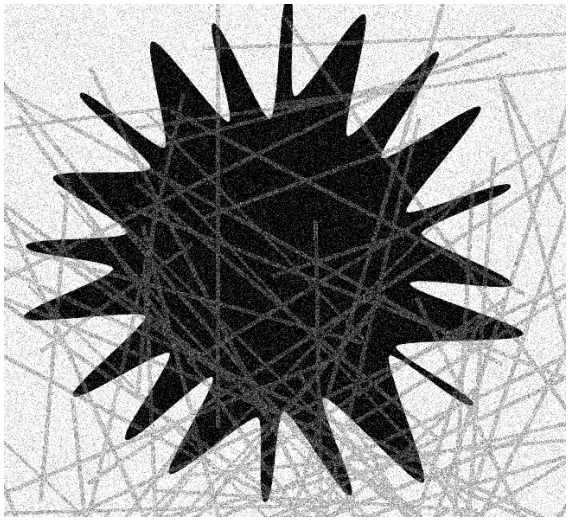


Figure 4.9: Sample results of pattern recognition using the described algorithm. White lines correspond to the superpixel segmentation, light and dark grey area represent the reference segmentation, dashed lines are patterns the algorithm matches against and the dotted line is the boundary found using the described method

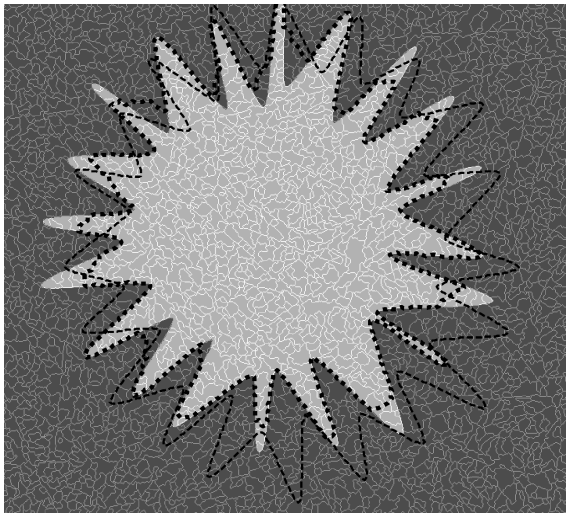
(a) Upper extremity of a femur



(b) Segmentation result



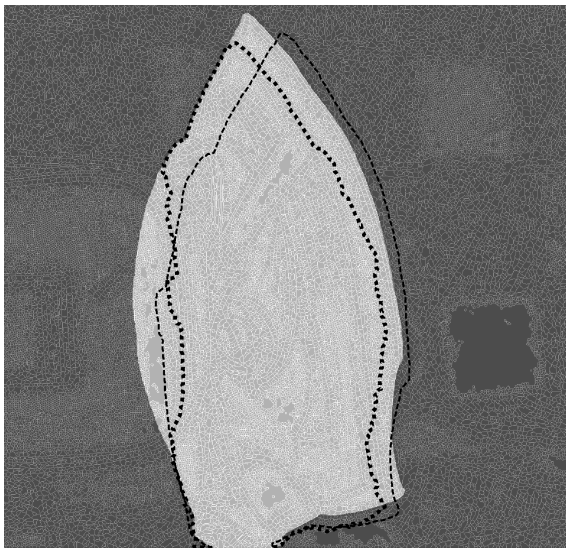
(c) A star-like shape



(d) Segmentation result

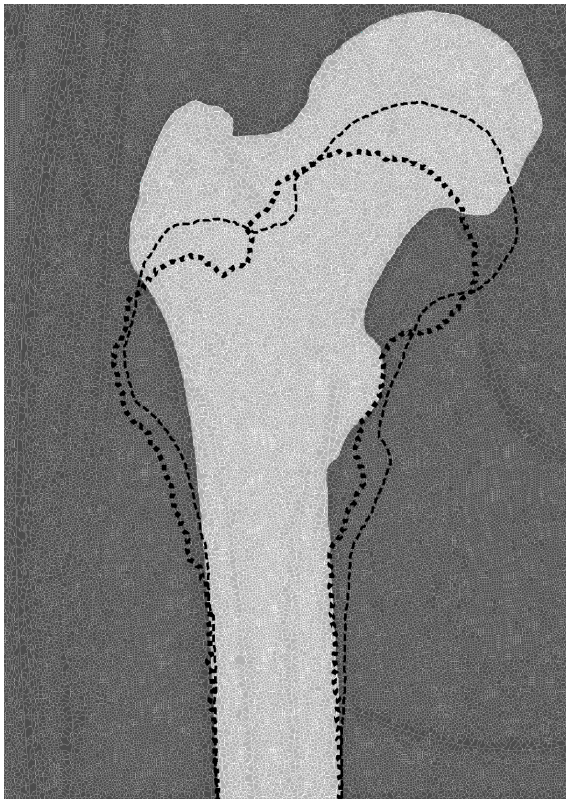
Figure 4.10: Recognition of several types of images. Left pictures are parts of input images containing searched objects. Their segmentations obtained using Algorithm 3 are presented on the right hand side. The same symbols are used as in Figure 4.9

(e) A partially occluded clothes iron



(f) Segmentation result

(g) Upper extremity of a femur



(h) Segmentation result

Figure 4.10: Recognition of several types of images. Left pictures are parts of input images containing searched objects. Their segmentations obtained using Algorithm 3 are presented on the right hand side. The same symbols are used as in Figure 4.9

It leaves some space for future improvements to the algorithm which could use more complicated cost functions, combining both more features of the image and a comprehensive statistical model of the searched shape. However even at this stage good accuracy has been achieved for images of different types.

The new algorithm, though, is not a competitor to existing segmentation algorithms. It is a framework for constructing segmentation algorithms of a new type. A competing algorithm would have to take into account other features of input images which would be incorporated into a more complex routing algebra. A statistical model should be used to describe representative shapes. Additionally a variant for closed shapes needs to be developed.

There are still a few open questions:

- Would it be feasible to perform a search without fixing the endpoints?
- Could the path straightening method of calculating the elastic distance (Srivastava et al., 2012) be extended to digraphs?
- What the guidelines for obtaining the superpixel segmentation should be?

However, considering solid theoretical foundations and encouraging experimental results, the new method is a good basis for developing new pattern recognition methods. The generality of the solved problem makes the algorithm likely to also find applications in other fields. Possible application to robot path planning is explained in Section 4.1.2. In syntactic pattern recognition and shape retrieval (Mehrotra and Gary, 1995, Tagougui et al., 2013) a collection of shapes, for example described by a set of chain codes (Freeman, 1961) or a grammar, can also be represented by a graph whose drawing represents superimposition of these shapes. It is also possible to apply the developed algorithm in alternative plan search (Felner et al., 2003). As demonstrated, the presented algorithm extends the applicability of an elastic metric to new problems.

Chapter 5

Machine learning approach to superpixel-constrained segmentation

This chapter describes a modification of the approach described in Chapter 4. The primary difference is introduction of a statistical model that describes the probability of a particular edge from the superpixels graph being a part of the contour (see Figure 4.8). This is modelled using probabilistic classifiers and probabilities obtained in this way are used to augment the edge weights in the σ-product of graphs used in Algorithm 3.

Section 5.1 describes the applied machine learning methodology and algorithm modifications. Section 5.2 details experiments conducted to assess the new method. Results of the experiments are discussed in Section 5.3 and the conclusions are drawn in Section 5.4.

Methods

There are two novel ideas that need to be developed to apply machine learning to the problem of image segmentation. First, a feature extraction procedure that describes edges in a superpixel graph with a sequence of numbers is needed. This is described in Subsection 5.1.1. Next, well-known classification algorithms are applied to the datasets obtained this way as described in Section 5.2. Finally, the classifiers are used to define a new weight function for the σ-product graph as described in Subsection 5.1.2.

Feature extraction for machine learning

One of the current trends in machine learning is automatic feature extraction, or representation learning. This is featured in recent work on deep learning (LeCun et al., 2015, Krizhevsky et al., 2012), including for example manifold learning (Lin and Zha, 2008, Lunga et al., 2014). Such methods, however, require a large amount of input data and the success of classification algorithms is generally dependent on using a right

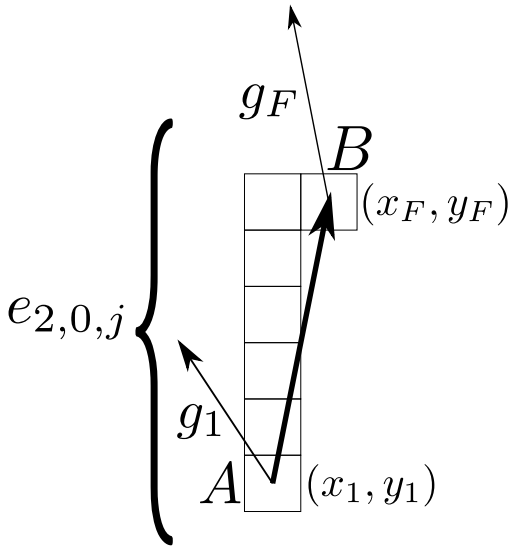


Figure 5.1: First step of feature extraction.

feature representation (Bengio et al., 2013). A custom feature extraction procedure is designed to make the algorithm faster and applicable to small sets of images.

Consider an edge $e \in E$ of the σ-product digraph $G_1 \times_\sigma G_2 = (V, E, w_N)$ where G_2 is constructed from a superpixel segmentation of an analysed image, as in Section 4.2. The edge e corresponds to a σ-pair $((e_{1,0,j})_{j=1}^{n_{1,0}}, (e_{2,0,j})_{j=1}^{n_{2,0}})$ that minimizes the value of Equation (4.2) for e. For each edge $e_{2,0,j}$ where $j = 1, 2, \ldots, n_{2,0}$ a number of features are determined.

Let $(x_1, y_1), (x_2, y_2), \ldots, (x_F, y_F)$ be the F coordinates of pixels that correspond to $e_{2,0,j}$, in the order determined by the drawing $\xi_{k,2}(e_{2,0,j})$, that is (x_1, y_1) is the pixel nearest to $A = \xi_{k,2}(e_{2,0,j})(0)$ and (x_F, y_F) is the pixel nearest to $B = \xi_{k,2}(e_{2,0,j})(1)$. At each pixel (x_j, y_j) for $j = 1, 2, \ldots, F$ a gradient vector g_j is calculated using the second order central difference formulas (see Figure 5.1). Each vector g_j is then expressed in terms of polar coordinates (r_j, θ_j) and a mean gradient vector g, expressed as (r, β) in polar coordinates, is calculated as follows:

$$r = \frac{1}{F} \sum_{j=1}^{F} r_j, \tag{5.1}$$

$$\beta = \arg \sum_{j=1}^{F} e^{i\theta_j}. \tag{5.2}$$

In the next step, the vector $\overrightarrow{AB}$ is cast onto a reference contour represented by points $(p_i)_{i=0}^{M-1}$. This is performed by separately casting points A and B onto the contour. Each of these points is cast using Algorithm 4. Casting the point A results in a point A' and a number t_A while casting the point B results in a point B' and a number t_B (see Figure 5.2). If the points A' and B' are different, then a vector $v = \overrightarrow{A'B'}$ is calculated. Otherwise the vector v is taken to be the forward first order finite difference tangent vector to the

closed polygonal polygonal chain $(p_i)_{i=0}^{M-1}$. The angles formed by expressing vectors $\overrightarrow{AB}$ and v are called, respectively, α and γ (both are marked on Figure 5.2).

Algorithm 4: Casting a point P onto a reference contour represented by points $(p_i)_{i=0}^{M-1}$.

Require: Point P, contour represented by points $(p_i)_{i=0}^{M-1}$.
$p_M \leftarrow p_0$
$p_c \leftarrow undefined$
$d \leftarrow \infty$
for $i \in \{0, 1, \ldots, M-1\}$ **do**
  if angles between $\overrightarrow{p_iP}$ and $\overrightarrow{p_ip_{i+1}}$ and between $\overrightarrow{p_{i+1}P}$ and $\overrightarrow{p_{i+1}p_i}$ are acute **then**
    $P_i' \leftarrow$ point from the line segment p_ip_{i+1} that is closest to P.
    $d_i \leftarrow \|P_i' - P\|$
    if $d_i < d$ **then**
      $p_c \leftarrow P_i'$
      $d \leftarrow d_i$
    end if
  end if
end for
if p_c is $undefined$ **then**
  for $i \in \{0, 1, \ldots, M-1\}$ **do**
    $d_i \leftarrow \|p_i - P\|$
    if $d_i < d$ **then**
      $p_c \leftarrow p_i$
      $d \leftarrow d_i$
    end if
  end for
end if
$t \leftarrow$ the distance between p_0 and p_c along the closed polygonal chain defined by the sequence of points $(p_i)_{i=0}^{M}$, in the direction indicated by indices.
return p_c, t

Finally, for each edge $e_{2,0,j}$ the following features are considered as input for the machine learning algorithms:

- The norm of mean gradient vector, $\|g\| = r$.
- A number indicating the position of the edge $e_{2,0,j}$ along the polygonal chain $(p_i)_{i=0}^{M-1}$, $t_{mean} = \frac{t_A+t_B}{2}$.
- The angles $\gamma - \alpha$ and $\gamma - \beta$.
- The mean distance $d_{mean} = \frac{\|A-A'\|+\|B-B'\|}{2}$.

Thus the feature extraction defines a function $f\colon E_2 \times \mathbb{R}^{2M} \to \mathbb{R}_{\geq 0} \times \mathbb{R}_{\geq 0} \times [0, 2\pi) \times [0, 2\pi) \times \mathbb{R}_{\geq 0}$ that transforms an edge in the graph G_2 given a reference contour to a sequence of feature values:

$$f(e_{2,0,j}, (p_i)_{i=0}^{M-1}) = (r, t_{mean}, \gamma - \alpha, \gamma - \beta, d_{mean}). \tag{5.3}$$

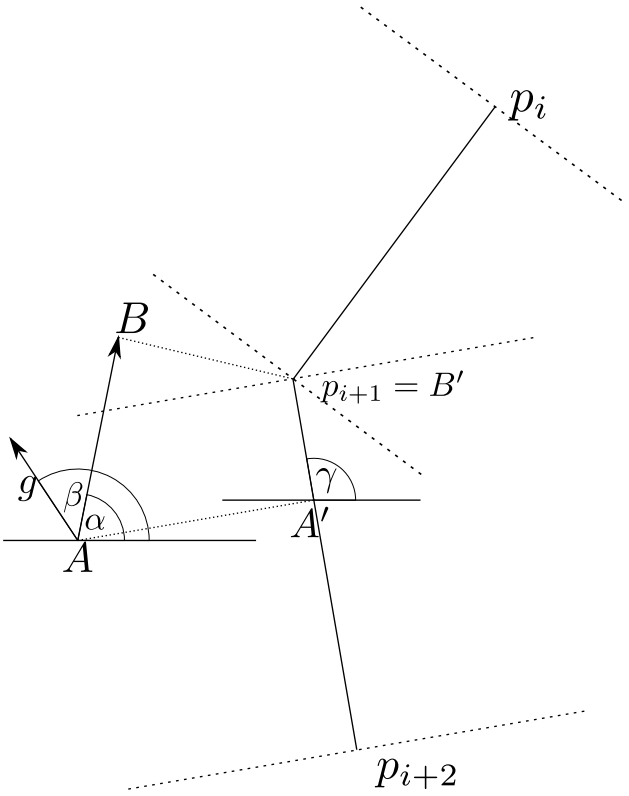


Figure 5.2: Second step of feature extraction.

A classifier c, $c\colon \mathbb{R}_{\geq 0} \times \mathbb{R}_{\geq 0} \times [0, 2\pi) \times [0, 2\pi) \times \mathbb{R}_{\geq 0} \to [0, 1]$ assigns the probability that an edge with given feature values does not belong to the real object boundary. A set of standard classification algorithms was considered as described in Section 5.2.

Modified edge weight function

The new weight function $w_N\colon E \to \mathbb{R}$ for the σ-product $G_1 \times_\sigma G_2 = (V, E, w_N)$ is given by

$$w_N(e) = h(w(e)) + \frac{w_{ML}}{n_{2,0}} \sum_{j=1}^{n_{2,0}} c(f(e_{2,0,j}, (p_i)_{i=0}^{M-1})), \tag{5.4}$$

where h is given by Equation (4.17), w is given by Equation (4.3), $w_{ML} \in \mathbb{R}_{\geq 0}$ is a parameter, the edge e corresponds to a σ-pair $((e_{1,0,j})_{j=1}^{n_{1,0}}, (e_{2,0,j})_{j=1}^{n_{2,0}})$ that minimizes the value of Equation (4.2) for e and $(p_i)_{i=0}^{M-1}$ is a sequence of points that correspond to the reference contour. If the drawing of G_1 is given by ξ_1 and ϕ_1 (see Section 4.1.1) and $G_1 = (V_1, E_1)$ is a path such that $V_1 = \{v_0, v_1, \ldots, v_{M-1}\}$ and $E_1 = \{(v_0, v_1), (v_1, v_2), \ldots, (v_{M-2}, v_{M-1})\}$, then $p_i = \phi_1(v_i)$ for $i = 0, 1, \ldots, M-1$.

Experiments

Two sets of images were used in the experiments: 50 standing frontal X-ray images of the knee and 30 standing frontal X-ray images of the hip bone area. Each image has approximately 1.0 to 1.5 megapixels. Images of the hip bone area were smoothed using an anisotropic diffusion filter (Perona and Malik, 1990) to reduce noise. In preprocessing, all images were rescaled to 50% of original width and height. The same superpixel segmentation approach was used as described in Section 4.2, with n_m equal to 10.

Selection of classification method

The following classification methods were compared:

- Naive Bayes (Russell and Norvig, 2009), denoted NB,
- Random Forest Classifier (Breiman, 2001), denoted RF,
- Random Trees embedding (Geurts et al., 2006) of original data, followed by Logistic Regression (Yu et al., 2011) on the transformed feature space, denoted RT+LR,
- Random Forest Classifier-based feature transformation, followed by the One Hot Encoder and Logistic Regression on the transformed feature space, denoted RF+LR,
- Gradient Boosting Classifier (Friedman, 2001), denoted GRD,
- Gradient Boosting Classifier-based feature transformation, followed by the One Hot Encoder and Logistic Regression on the transformed feature space, denoted GRD+LR.

Applying Support Vector Machine with Platt's scaling (Platt, 1999) for obtaining class probabilities was also attempted but this method was rejected due to very poor performance. In cases where Logistic Regression was applied, the training set was split into two equal parts: one for training the classifier performing the embedding into a higher-dimensional feature space and the second half for training the Logistic Regression classifier.

The machine learning was performed using algorithms from the scikit-learn library (Pedregosa et al., 2011). The number of trees in Random Forest and Random Trees classifiers was equal to 300, maximum depth of trees was equal to 6, the number of boosting stages for the Gradient Boosting Classifier was equal to 300 and other parameters were set to their default values.

Testing of image segmentation

To reduce resource utilization every third pixel was retained from the target pattern and every third node with two white neighbours was retained from the superpixel digraph (see Figure 4.8). The values of Δt follow a constant-speed parametrization of the target boundary. In the experiments the set σ was equal to $\{(1,1),(1,2),(2,1),(1,3),(2,3),(3,1),(3,2),(1,4),(3,4),(4,1),(4,3)\}$.

Table 5.1: AUC coefficients for ROC curves depicted in Figures 5.3 and 5.4.

dataset	knee	hip bone
NB	0.9917	0.9887
RF	0.9943	0.9911
RT+LR	0.9953	0.9920
RF+LR	0.9953	0.9921
GRD	0.9974	0.9944
GRD+LR	0.9968	0.9954

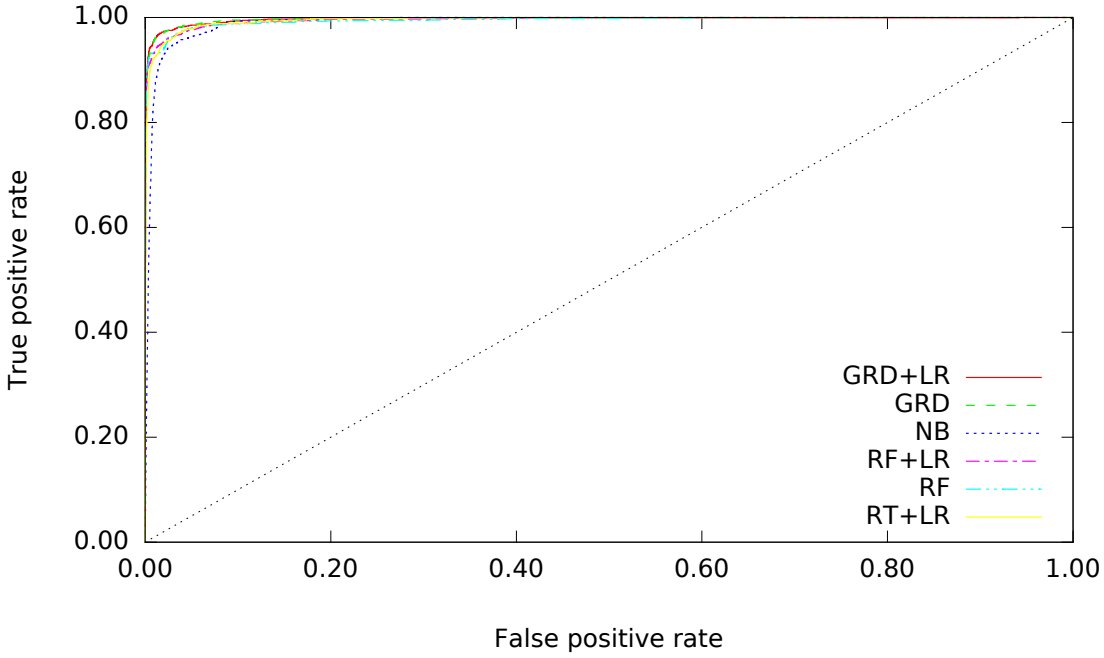
Ten-fold cross-validation was applied to evaluate the proposed algorithm. A given set of images was divided into ten parts with equal number of images. In each fold of the cross-validation images from nine of the sets are used to train probabilistic classifiers and find the optimal value of the w_{ML} parameter. The tenth set is used to assess performance of the method.

Optimization of the w_{ML} parameter is performed using the SciPy (Jones et al., 2001–) brute function on a uniform n_{opt}-point grid on the interval $[0, 0.2]$ with the default local optimization function. The parameter n_{opt} was equal to 11 for both sets of images. The optimization objective is maximization of the average Dice coefficient over all test images, each one analysed using five different contours selected from reference segmentations of training images.

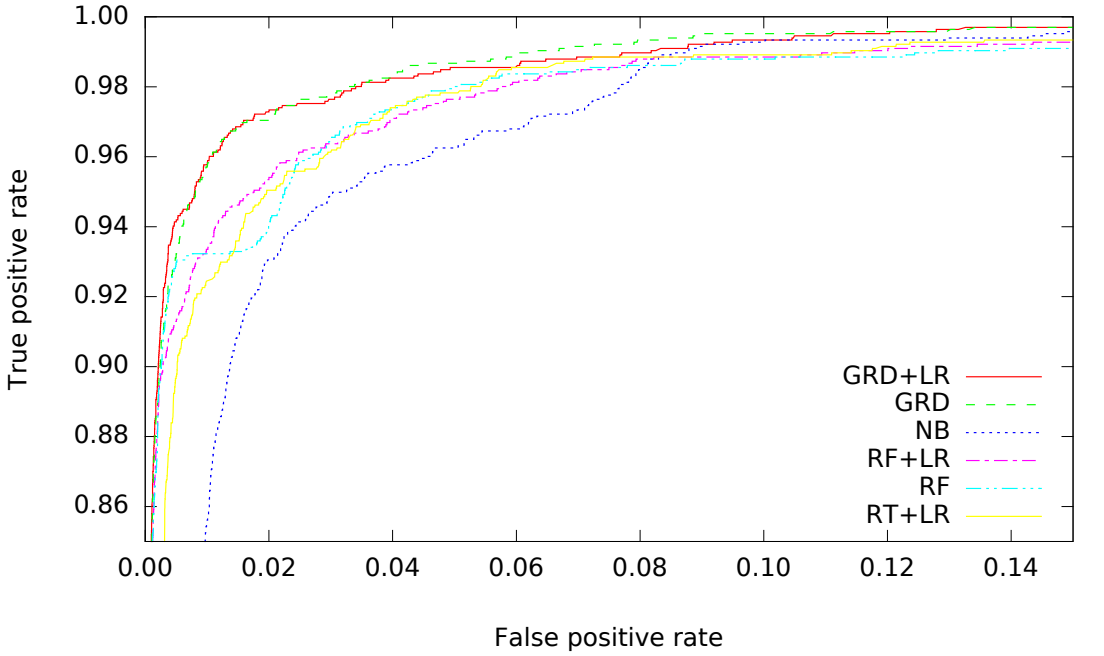
Results

Figures 5.3 and 5.4 depict ROC curves of classifiers described in Section 5.2.1 tested on, respectively, knee and hip bone datasets. The ROC curves were obtained by varying the threshold $T \in [0, 1]$ for classifying an edge $e_{2,0,j}$ as not belonging to the contour, that is when $c(f(e_{2,0,j}, (p_i)_{i=0}^{M-1})) \geq T$. For both knee and hip bone datasets the data extracted from training images for the first fold of cross-validation were used. 80% of the training data was used to train classifiers and 20% was used to construct the ROC curves. As indicated in Table 5.1, the NB classifier has the lowest area under the ROC curve (AUC) while GRD and GRD+LR classifiers have the highest AUC on both datasets. Based on these results, the GRD+LR classifier was selected for edge classification.

Figure 5.7 shows box plots for different values of the w_{ML} parameter for both datasets of images. For each value of w_{ML} a full cross-validation was performed as described in Section 5.2.2. The box plots illustrate statistical distributions of the mean Dice coefficient across different folds. Horizontal lines in each plot denote respectively, from bottom to top, minimum, first quartile, median, third quartile and maximum value of mean Dice coefficients for a given value of w_{ML}. Minimum values and first quartiles as well as third quartiles and maximum values are connected by single centred vertical lines. First and third quartiles are

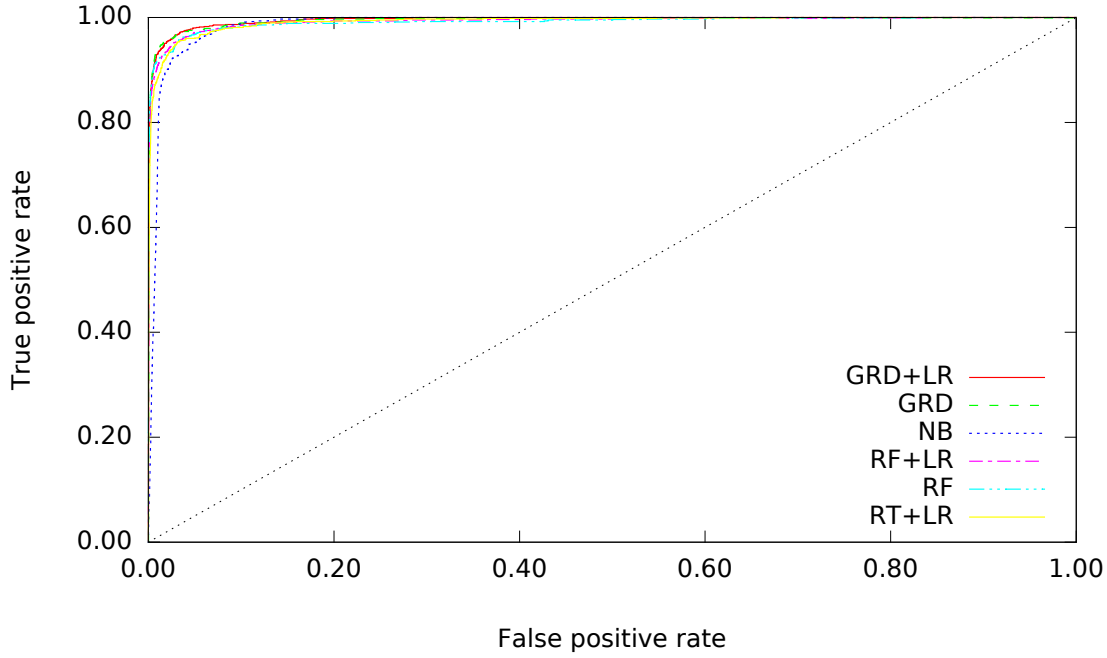


(a)

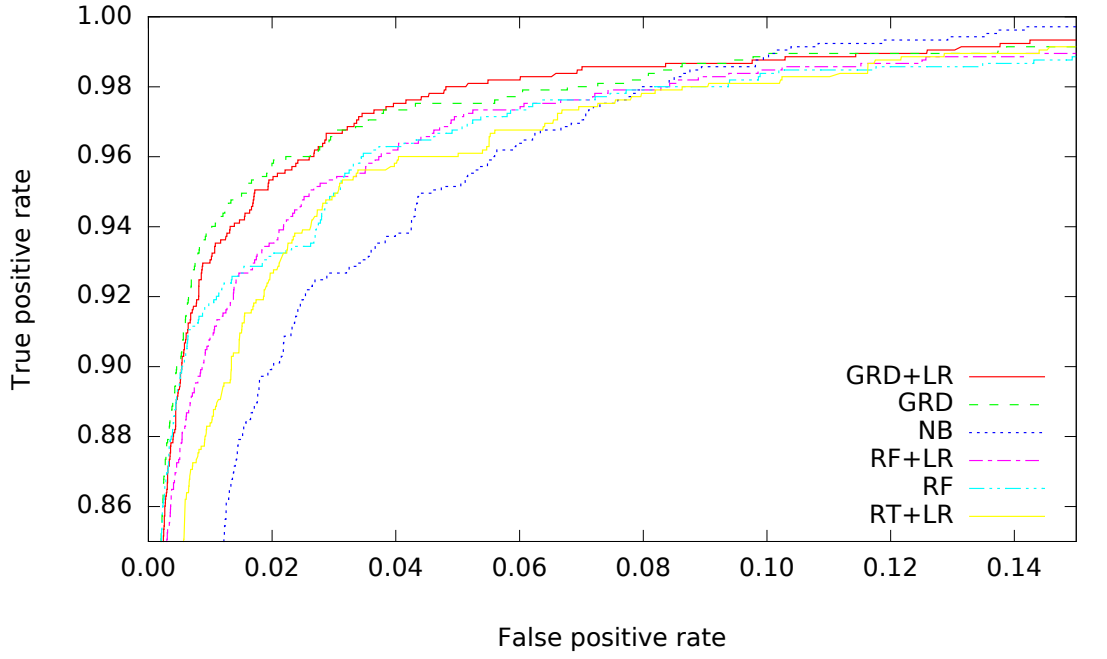


(b)

Figure 5.3: ROC curves of the tested machine learning algorithms (see Section 5.2.1) on the first fold of cross-validation on the knee dataset, with 80% of training data used to train classifiers and 20% used to construct the ROC curves. Subfigure 5.3a depicts entire ROC curves and Subfigure 5.3b shows the top left corner of the ROC plot.



(a)



(b)

Figure 5.4: ROC curves the tested machine learning algorithms (see Section 5.2.1) on the first fold of cross-validation on the hip bone dataset, with 80% of training data used to train classifiers and 20% used to construct the ROC curves. Subfigure 5.4a depicts entire ROC curves and Subfigure 5.4b shows the top left corner of the ROC plot.

additionally connected by pairs of vertical lines that form rectangles. Squares on plots in Figure 5.7 denote the average of lowest Dice coefficients obtained in different folds while triangles stand for the lowest Dice coefficient across all folds, for the value of w_{ML} indicated by the horizontal axis.

Sample results of the machine learning classifier are depicted in Figure 5.5. The edges of superpixel segmentation are coloured according to the probability of belonging to the shape boundary, as predicted by classifiers constructed in the first fold of cross-validation. Images used in this Figure were taken from respective test sets. Predicted probabilities closely match true boundaries of the objects in selected images. In certain areas the predictions are less accurate due to for example lower edge gradient or distracting features in the image.

Local optimization of the w_{ML} parameter resulted in a small further improvement. In the knee dataset, the average Dice coefficient (averaged across all folds of cross-validation) increased by 5.483×10^{-5}, and by 8.518×10^{-5} for the hip bone dataset. For each fold, the difference in average Dice coefficient ranged between 4.000×10^{-7} and 5.972×10^{-4}.

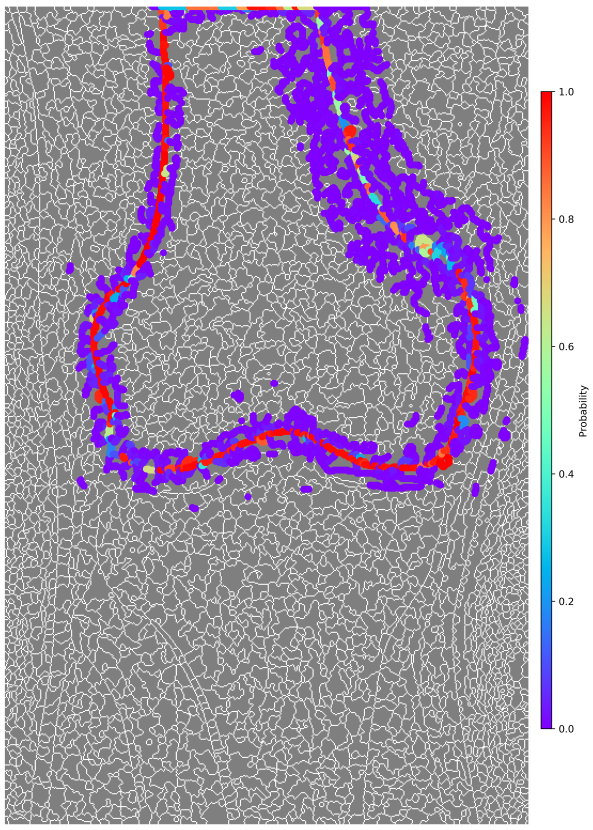
The performance of the machine learning solution is satisfactory. It takes about 90 to 100 seconds to construct a GRD+LR classifier with approximately $267\,000$ edges in the training set. Calculating probabilities for edges in a single image using the constructed classifier takes one to two seconds. The time it takes to perform shape matching using Algorithm 3 for graph with modified edge weights varies between 52 and 121 seconds per image for the knee dataset and between 98 and 227 seconds per image for the hip bone dataset. The time is typically lower for the optimal value of w_{ML} than for the pure ESA algorithm (the case when $w_{ML} = 0$). The tests were performed on an Intel Core i7 processor.

Conclusions

The obtained results indicate that the proposed approach to augmenting the Elastic Shape Analysis-based image segmentation with additional parameters computed using probabilistic classification algorithms significantly improve obtained results. As indicated in Figure 5.7, the new approach offers significantly better average and worst-case performance than the pure ESA-based algorithm while taking a comparable amount of time to process a single image. What is also important, decreased Dice coefficients for large values of w_{ML} show that the weight component originating from probabilistic classification alone would not be enough to achieve satisfactory results. Combining ESA-based pattern matching and machine learning is necessary to achieve accurate boundary delineation.

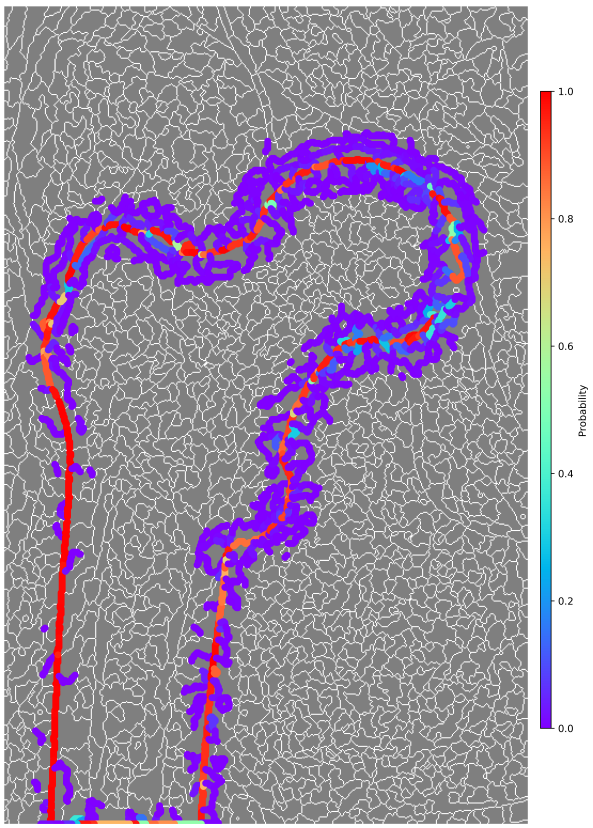
The presented algorithm is not fully invariant with respect to scaling and rotations. This issue is inherited from the framework developed in Chapter 4, where the relative rotation and scale between the given pattern and shape to be found in an image is selected in an arbitrary way. Nevertheless, the algorithm proved to be quite robust to non-optimal selection scale and rotation, as shown e.g. in Figure 4.9. In certain cases (see

(a)



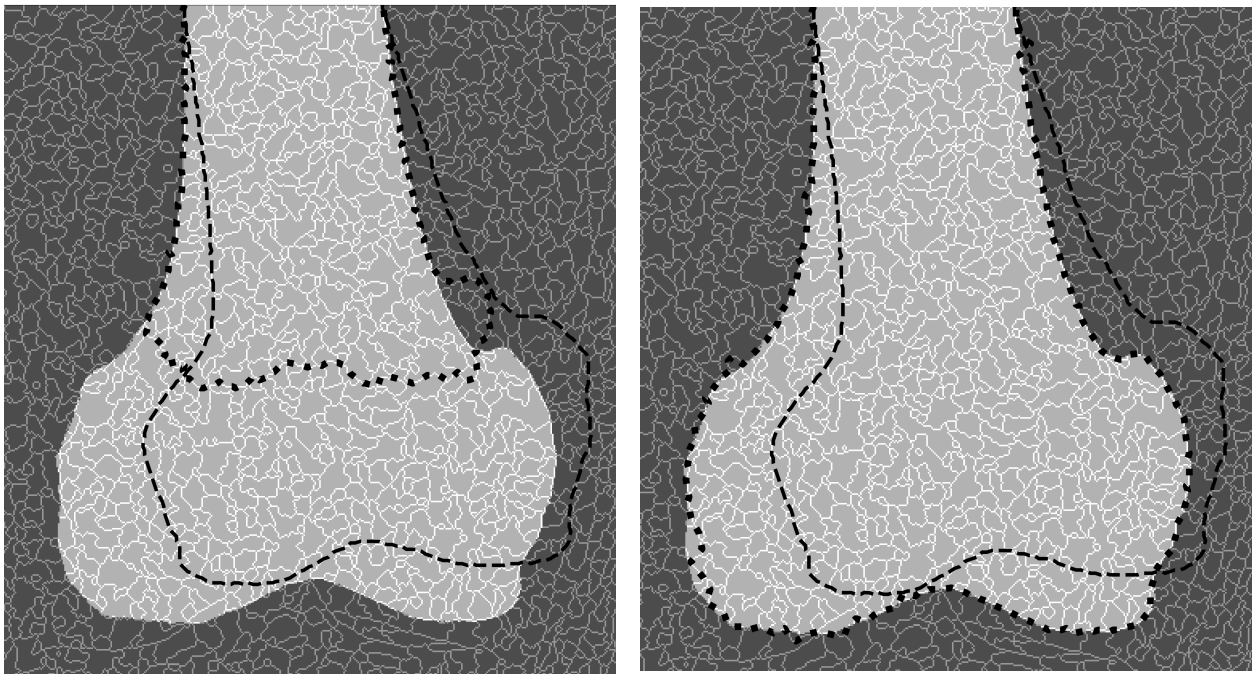
(b)

(c)



(d)

Figure 5.5: Sample segmentations of images from the knee dataset (Subfigures 5.5a and 5.5b) and the hip bone dataset (Subfigures 5.5c and 5.5d). Edges with probability of belonging to the shape boundary higher than 0.001 are coloured according to the probability, as denoted by the vertical bars.

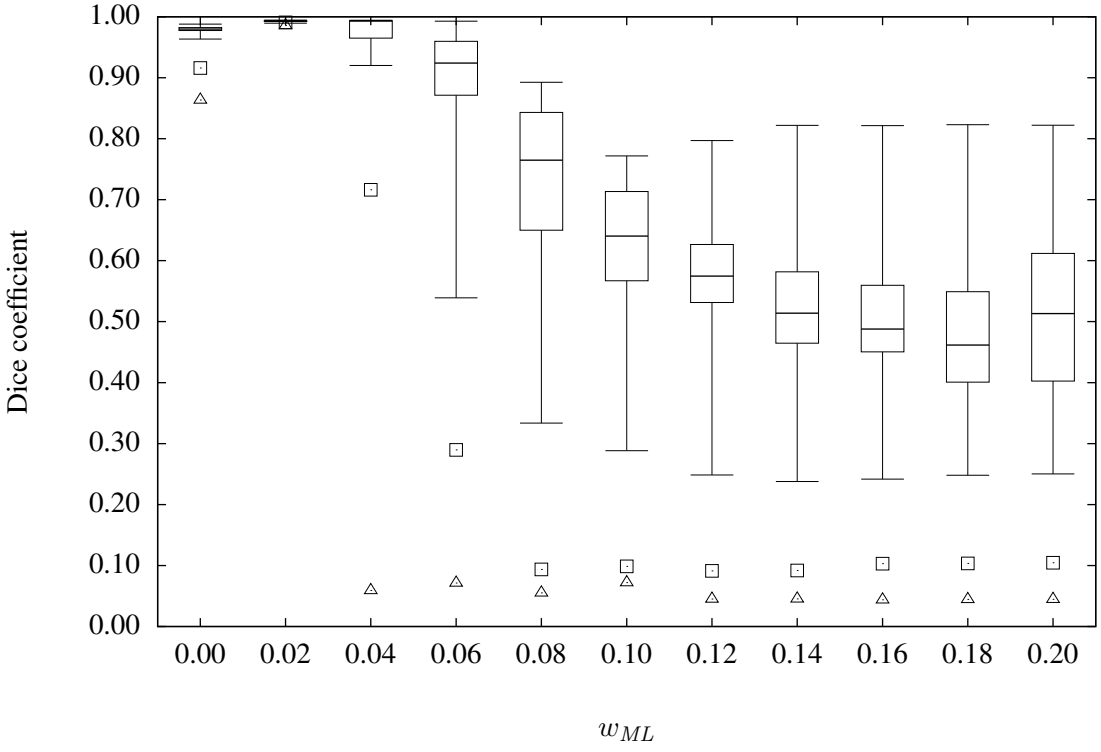


(a) (b)

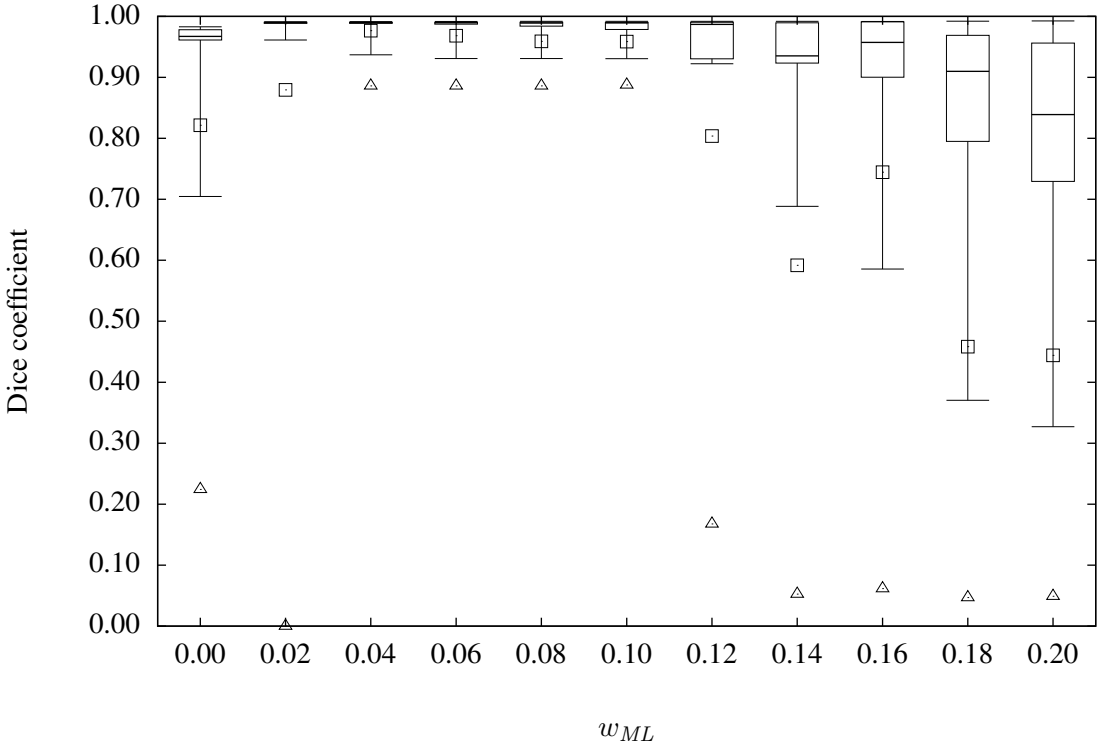
Figure 5.6: A comparison of results of segmentation using the pure ESA-based algorithm described in Chapter 4 (Subfigure 5.6a) and machine learning-based segmentation with $w_{ML} = 0.02$ (Subfigure 5.6b) for a sample image from the knee dataset. Notation is described in the caption for Figure 4.9.

Figure 5.6) the machine learning-based weight calculation method further reduces the issue of suboptimal scale and rotation adjustment. Full solution of this problem is left for future work on the algorithm.

One of the drawbacks of using complex machine learning methods is poor comprehensibility of constructed shape models. This was not considered as a requirement for this research, however in the future it is possible to limit the choice of machine learning algorithms that only use simple to understand models of data. For example, the AAM and ASM algorithms use PCA for modelling of shape variability.



(a)



(b)

Figure 5.7: Box plots for different values of w_{ML} in cross-validation for the knee dataset (Subfigure 5.7a) and hip bone dataset (Subfigure 5.7b).

Chapter 6

Conclusions

Concluding remarks

The main objective of this work was to show that it is possible to effectively apply global optimization methods to the problem of boundary delineation by constraining the search to borders of a superpixel segmentation of the image. This goal was achieved by proving three research hypotheses in subsequent chapters.

First it was shown that it is possible to achieve high segmentation accuracy when the search space is constrained by the input image. This was proved by explicitly constructing an algorithm for this task and testing it on four different sets of images. One of the major flaws of this approach was a considerable number of free parameters that need to be separately determined for each particular set of images.

In the next chapter an algorithm based on Elastic Shape Analysis is described that provides a principled, global and, for many practical situations, exact way to find the best match in a superpixel graph for a given pattern. The algorithms has few free parameters that have relatively weak influence on the quality of the resulting segmentation. The main drawback of this method is worse accuracy compared to state-of-the-art algorithms.

Finally, the ESA-based algorithm is combined with a statistical shape model constructed using machine learning methods. This is a natural extension the framework that introduces only one free parameter with significant impact on the segmentation accuracy. The influence of this parameter is analysed on two sets of images proving that the algorithm provides average accuracy comparable to state-of-the-art methods while avoiding the need to introduce a Point Distribution Model in an arbitrary way and requiring only very limited initialization data. The algorithm is at the same time quite efficient.

Obtained results suggest that the proposed approach to boundary delineation is very promising. It has good theoretical properties and the results of experiments are encouraging. The next section describes a few directions for future improvements of the presented method.

Future work

One of the sources of plans for future research is weakening the assumptions given in Section 1.1. This includes the following:

- Extending described algorithms to colour and hyperspectral images (Lunga et al., 2014).
- Combining the algorithm with feature detectors and descriptors capable of automatic extraction of boundary endpoints. This extension would make the algorithm perform a fully global search.
- Segmentation of objects that are not simply connected (Mikolajczyk et al., 2003). This would require simultaneous detection of several distinct curves corresponding to different parts of the object boundary.

Furthermore, it is not clear how the algorithm should be modified to handle boundaries of objects that do not align with boundaries between superpixels. This may issue may arise for example when an object is partially occluded. Figure 4.10e depicts such a case. Feature detectors, considered in this thesis, would not improve the outcome as the features themselves are not visible in the picture.

Finally, one of the drawbacks of the methods presented in Chapters 4 and 5 is the lack of rotational and scale invariance. It is worth checking to what extent the algorithms can be extended to feature these two properties while remaining computationally efficient. Affine and projective invariance may also be considered (Bryner et al., 2014). Certain additional directions for further work are described in conclusions to Chapters 3, 4 and 5.

Appendix A

Properties of the routing algebra $\mathcal{A}$

The routing algebra $\mathcal{A}$ is defined by a triple $(W, \oplus, \preceq)$ where $W = \{(x, y) \colon x \in \mathbb{R}_{\geq 0}, y \in \mathbb{R}_{+}\} \cup \{(0,0), (\infty, \infty)\}$, $\oplus$ is defined in Table 4.1 and $\preceq$ is the lexicographic order, that is for any two elements of the algebra (a, b) and (c, d) the inequality $(a, b) \preceq (c, d)$ is true if and only if $a < c$ or $a = c$ and $b \leq d$, assuming that ∞ is greater than any real number and $\infty = \infty$. Let us verify that $\mathcal{A}$ possesses the properties of a (commutative) routing algebra (Definition 8).

- *Closure.* Let $w_1 = (a, b), w_2 = (c, d)$ be any elements of W. The value of $w_1 \oplus w_2$ is either $(0, 0)$, w_1, w_2, (∞, ∞) or $\left(\frac{ab+cd}{b+d}, b + d\right)$. Using basic arithmetic one can show that in all of these cases the pair belongs to W. □
- *Commutativity.* Let $w_1 = (a, b), w_2 = (c, d)$ be any elements of W. Commutativity ($w_1 \oplus w_2 = w_2 \oplus w_1$) is obvious from the symmetry of Table 4.1 and commutativity of the addition and multiplication of real numbers. □
- *Associativity.* Let $w_1 = (a, b)$,$w_2 = (c, d)$ and $w_3 = (e, f)$ be any elements of W. Associativity is proved case-by-case in Table A.1. Note that rows and columns are interchangeable since $\mathcal{A}$ is commutative. □
- *Identity element.* The identity element of $\mathcal{A}$ is $(0, 0)$. It can be easily verified using Table 4.1. □
- *Zero element.* The zero element of $\mathcal{A}$ is (∞, ∞). It can be easily verified using Table 4.1. □
- *Reflexivity.* There are three cases:
  - $(0, 0) \preceq (0, 0)$ is true since $0 = 0$ and $0 \leq 0$.
  - $(a, b) \preceq (a, b)$ where a is a nonnegative real number and b is a positive real number is true because $a = a$ and $b \leq b$.
  - $(\infty, \infty) \preceq (\infty, \infty)$ is true because, as assumed, $\infty = \infty$ and $\infty \leq \infty$.

  □
- *Antisymmetry.* Let $w_1 = (a, b)$ and $w_2 = (c, d)$ be two elements of W. From the assumption it is known that $w_1 \preceq w_2$ and $w_2 \preceq w_1$. This is equivalent to $(a < c \vee (a = c \wedge b \leq d)) \wedge (c < a \vee (a =$

Table A.1: Validation of associativity of the routing algebra $\mathcal{A}$

(a) Values of $w_1 \oplus (w_2 \oplus w_3)$. Columns contain values of w_1 and rows contain values of $w_2 \oplus w_3$

$\oplus$	$(0,0)$	(a,b)	(∞,∞)
$(0,0) \oplus (0,0)$	$(0,0)$	(a,b)	(∞,∞)
$(c,d) \oplus (0,0)$	(c,d)	$\left(\frac{ab+cd}{b+d}, b+d\right)$	(∞,∞)
$(\infty,\infty) \oplus (0,0)$	(∞,∞)	(∞,∞)	(∞,∞)
$(0,0) \oplus (e,f)$	(e,f)	$\left(\frac{ab+ef}{b+f}, b+f\right)$	(∞,∞)
$(c,d) \oplus (e,f)$	$\left(\frac{cd+ef}{d+f}, d+f\right)$	$\left(\frac{ab+cd+ef}{b+d+f}, b+d+f\right)$	(∞,∞)
$(\infty,\infty) \oplus (e,f)$	(∞,∞)	(∞,∞)	(∞,∞)
$(0,0) \oplus (\infty,\infty)$	(∞,∞)	(∞,∞)	(∞,∞)
$(c,d) \oplus (\infty,\infty)$	(∞,∞)	(∞,∞)	(∞,∞)
$(\infty,\infty) \oplus (\infty,\infty)$	(∞,∞)	(∞,∞)	(∞,∞)

(b) Values of $(w_1 \oplus w_2) \oplus w_3$. Columns contain values of w_3 and rows contain values of $w_1 \oplus w_2$

$\oplus$	$(0,0)$	(e,f)	(∞,∞)
$(0,0) \oplus (0,0)$	$(0,0)$	(e,f)	(∞,∞)
$(a,b) \oplus (0,0)$	(a,b)	$\left(\frac{ab+ef}{b+f}, b+f\right)$	(∞,∞)
$(\infty,\infty) \oplus (0,0)$	(∞,∞)	(∞,∞)	(∞,∞)
$(0,0) \oplus (c,d)$	(c,d)	$\left(\frac{cd+ef}{d+f}, d+f\right)$	(∞,∞)
$(a,b) \oplus (c,d)$	$\left(\frac{ab+cd}{b+d}, b+d\right)$	$\left(\frac{ab+cd+ef}{b+d+f}, b+d+f\right)$	(∞,∞)
$(\infty,\infty) \oplus (c,d)$	(∞,∞)	(∞,∞)	(∞,∞)
$(0,0) \oplus (\infty,\infty)$	(∞,∞)	(∞,∞)	(∞,∞)
$(a,b) \oplus (\infty,\infty)$	(∞,∞)	(∞,∞)	(∞,∞)
$(\infty,\infty) \oplus (\infty,\infty)$	(∞,∞)	(∞,∞)	(∞,∞)

$c \wedge d \leq b)$). If $a < c$ then the second part of conjunction is false and if $c < a$ then the first part of conjunction is false, thus $a = c$. Now we can infer that $b \leq d$ and $d \leq b$. If both b and d are finite real numbers then $b = d$ from antisymmetry of standard ordering relation $\leq$ for real numbers. If one of these numbers is equal to ∞ then the other one is also equal to ∞ because from our assumptions ∞ is the only solution to $\infty \leq x$ where $x \in \mathbb{R}_{\geq 0} \cup \{\infty\}$. □

- *Transitivity.* Let $w_1 = (a, b), w_2 = (c, d), w_3 = (e, f) \in W$ such that $w_1 \preceq w_2$ and $w_2 \preceq w_3$. The possible cases are:
  - $a = c = e$. In this case $b \leq d$ and $d \leq f$. If b, d and f are finite, then from the transitivity of standard ordering $\leq$ of real numbers $b \leq d$ and $w_1 \preceq w_3$ follows. If any of b, d, f is equal to ∞ then so must be f. As a result $b \leq f$ and therefore $w_1 \preceq w_3$.
  - $a = c < e$, $a < c = e$ or $a < c < e$. In these cases $a < e$ and thus $w_1 \preceq w_3$.

  □
- *Totality.* Let $w_1 = (a, b)$, $w_2 = (c, d)$ be elements of W. If $a = \infty$ ($c = \infty$), then $w_2 \preceq w_1$ ($w_1 \preceq w_2$). Otherwise, there are three cases:
  - $a < c$. In this case $w_1 \preceq w_2$.
  - $a = c$. Using the totality of the standard ordering of real numbers extended by positive infinity either $b \leq d$ or $d \leq b$, thus either $w_1 \preceq w_2$ or $w_2 \preceq w_1$.
  - $a > c$. In this case $w_2 \preceq w_1$.

  □
- *Maximal element.* The maximal element is (∞, ∞). □
- *Equality of maximal element and zero element.* The maximal element (∞, ∞) is equal to the zero element (∞, ∞) directly from the assumptions. □

Monotonicity (Definition 10) and isotonicity (Definition 11) do not hold for $\mathcal{A}$. A simple counterexample for monotonicity is $w_1 = (2, 1)$ and $w_2 = (1, 1)$. The condition for isotonicity is false when $w_1 = (1, 1)$, $w_2 = (1, 2)$ and $w_3 = (2, 1)$.

Appendix B

Proof of Theorem 1

Let $G_1 = (V_1, E_1), G_2 = (V_2, E_2)$ be digraphs with drawings on $\mathbb{R}^2$ defined by $\phi_1, \xi_1, \phi_2, \xi_2$. The following lemma is used in the proof of Theorem 1.

Lemma 1. *If $p = (p_{1,i}, p_{2,i})_{i=1}^{N}$ is a sequence of σ-pairs of paths in G_1, G_2 between $(v_{b,1}, v_{b,2})$ and $(v_{e,1}, v_{e,2})$, then there is a sequence $\hat{p} = (\hat{p}_{1,i}, \hat{p}_{2,i})_{i=1}^{N}$ of σ-pairs of paths in G_1, G_2 between the same pairs of nodes such that $d_{pre}^{c}(\rho_1^q(\hat{p}), \rho_2^q(\hat{p}) \cdot \gamma(\hat{p})) \leq d_{pre}^{c}(\rho_1^q(p), \rho_2^q(p) \cdot \gamma(p))$ and for each $i = 1, 2, \ldots, N$ the pair $(\hat{p}_{1,i}, \hat{p}_{2,i})$ minimizes the value of Equation* (4.2) *for a certain edge in $G_1 \times_\sigma G_2$.*

Proof. From the third condition in Definition 31 for each $i = 1, 2, \ldots, N$ either $p_{1,i}, p_{2,i}$ is a σ-pair of paths that corresponds to an edge in $G_1 \times_\sigma G_2$ or there is a σ-pair of paths $\bar{p}_{1,i}, \bar{p}_{1,i}$ between the same pairs of nodes that gives a lower value of I. Sequence of σ-pairs $\hat{p} = (\hat{p}_{1,i}, \hat{p}_{2,i})_{i=1}^{N}$ where $\hat{p}_{j,i}$ is $p_{j,i}$ in the first case and $\bar{p}_{j,i}$ in the second case, for $j = 1, 2$, is the required sequence. □

A proof of Theorem 1 is given below.

Proof. From Lemma 1 it is known that the closest sequence of σ-pairs of paths must correspond to a path in $G_1 \times_\sigma G_2$. On the other hand, since the optimal path finding procedure is exact, it returns the minimum weighted average path between a given pair of nodes. Thus there can be no sequence of σ-pairs of paths with smaller elastic distance. □

Bibliography

I.E. Abdou and W.K. Pratt. Quantitative design and evaluation of enhancement/thresholding edge detectors. *Proceedings of the IEEE*, 67(5):753–763, 1979. ISSN 0018-9219. doi: 10.1109/PROC.1979.11325. URL http://ieeexplore.ieee.org/document/1455594/.

Ralph Abraham, J. E. Marsden, and Tudor Ratiu. *Manifolds, Tensor Analysis, and Applications*. Springer-Verlag Publishing Company, Inc., New York, 3rd edition, 2001. ISBN 0-201-10168-S.

Radhakrishna Achanta, Appu Shaji, Kevin Smith, Aurélien Lucchi, Pascal Fua, and Sabine Süsstrunk. SLIC Superpixels. Technical report, EPFL, EPFL, 2010.

Radhakrishna Achanta, Appu Shaji, Kevin Smith, Aurelien Lucchi, Pascal Fua, and Sabine Susstrunk. SLIC Superpixels Compared to State-of-the-Art Superpixel Methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(11):2274–2282, November 2012. ISSN 0162-8828. doi: 10.1109/TPAMI.2012.120.

Alberto S. Aguado, Eugenia Montiel, and Mark S. Nixon. Invariant characterisation of the Hough transform for pose estimation of arbitrary shapes. *Pattern Recognition*, 35(5):1083–1097, May 2002. ISSN 0031-3203. doi: 10.1016/S0031-3203(01)00099-1. URL http://www.sciencedirect.com/science/article/pii/S0031320301000991.

R. V. Ambartzumian. *Factorization Calculus and Geometric Probability*. Cambridge University Press, Cambridge ; New York, 1st edition, September 1990. ISBN 978-0-521-34535-4.

Jesús Angulo and Dominique Jeulin. Stochastic watershed segmentation. In *Proc. of the 8th Int. Symposium on Mathematical Morphology*, pages 265–276, Rio de Janeiro, Brazil, 2007.

Carlos Ansótegui, Meinolf Sellmann, and Kevin Tierney. A Gender-Based Genetic Algorithm for the Automatic Configuration of Algorithms. In *Principles and Practice of Constraint Programming - CP 2009*, pages 142–157. Springer, Berlin, Heidelberg, September 2009. doi: 10.1007/978-3-642-04244-7_14. URL https://link.springer.com/chapter/10.1007/978-3-642-04244-7_14.

K. S. Arun, T. S. Huang, and S. D. Blostein. Least-Squares Fitting of Two 3-D Point Sets. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-9(5):698–700, September 1987. ISSN 0162-8828. doi: 10.1109/TPAMI.1987.4767965.

Suyash P. Awate, Yen-Yun Yu, and Ross T. Whitaker. Kernel Principal Geodesic Analysis. In Toon Calders, Floriana Esposito, Eyke Hüllermeier, and Rosa Meo, editors, *Machine Learning and Knowledge Discovery in Databases*, number 8724 in Lecture Notes in Computer Science, pages 82–98. Springer Berlin Heidelberg, September 2014. ISBN 978-3-662-44847-2 978-3-662-44848-9. doi: 10.1007/978-3-662-44848-9_6. URL http://link.springer.com/chapter/10.1007/978-3-662-44848-9_6.

E. Bae, J. Shi, and X. C. Tai. Graph Cuts for Curvature Based Image Denoising. *IEEE Transactions on Image Processing*, 20(5):1199–1210, May 2011. ISSN 1057-7149. doi: 10.1109/TIP.2010.2090533.

D. H. Ballard. Generalizing the Hough transform to detect arbitrary shapes. *Pattern Recognition*, 13(2): 111–122, January 1981. ISSN 0031-3203. doi: 10.1016/0031-3203(81)90009-1. URL `http://www.sciencedirect.com/science/article/pii/0031320381900091`.

Dhruv Batra, Payman Yadollahpour, Abner Guzman-Rivera, and Gregory Shakhnarovich. Diverse M-Best Solutions in Markov Random Fields. In Andrew Fitzgibbon, Svetlana Lazebnik, Pietro Perona, Yoichi Sato, and Cordelia Schmid, editors, *Computer Vision – ECCV 2012: 12th European Conference on Computer Vision, Florence, Italy, October 7-13, 2012, Proceedings, Part V*, pages 1–16, Berlin, Heidelberg, 2012. Springer Berlin Heidelberg. ISBN 978-3-642-33715-4. doi: 10.1007/978-3-642-33715-4_1. URL `https://doi.org/10.1007/978-3-642-33715-4_1`.

Y. Bengio, A. Courville, and P. Vincent. Representation Learning: A Review and New Perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(8):1798–1828, August 2013. ISSN 0162-8828. doi: 10.1109/TPAMI.2013.50.

C. Berge. *Hypergraphs: Combinatorics of Finite Sets*. Elsevier, May 1984. ISBN 978-0-08-088023-5.

Javier Bernal, Günay Doğan, and Charles R. Hagwood. Fast Dynamic Programming for Elastic Registration of Curves. In *2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1066–1073, June 2016. doi: 10.1109/CVPRW.2016.137.

P. J. Besl and N. D. McKay. A method for registration of 3-D shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(2):239–256, February 1992. ISSN 0162-8828. doi: 10.1109/34.121791.

S. Beucher. Watersheds of functions and picture segmentation. In *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP '82.*, volume 7, pages 1928–1931, May 1982. doi: 10.1109/ICASSP.1982.1171424.

S. Beucher and C. Lantuéjoul. Use of watershed in contour detection. In *International Workshop on Image Processing: Real-time Edge and Motion Detection/Estimation*, September 1979.

Serge Beucher and Fernand Meyer. The morphological approach to segmentation: The watershed transformation. *Mathematical Morphology in Image Processing*, 34:433–481, 1993.

Mauro Birattari, Zhi Yuan, Prasanna Balaprakash, and Thomas Stützle. F-Race and Iterated F-Race: An Overview. In Thomas Bartz-Beielstein, Marco Chiarandini, Luís Paquete, and Mike Preuss, editors, *Experimental Methods for the Analysis of Optimization Algorithms*. Springer-Verlag Berlin Heidelberg, 2010. ISBN 978-3-642-02537-2. URL `http://www.springer.com/us/book/9783642025372`.

Christopher Bishop. *Pattern Recognition and Machine Learning*. Information Science and Statistics. Springer-Verlag New York, 1st edition, 2006. ISBN 978-0-387-31073-2.

J.A. Bondy and U.S.R. Morty. *Graph Theory with Applications*. North-Holland Publishing Co., 1982.

H. Boussaid, I. Kokkinos, and N. Paragios. Discriminative learning of deformable contour models. In *Biomedical Imaging (ISBI), 2014 IEEE 11th International Symposium on*, pages 624–628, Beijing, China, April 2014. IEEE. ISBN 978-1-4673-1961-4. doi: 10.1109/ISBI.2014.6867948.

Y. Y. Boykov and M. P. Jolly. Interactive graph cuts for optimal boundary & region segmentation of objects in N-D images. In *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*, volume 1, pages 105–112, Vancouver, BC, Canada, 2001. IEEE. doi: 10.1109/ICCV.2001.937505.

Andrew P. Bradley. The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognition*, 30(7):1145–1159, July 1997. ISSN 0031-3203. doi: 10.1016/S0031-3203(96)00142-2. URL http://www.sciencedirect.com/science/article/pii/S0031320396001422.

Leo Breiman. Bagging Predictors. *Machine Learning*, 24(2):123–140, August 1996. doi: 10.1023/A:1018054314350.

Leo Breiman. Random Forests. *Machine Learning*, 45(1):5–32, October 2001. doi: 10.1023/A:1010933404324.

M. Brejl and M. Sonka. Object localization and border detection criteria design in edge-based image segmentation: automated learning from examples. *IEEE Transactions on Medical Imaging*, 19(10):973–985, October 2000. ISSN 0278-0062. doi: 10.1109/42.887613.

Alain Bretto. Introduction to Hypergraph Theory and Its Use in Engineering and Image Processing. *Advances in Imaging and Electron Physics*, 131:1–64, January 2004. ISSN 1076-5670. doi: 10.1016/S1076-5670(04)31001-3. URL http://www.sciencedirect.com/science/article/pii/S1076567004310013.

Alain Bretto and Luc Gillibert. Hypergraph-Based Image Representation. In *Proceedings of the 5th IAPR International Conference on Graph-Based Representations in Pattern Recognition*, GbRPR'05, pages 1–11, Berlin, Heidelberg, 2005. Springer-Verlag. ISBN 978-3-540-25270-2. doi: 10.1007/978-3-540-31988-7_1.

Christopher Brignell. *Shape analysis and statistical modelling in brain imaging*. PhD Thesis, University of Nottingham, Nottingham, July 2007. URL http://eprints.nottingham.ac.uk/12106/.

P. A. Bromiley, C. Lindner, J. Thomson, M. Wrigley, and T. F. Cootes. Multi-point Regression Voting for Shape Model Matching. *Procedia Computer Science*, 90(Supplement C):48–53, January 2016. ISSN 1877-0509. doi: 10.1016/j.procs.2016.07.009. URL http://www.sciencedirect.com/science/article/pii/S1877050916311875.

Andrés Bruhn, Thomas Pock, and Xue-Cheng Tai, editors. *Efficient Algorithms for Global Optimization Methods in Computer Vision*. Number 8293 in Image Processing, Computer Vision, Pattern Recognition, and Graphics. Springer-Verlag Berlin Heidelberg, 2014. URL //www.springer.com/gp/book/9783642547737.

D. Bryner, E. Klassen, H. Le, and A. Srivastava. 2d Affine and Projective Shape Analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(5):998–1011, May 2014. ISSN 0162-8828. doi: 10.1109/TPAMI.2013.199.

Edmund K. Burke, Michel Gendreau, Matthew Hyde, Graham Kendall, Gabriela Ochoa, Ender Özcan, and Rong Qu. Hyper-heuristics: a survey of the state of the art. *Journal of the Operational Research Society*, 64(12):1695–1724, December 2013. ISSN 0160-5682, 1476-9360. doi: 10.1057/jors.2013.71. URL https://link.springer.com/article/10.1057/jors.2013.71.

Qifeng Chen and Vladlen Koltun. Full Flow: Optical Flow Estimation By Global Optimization Over Regular Grids. In *Computer Vision and Pattern Recognition (CVPR), 2016 IEEE Conference on*, pages 4706–4714, Las Vegas, NV, USA, 2016. IEEE. doi: 10.1109/CVPR.2016.509.

Siqi Chen, D. Michael Lovelock, and Richard J. Radke. Segmenting the prostate and rectum in CT imagery using anatomical constraints. *Medical Image Analysis*, 15(1):1–11, February 2011. ISSN 1361-8415. doi: 10.1016/j.media.2010.06.004. URL http://www.sciencedirect.com/science/article/pii/S1361841510000678.

Xinjian Chen, Jayaram K. Udupa, Abass Alavi, and Drew A. Torigian. GC-ASM: Synergistic integration of graph-cut and active shape model strategies for medical image segmentation. *Computer Vision and Image Understanding*, 117(5):513–524, May 2013. ISSN 1077-3142. doi: 10.1016/j.cviu.2012.12.001. URL http://www.sciencedirect.com/science/article/pii/S1077314212001956.

Yang Chen and Gérard Medioni. Object modelling by registration of multiple range images. *Image and Vision Computing*, 10(3):145–155, April 1992. ISSN 0262-8856. doi: 10.1016/0262-8856(92)90066-C. URL http://www.sciencedirect.com/science/article/pii/026288569290066C.

Weiwei Cheng and Eyke Hüllermeier. Combining instance-based learning and logistic regression for multilabel classification. *Machine Learning*, 76(2-3):211–225, September 2009. doi: 10.1007/s10994-009-5127-5.

Erhan Çınlar. *Probability and Stochastics*. Number 261 in Graduate Texts in Mathematics. Springer, New York ; London, 1st edition, February 2011. ISBN 978-0-387-87858-4.

Krzysztof J. Cios, Witold Pedrycz, Roman W. Swiniarski, and Lukasz A. Kurgan. *Data Mining: A Knowledge Discovery Approach*. Springer, New York, NY, February 2007. ISBN 978-0-387-33333-5.

Isaac Cohen, Nicholas Ayache, and Patrick Sulger. Tracking points on deformable objects using curvature information. In G. Sandini, editor, *Computer Vision — ECCV'92*, Lecture Notes in Computer Science. Springer Berlin Heidelberg, May 1992. ISBN 978-3-540-55426-4 978-3-540-47069-4. doi: 10.1007/3-540-55426-2_51.

Jacob Cohen. A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, 20(1):37–46, April 1960. ISSN 0013-1644. doi: 10.1177/001316446002000104.

D. Comaniciu and P. Meer. Mean shift: a robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(5):603–619, May 2002. ISSN 0162-8828. doi: 10.1109/34.1000236.

T. F. Cootes and C. J. Taylor. Statistical Models of Appearance for Computer Vision. Technical Report, Wolfson Image Analysis Unit, University of Manchester, Manchester, U.K., 1999.

T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham. Active Shape Models-Their Training and Application. *Computer Vision and Image Understanding*, 61(1):38–59, January 1995. ISSN 1077-3142. doi: 10.1006/cviu.1995.1004.

Timothy F. Cootes, Gareth J. Edwards, and Christopher J. Taylor. Active Appearance Models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(6):681–685, June 2001. ISSN 0162-8828. doi: 10.1109/34.927467.

Emil Cornea, Hongtu Zhu, Peter Kim, and Joseph G. Ibrahim. Regression models on Riemannian symmetric spaces. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 2016. doi: 10.1111/rssb.12169. URL http://onlinelibrary.wiley.com/doi/10.1111/rssb.12169/full.

Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine Learning*, 20(3):273–297, September 1995. doi: 10.1007/BF00994018.

F. Arámbula Cosío. Automatic initialization of an active shape model of the prostate. *Medical Image Analysis*, 12(4):469–483, August 2008. ISSN 1361-8415. doi: 10.1016/j.media.2008.02.001. URL http://www.sciencedirect.com/science/article/pii/S1361841508000212.

C. Couprie, L. Grady, L. Najman, and H. Talbot. Power Watershed: A Unifying Graph-Based Optimization Framework. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(7):1384–1399, July 2011. ISSN 0162-8828. doi: 10.1109/TPAMI.2010.200.

David Cristinacce and Tim Cootes. Feature detection and tracking with constrained local models. In *BMVC 2006 - Proceedings of the British Machine Vision Conference 2006*, pages 929–938, Edinburgh, 2006. British Machine Vision Association, BMVA.

David Cristinacce and Tim Cootes. Automatic feature localisation with constrained local models. *Pattern Recognition*, 41(10):3054–3067, October 2008. ISSN 0031-3203. doi: 10.1016/j.patcog.2008.01.024. URL http://www.sciencedirect.com/science/article/pii/S0031320308000630.

C. Davatzikos, Xiaodong Tao, and Dinggang Shen. Hierarchical active shape models, using the wavelet transform. *IEEE Transactions on Medical Imaging*, 22(3):414–423, March 2003. ISSN 0278-0062. doi: 10.1109/TMI.2003.809688.

G. G. Demisse, D. Aouada, and B. Ottersten. Similarity Metric for Curved Shapes in Euclidean Space. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5042–5050, June 2016. doi: 10.1109/CVPR.2016.545.

G. G. Demisse, D. Aouada, and B. Ottersten. Deformation Based Curved Shape Representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PP(99):1–1, 2017. ISSN 0162-8828. doi: 10.1109/TPAMI.2017.2711607.

Janez Demšar. On the Appropriateness of Statistical Tests in Machine Learning. In *Proceedings of the ICML'08 Third Workshop on Evaluation Methods for Machine Learning*, New York, 2008. Association for Computing Machinery.

C. Lakshmi Devasena, T. Sumathi, V. V. Gomathi, and M. Hemalatha. Effectiveness Evaluation of Rule Based Classifiers for the Classification of Iris Data Set. *Bonfring International Journal of Man Machine Interface*, 1(Inaugural Special Issue):5–9, December 2011. ISSN 2277-5064. doi: 10.9756/BIJMMI.1002. URL http://journal.bonfring.org/abstract.php?id=4&archiveid=104.

Lee R. Dice. Measures of the Amount of Ecologic Association Between Species. *Ecology*, 26(3):297–302, 1945. ISSN 0012-9658. doi: 10.2307/1932409. URL http://www.jstor.org/stable/1932409.

H. Digabel and C. Lantuéjoul. Iterative Algorithms. In *Proceedings of the 2nd European Symposium Quantitative Analysis of Microstructures in Material Science, Biology and Medicine*, pages 85–89, 1978.

E. W. Dijkstra. A note on two problems in connexion with graphs. *Numerische Mathematik*, 1(1):269–271, December 1959. ISSN 0029-599X, 0945-3245. doi: 10.1007/BF01386390.

Günay Doğan, Javier Bernal, and Charles R. Hagwood. A Fast Algorithm for Elastic Shape Distances Between Closed Planar Curves. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4222–4230, Boston, MA, USA, 2015. IEEE. doi: 10.1109/CVPR.2015.7299050.

Stephan Dreiseitl and Lucila Ohno-Machado. Logistic regression and artificial neural network classification models: a methodology review. *Journal of Biomedical Informatics*, 35(5):352–359, October 2002. ISSN 1532-0464. doi: 10.1016/S1532-0464(03)00034-0.

Chris Drummond. Machine learning as an experimental science (revised). In *Proceedings of the AAAI'06 Workshop on Evaluation Methods for Machine Learning I*, Menlo Park, CA, 2006. The AAAI Press.

Ian L. Dryden and Kanti V. Mardia. *Statistical Shape Analysis*. Wiley, Chichester ; New York, 1st edition, September 1998. ISBN 978-0-471-95816-1.

Paul Dupuis, Ulf Grenander, and Michael I. Miller. Variational Problems on Flows of Diffeomorphisms for Image Matching. *Quarterly of Applied Mathematics*, 56(3):587–600, September 1998. ISSN 0033-569X.

N. Duta and M. Sonka. Segmentation and interpretation of MR brain images. An improved active shape model. *IEEE Transactions on Medical Imaging*, 17(6):1049–1062, December 1998. ISSN 0278-0062. doi: 10.1109/42.746716.

Bradley Efron and Gail Gong. A Leisurely Look at the Bootstrap, the Jackknife, and Cross-Validation. *The American Statistician*, 37(1):36–48, 1983. ISSN 0003-1305. doi: 10.1080/00031305.1983.10483087. URL `http://www.tandfonline.com/doi/abs/10.1080/00031305.1983.10483087`.

N. Y. El-Zehiry and L. Grady. Fast global optimization of curvature. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 3257–3264, San Francisco, CA, USA, June 2010. IEEE. doi: 10.1109/CVPR.2010.5540057.

J. H. Elder, A. Krupnik, and L. A. Johnston. Contour grouping with prior models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(6):661–674, June 2003. ISSN 0162-8828. doi: 10.1109/TPAMI.2003.1201818.

István Fáry. On straight line representation of planar graphs. *Acta Scientiarum Mathematicarum (Szeged)*, 11(4-4):229–233, 1948.

Ariel Felner, Alex Pomeransky, and Jeffrey S. Rosenschein. Searching for an Alternative Plan. In *Proceedings of the Second International Joint Conference on Autonomous Agents and Multiagent Systems*, AAMAS '03, pages 33–40, New York, NY, USA, 2003. ACM. ISBN 978-1-58113-683-8. doi: 10.1145/860575.860582.

Pedro F. Felzenszwalb and Daniel P. Huttenlocher. Efficient Graph-Based Image Segmentation. *International Journal of Computer Vision*, 59(2):167–181, September 2004. ISSN 0920-5691, 1573-1405. doi: 10.1023/B:VISI.0000022288.19776.77. URL `https://link.springer.com/article/10.1023/B:VISI.0000022288.19776.77`.

Peter Flach. *Machine Learning: The Art and Science of Algorithms that Make Sense of Data*. Cambridge University Press, Cambridge ; New York, November 2012. ISBN 978-1-107-42222-3.

J. Fleiss. Measuring nominal scale agreement among many raters. *Psychological Bulletin*, 76(5):378–382, 1971. doi: 10.1037/h0031619.

P. Thomas Fletcher. Geodesic Regression and the Theory of Least Squares on Riemannian Manifolds. *International Journal of Computer Vision*, 105(2):171–185, November 2013. ISSN 0920-5691, 1573-1405. doi: 10.1007/s11263-012-0591-y. URL `http://link.springer.com/article/10.1007/s11263-012-0591-y`.

P. Thomas Fletcher, Conglin Lu, and Sarang C. Joshi. Statistics of Shape via Principal Geodesic Analysis on Lie Groups. In *2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings.*, volume 1, pages 95–101, Madison, WI, USA, June 2003. IEEE. doi: 10.1109/CVPR.2003.1211342.

P. Thomas Fletcher, Suresh Venkatasubramanian, and Sarang Joshi. The geometric median on Riemannian manifolds with application to robust atlas estimation. *NeuroImage*, 45(1 Suppl):S143–152, March 2009. ISSN 1095-9572. doi: 10.1016/j.neuroimage.2008.10.052.

P.T. Fletcher, C. Lu, S.M. Pizer, and S. Joshi. Principal geodesic analysis for the study of nonlinear statistics of shape. *IEEE Transactions on Medical Imaging*, 23(8):995–1005, August 2004. ISSN 0278-0062. doi: 10.1109/TMI.2004.831793.

L.R. Foulds. *Graph Theory Applications*. Universitext. Springer New York, 1992. ISBN 978-1-4612-0933-1. URL `http://www.springer.com/gp/book/9780387975993`.

J. R. Fram and E. S. Deutsch. On the Quantitative Evaluation of Edge Detection Schemes and their Comparison with Human Performance. *IEEE Transactions on Computers*, C-24(6):616–628, June 1975. ISSN 0018-9340. doi: 10.1109/T-C.1975.224274.

Maurice Fréchet. Les éléments aléatoires de nature quelconque dans un espace distancié. *Annales de l'institut Henri Poincaré*, 10(4):215–310, 1948. ISSN 0365-320X. URL `https://eudml.org/doc/79021`.

Michael L. Fredman and Robert Endre Tarjan. Fibonacci Heaps and Their Uses in Improved Network Optimization Algorithms. *Journal of the ACM*, 34(3):596–615, July 1987. ISSN 0004-5411. doi: 10.1145/28869.28874.

H. Freeman. On the Encoding of Arbitrary Geometric Configurations. *IRE Transactions on Electronic Computers*, EC-10(2):260–268, June 1961. ISSN 0367-9950. doi: 10.1109/TEC.1961.5219197.

J. Freixenet, X. Muñsoz, D. Raba, J. Martí, and X. Cufí. Yet Another Survey on Image Segmentation: Region and Boundary Information Integration. In *Computer Vision — ECCV 2002*, volume 2352 of *Lecture Notes in Computer Science*, pages 408–422. Springer, Berlin, Heidelberg, May 2002. ISBN 978-3-540-43746-8 978-3-540-47977-2. doi: 10.1007/3-540-47977-5_27.

Max Frenkel and Ronen Basri. Curve Matching Using the Fast Marching Method. In Anand Rangarajan, Mário Figueiredo, and Josiane Zerubia, editors, *Energy Minimization Methods in Computer Vision and Pattern Recognition*, Lecture Notes in Computer Science, pages 35–51. Springer Berlin Heidelberg, July 2003. ISBN 978-3-540-40498-9 978-3-540-45063-4. doi: 10.1007/978-3-540-45063-4_3. URL `http://link.springer.com/chapter/10.1007/978-3-540-45063-4_3`.

Jerome H. Friedman. Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, 29(5):1189–1232, 2001. ISSN 0090-5364, 2168-8966. doi: 10.1214/aos/1013203451. URL https://projecteuclid.org/euclid.aos/1013203451.

K. S. Fu and J. K. Mui. A survey on image segmentation. *Pattern Recognition*, 13(1):3–16, January 1981. ISSN 0031-3203. doi: 10.1016/0031-3203(81)90028-5.

B. Fulkerson, A. Vedaldi, and S. Soatto. Class segmentation and object localization with superpixel neighborhoods. In *2009 IEEE 12th International Conference on Computer Vision*, pages 670–677, September 2009. doi: 10.1109/ICCV.2009.5459175.

J. Gall and V. Lempitsky. Class-specific Hough forests for object detection. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1022–1029, June 2009. doi: 10.1109/CVPR.2009.5206740.

Anna Galluccio, Luisa Gargano, János Körner, and Gábor Simonyi. Different capacities of a digraph. *Graphs and Combinatorics*, 10(2-4):105–121, June 1994. ISSN 0911-0119, 1435-5914. doi: 10.1007/BF02986655. URL https://link.springer.com/article/10.1007/BF02986655.

Xinbo Gao, Ya Su, Xuelong Li, and Dacheng Tao. A Review of Active Appearance Models. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, 40(2):145–158, March 2010. ISSN 1094-6977. doi: 10.1109/TSMCC.2009.2035631.

Pierre Geurts, Damien Ernst, and Louis Wehenkel. Extremely randomized trees. *Machine Learning*, 63(1):3–42, April 2006. ISSN 0885-6125, 1573-0565. doi: 10.1007/s10994-006-6226-1.

Polina Golland and Bruce Fischl. Permutation Tests for Classification: Towards Statistical Significance in Image-Based Studies. In *Information Processing in Medical Imaging*, Lecture Notes in Computer Science, pages 330–341. Springer, Berlin, Heidelberg, July 2003. ISBN 978-3-540-40560-3 978-3-540-45087-0. doi: 10.1007/978-3-540-45087-0_28.

Rafael C. Gonzalez and Richard E. Woods. *Digital Image Processing*. Pearson, Upper Saddle River, N.J, 3rd edition, August 2007. ISBN 978-0-13-168728-8.

Colin Goodall. Procrustes Methods in the Statistical Analysis of Shape. *Journal of the Royal Statistical Society. Series B (Methodological)*, 53(2):285–339, 1991. ISSN 0035-9246. URL http://www.jstor.org/stable/2345744.

J. C. Gower. Generalized procrustes analysis. *Psychometrika*, 40(1):33–51, March 1975. ISSN 0033-3123, 1860-0980. doi: 10.1007/BF02291478.

L. Grady. *Space-Variant Computer Vision: a Graph-Theoretic Approach*. PhD Thesis, Boston University, Boston, 2004.

DM Greig, BT Porteous, and AH Seheult. Exact maximum a posteriori estimation for binary images. *Journal of the Royal Statistical Society. Series B (Methodological)*, 51(2):271–279, 1989.

U. Grenander and D. Keenan. On the Shape of Plane Images. *SIAM Journal on Applied Mathematics*, 53(4):1072–1094, August 1993. ISSN 0036-1399. doi: 10.1137/0153054. URL http://epubs.siam.org/doi/abs/10.1137/0153054.

U. Grenander, M. I. Miller, and A. Srivastava. Hilbert-Schmidt lower bounds for estimators on matrix lie groups for ATR. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(8):790–802, August 1998. ISSN 0162-8828. doi: 10.1109/34.709572.

Ulf Grenander. *General Pattern Theory: A Mathematical Study of Regular Structures*. Oxford Mathematical Monographs. Oxford University Press, Oxford, New York, January 1994. ISBN 978-0-19-853671-0.

Ulf Grenander and Michael I. Miller. Computational Anatomy: an Emerging Discipline. *Quarterly of Applied Mathematics*, 56(4):617–694, 1998. ISSN 0033-569X. URL http://www.jstor.org/stable/43638257.

Ge Guo, TingTing Jiang, YiZhou Wang, and Wen Gao. 2-D shape completion with shape priors. *Chinese Science Bulletin*, 58(27):3430–3436, August 2013. ISSN 1001-6538, 1861-9541. doi: 10.1007/s11434-013-5946-4. URL http://link.springer.com/article/10.1007/s11434-013-5946-4.

Isabelle Guyon, Masoud Nikravesh, Steve Gunn, Lotfi A. Zadeh, and Janusz Kacprzyk, editors. *Feature Extraction*. Number 207 in Studies in Fuzziness and Soft Computing. Springer Berlin Heidelberg, Berlin, Heidelberg, 2006. ISBN 978-3-540-35488-8. URL http://link.springer.com/10.1007/978-3-540-35488-8. DOI: 10.1007/978-3-540-35488-8.

Robert M. Haralick and Linda G. Shapiro. *Computer and Robot Vision*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 1st edition, 1992. ISBN 978-0-201-56943-8.

Frank Harary. *Graph Theory*. Addison-Wesley Series in Mathematics. Addison-Wesley Publishing Company, 1969.

Frank Harary and Charles A. Trauth Jr. Connectedness of Products of Two Directed Graphs. *SIAM Journal on Applied Mathematics*, 14(2):250–254, March 1966. ISSN 0036-1399. doi: 10.1137/0114024. URL http://epubs.siam.org/doi/abs/10.1137/0114024.

Richard Hartley, Jochen Trumpf, Yuchao Dai, and Hongdong Li. Rotation Averaging. *International Journal of Computer Vision*, 103(3):267–305, July 2013. ISSN 0920-5691, 1573-1405. doi: 10.1007/s11263-012-0601-0. URL http://link.springer.com/article/10.1007/s11263-012-0601-0.

S. Hauberg, A. Feragen, and M. J. Black. Grassmann Averages for Scalable Robust PCA. In *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pages 3810–3817, June 2014. doi: 10.1109/CVPR.2014.481.

Ran He, Bao-Gang Hu, Wei-Shi Zheng, and Xiang-Wei Kong. Robust Principal Component Analysis Based on Maximum Correntropy Criterion. *IEEE Transactions on Image Processing*, 20(6):1485–1494, June 2011. ISSN 1057-7149. doi: 10.1109/TIP.2010.2103949.

Tobias Heimann and Hans-Peter Meinzer. Statistical shape models for 3d medical image segmentation: A review. *Medical Image Analysis*, 13(4):543–563, August 2009. ISSN 1361-8415. doi: 10.1016/j.media.2009.05.004. URL http://www.sciencedirect.com/science/article/pii/S1361841509000425.

Marc Hellmuth and Tilen Marc. On the Cartesian skeleton and the factorization of the strong product of digraphs. *Theoretical Computer Science*, 565:16–29, February 2015. ISSN 0304-3975. doi: 10.

1016/j.tcs.2014.10.045. URL http://www.sciencedirect.com/science/article/pii/S0304397514008378.

Wen Huang, Kyle A. Gallivan, Anuj Srivastava, and Pierre-Antoine Absil. Riemannian Optimization for Registration of Curves in Elastic Shape Analysis. *Journal of Mathematical Imaging and Vision*, 54(3): 320–343, March 2016. ISSN 0924-9907, 1573-7683. doi: 10.1007/s10851-015-0606-8. URL http://link.springer.com/article/10.1007/s10851-015-0606-8.

Peter J. Huber and Elvezio M. Ronchetti. *Robust Statistics*. Wiley, 2nd edition, March 2009. ISBN 978-0-470-12990-6.

Stephan Huckemann and Herbert Ziezold. Principal component analysis for Riemannian manifolds, with an application to triangular shape spaces. *Advances in Applied Probability*, 38(2):299–319, June 2006. ISSN 0001-8678, 1475-6064. doi: 10.1239/aap/1151337073. URL http://projecteuclid.org/euclid.aap/1151337073.

Stephan Huckemann, Thomas Hotz, and Axel Munk. Intrinsic Shape Analysis: Geodesic PCA for Riemannian Manifolds Module Isometric Lie Group Actions. *Statistica Sinica*, 20(1):1–58, 2010. ISSN 1017-0405. URL http://www.jstor.org/stable/24308976.

Yan Chai Hum, Khin Wee Lai, Nugraha Priya Utama, Maheza Irna Mohamad Salim, and Yin Mon Myint. Review on Segmentation of Computer-Aided Skeletal Maturity Assessment. In Khin Wee Lai, Yan Chai Hum, Maheza Irna Mohamad Salim, Sang-Bing Ong, Nugraha Priya Utama, Yin Mon Myint, Norliza Mohd Noor, and Eko Supriyanto, editors, *Advances in Medical Diagnostic Technology*, Lecture Notes in Bioengineering, pages 23–51. Springer Singapore, 1st edition, 2014. ISBN 978-981-4585-71-2.

Xiaoming Huo. *Spase image representation via combined transforms*. PhD Thesis, Stanford University, August 1999.

F. Hutter, H. H. Hoos, and K. Leyton-Brown. Sequential model-based optimization for general algorithm configuration. In *Proc. of LION-5*, pages 507–523, 2011.

Frank Hutter, Holger H. Hoos, Kevin Leyton-Brown, and Thomas Stützle. ParamILS: An Automatic Algorithm Configuration Framework. *J. Artif. Int. Res.*, 36(1):267–306, September 2009. ISSN 1076-9757.

Paul Jaccard. Distribution comparée de la flore alpine dans quelques régions des Alpes occidentales et orientales. *Bulletin de la Société Vaudoise des Sciences Naturelles*, 37(140):241–272, 1901.

Bernd Jähne. *Digital Image Processing*. Springer-Verlag Berlin Heidelberg, 5th edition, 2002. ISBN 978-3-662-04781-1. URL http://www.springer.com/gp/book/9783662047811.

Nathalie Japkowicz and Mohak Shah. *Evaluating Learning Algorithms: A Classification Perspective*. Cambridge University Press, 1st edition, 2011. ISBN 978-0-521-19600-0.

I.T. Jolliffe. *Principal Component Analysis*. Springer Series in Statistics. Springer-Verlag New York, 2nd edition, 2002. ISBN 978-0-387-95442-4.

Eric Jones, Travis Oliphant, Pearu Peterson, et al. SciPy: Open source scientific tools for Python, 2001–. URL http://www.scipy.org/. [Online; accessed 2017-11-28].

Shantanu H. Joshi, Eric Klassen, Anuj Srivastava, and Ian Jermyn. A Novel Representation for Riemannian Analysis of Elastic Curves in Rn. In *Proceedings / CVPR, IEEE Computer Society Con-*

ference on Computer Vision and Pattern Recognition, pages 1–7, Minneapolis, July 2007. IEEE. doi: 10.1109/CVPR.2007.383185.

Sungkyu Jung, Ian L. Dryden, and J. S. Marron. Analysis of principal nested spheres. *Biometrika*, 99(3): 551–568, September 2012. ISSN 0006-3444. doi: 10.1093/biomet/ass022. URL http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3635703/.

Matti Kääriäinen and John Langford. A Comparison of Tight Generalization Error Bounds. In *Proceedings of the 22nd International Conference on Machine Learning*, ICML '05, pages 409–416, New York, NY, USA, 2005. ACM. ISBN 978-1-59593-180-1. doi: 10.1145/1102351.1102403.

H. Karcher. Riemannian center of mass and mollifier smoothing. *Communications on Pure and Applied Mathematics*, 30(5):509–541, September 1977. ISSN 1097-0312. doi: 10.1002/cpa.3160300502. URL http://onlinelibrary.wiley.com/doi/10.1002/cpa.3160300502/abstract.

Michael Kass, Andrew Witkin, and Demetri Terzopoulos. Snakes: Active contour models. *International Journal of Computer Vision*, 1(4):321–331, January 1988. ISSN 0920-5691, 1573-1405. doi: 10.1007/BF00133570. URL https://link.springer.com/article/10.1007/BF00133570.

D. G. Kendall, D. Barden, T. K. Carne, and H. Le. *Shape and Shape Theory*. Wiley, New York, 1st edition, October 1999. ISBN 978-0-471-96823-8.

David G. Kendall. Shape Manifolds, Procrustean Metrics, and Complex Projective Spaces. *Bulletin of the London Mathematical Society*, 16(2):81–121, March 1984. ISSN 0024-6093, 1469-2120. doi: 10.1112/blms/16.2.81. URL http://blms.oxfordjournals.org/content/16/2/81.

David G. Kendall. A Survey of the Statistical Theory of Shape. *Statistical Science*, 4(2):87–99, May 1989. ISSN 0883-4237, 2168-8745. doi: 10.1214/ss/1177012582. URL http://projecteuclid.org/euclid.ss/1177012582.

Wilfrid S. Kendall. Probability, Convexity, and Harmonic Maps with Small Image I: Uniqueness and Fine Existence. *Proceedings of the London Mathematical Society*, s3-61(2):371–406, September 1990. ISSN 0024-6115, 1460-244X. doi: 10.1112/plms/s3-61.2.371. URL http://plms.oxfordjournals.org/content/s3-61/2/371.

N. Khaneja, M. I. Miller, and U. Grenander. Dynamic programming generation of curves on brain surfaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(11):1260–1265, November 1998. ISSN 0162-8828. doi: 10.1109/34.730559.

Davis E. King. Dlib-ml: A Machine Learning Toolkit. *Journal of Machine Learning Research*, 10:1755–1758, 2009.

E. Klassen, A. Srivastava, M. Mio, and S. H. Joshi. Analysis of planar shapes using geodesic paths on shape spaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(3):372–383, March 2004. ISSN 0162-8828. doi: 10.1109/TPAMI.2004.1262333.

Ron Kohavi. A Study of Cross-validation and Bootstrap for Accuracy Estimation and Model Selection. In *Proceedings of the 14th International Joint Conference on Artificial Intelligence - Volume 2*, pages 1137–1143, San Francisco, CA, USA, 1995. Morgan Kaufmann Publishers Inc. ISBN 1-55860-363-8.

Martin A. Koschat and Deborah F. Swayne. A weighted procrustes criterion. *Psychometrika*, 56(2):229–239, June 1991. ISSN 0033-3123, 1860-0980. doi: 10.1007/BF02294460. URL http://link.springer.com/article/10.1007/BF02294460.

Andreas Kriegl and Peter W Michor. Aspects of the theory of infinite dimensional manifolds. *Differential Geometry and its Applications*, 1(2):159–176, September 1991. ISSN 0926-2245. doi: 10.1016/0926-2245(91)90029-9. URL http://www.sciencedirect.com/science/article/pii/0926224591900299.

Andreas Kriegl and Peter W. Michor. *The Convenient Setting of Global Analysis*. Number 53 in Mathematical Surveys and Monographs. American Mathematical Society, Providence, R.I, September 1997. ISBN 978-0-8218-0780-4.

Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. ImageNet Classification with Deep Convolutional Neural Networks. In F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 25*, pages 1097–1105. Curran Associates, Inc., 2012.

Casimir Kuratowski. Sur le problème des courbes gauches en Topologie. *Fundamenta Mathematicae*, 15 (1):271–283, 1930. ISSN 0016-2736. URL https://eudml.org/doc/212352.

Sebastian Kurtek, Anuj Srivastava, Eric Klassen, and Zhaohua Ding. Statistical Modeling of Curves Using Shapes and Related Features. *Journal of the American Statistical Association*, 107(499):1152–1165, September 2012. ISSN 0162-1459. doi: 10.1080/01621459.2012.699770. URL http://amstat.tandfonline.com/doi/abs/10.1080/01621459.2012.699770.

Sebastian Kurtek, Anuj Srivastava, Eric Klassen, and Hamid Laga. Landmark-Guided Elastic Shape Analysis of Spherically-Parameterized Surfaces. *Computer Graphics Forum*, 32(2pt4):429–438, May 2013. ISSN 1467-8659. doi: 10.1111/cgf.12063. URL http://onlinelibrary.wiley.com/doi/10.1111/cgf.12063/abstract.

N. Kwak. Principal Component Analysis Based on L1-Norm Maximization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(9):1672–1680, September 2008. ISSN 0162-8828. doi: 10.1109/TPAMI.2008.114.

John D. Lafferty, Andrew McCallum, and Fernando C. N. Pereira. Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data. In *Proceedings of the Eighteenth International Conference on Machine Learning*, ICML '01, pages 282–289, San Francisco, CA, USA, 2001. Morgan Kaufmann Publishers Inc. ISBN 978-1-55860-778-1.

Serge Lang. *Real analysis*. Addison-Wesley, Advanced Book Program/World Science Division, 1983. ISBN 978-0-201-14179-5.

Serge Lang. *Differential Manifolds*. Springer, New York, 2nd edition, January 1985. ISBN 978-0-387-96113-2.

Serge Lang. *Differential and Riemannian Manifolds*. Springer, New York, 3rd edition, April 1996. ISBN 978-0-387-94338-1.

John Langford. Tutorial on Practical Prediction Theory for Classification. *Journal of Machine Learning Research*, 6:273–306, December 2005. ISSN 1532-4435. URL http://www.jmlr.org/papers/v6/langford05a.html.

Drew Lazar and Lizhen Lin. Scale and curvature effects in principal geodesic analysis. *Journal of Multivariate Analysis*, 153:64–82, January 2017. ISSN 0047-259X. doi: 10.1016/j.jmva.2016.09.009. URL http://www.sciencedirect.com/science/article/pii/S0047259X16300926.

H. Le and A. Kume. Detection of shape changes in biological features. *Journal of Microscopy*, 200(Pt 2): 140–147, November 2000. ISSN 0022-2720.

Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *Nature*, 521:436–444, May 2015. ISSN 1476-4687. doi: 10.1038/nature14539.

Reiner Lenz. *Group Theoretical Methods in Image Processing*. Number 413 in Lecture Notes in Computer Science. Springer Berlin Heidelberg, 1990. ISBN 978-3-540-52290-4. URL http://www.springer.com/gp/book/9783540522904.

A. Levinshtein, A. Stere, K. N. Kutulakos, D. J. Fleet, S. J. Dickinson, and K. Siddiqi. TurboPixels: Fast Superpixels Using Geometric Flows. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(12):2290–2297, December 2009. ISSN 0162-8828. doi: 10.1109/TPAMI.2009.96.

X. Li, Y. Li, C. Shen, A. Dick, and A. V. D. Hengel. Contextual Hypergraph Modeling for Salient Object Detection. In *2013 IEEE International Conference on Computer Vision*, pages 3328–3335, December 2013. doi: 10.1109/ICCV.2013.413.

Yin Li, Jian Sun, Chi-Keung Tang, and Heung-Yeung Shum. Lazy Snapping. In *ACM SIGGRAPH 2004 Papers*, SIGGRAPH '04, pages 303–308, New York, NY, USA, 2004. ACM. doi: 10.1145/1186562.1015719.

Z. Li, X. M. Wu, and S. F. Chang. Segmentation using superpixels: A bipartite graph partitioning approach. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pages 789–796, June 2012. doi: 10.1109/CVPR.2012.6247750.

C. E. Liedtke, T. Gahm, F. Kappei, and B. Aeikens. Segmentation of microscopic cell scenes. *Analytical and Quantitative Cytology and Histology*, 9(3):197–211, June 1987. ISSN 0884-6812.

T. Lin and H. Zha. Riemannian Manifold Learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(5):796–809, May 2008. ISSN 0162-8828. doi: 10.1109/TPAMI.2007.70735.

C. Lindner, P. A. Bromiley, M. C. Ionita, and T. F. Cootes. Robust and Accurate Shape Model Matching Using Random Forest Regression-Voting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(9):1862–1874, September 2015. ISSN 0162-8828. doi: 10.1109/TPAMI.2014.2382106.

Marion Lineberry. Image Segmentation By Edge Tracing. In *Proc. SPIE*, volume 0359, pages 361–368, 1983. doi: 10.1117/12.965986.

Wei Liu, Anuj Srivastava, and Jinfeng Zhang. Protein Structure Alignment Using Elastic Shape Analysis. In *Proceedings of the First ACM International Conference on Bioinformatics and Computational Biology*, BCB '10, pages 62–70, New York, NY, USA, 2010. ACM. ISBN 978-1-4503-0438-2. doi: 10.1145/1854776.1854790.

D. Lunga, S. Prasad, M. M. Crawford, and O. Ersoy. Manifold-Learning-Based Feature Extraction for Classification of Hyperspectral Data: A Review of Advances in Manifold Learning. *IEEE Signal Processing Magazine*, 31(1):55–66, January 2014. ISSN 1053-5888. doi: 10.1109/MSP.2013.2279894.

P. Mainali, G. Lafruit, K. Tack, L. Van Gool, and R. Lauwereins. Derivative-Based Scale Invariant Image Feature Detector With Error Resilience. *IEEE Transactions on Image Processing*, 23(5):2380–2391, May 2014. ISSN 1057-7149. doi: 10.1109/TIP.2014.2315959.

M. Mani, S. Kurtek, C. Barillot, and A. Srivastava. A comprehensive riemannian framework for the analysis of white matter fiber tracts. In *2010 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pages 1101–1104, April 2010. doi: 10.1109/ISBI.2010.5490185.

Oded Maron and Andrew W. Moore. Hoeffding Races: Accelerating Model Selection Search for Classification and Function Approximation. In J. D. Cowan, G. Tesauro, and J. Alspector, editors, *Advances in Neural Information Processing Systems 6*, pages 59–66. Morgan-Kaufmann, 1994.

Alberto Martelli. An Application of Heuristic Search Methods to Edge and Contour Detection. *Commun. ACM*, 19(2):73–83, February 1976. ISSN 0001-0782. doi: 10.1145/359997.360004.

P. Martins, J. F. Henriques, R. Caseiro, and J. Batista. Bayesian Constrained Local Models Revisited. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(4):704–716, April 2016. ISSN 0162-8828. doi: 10.1109/TPAMI.2015.2462343.

M. H. McAndrew. On the Product of Directed Graphs. *Proceedings of the American Mathematical Society*, 14(4):600–606, 1963. ISSN 0002-9939. doi: 10.2307/2034282. URL `http://www.jstor.org/stable/2034282`.

R. Mehrotra and J. E. Gary. Similar-shape retrieval in shape data management. *Computer*, 28(9):57–62, September 1995. ISSN 0018-9162. doi: 10.1109/2.410154.

Andrea C. G. Mennucci. Metrics of Curves in Shape Optimization and Analysis. In *Level Set and PDE Based Reconstruction Methods in Imaging*, Lecture Notes in Mathematics, pages 205–319. Springer, Cham, 2013. ISBN 978-3-319-01711-2 978-3-319-01712-9. doi: 10.1007/978-3-319-01712-9_4. URL `https://link.springer.com/chapter/10.1007/978-3-319-01712-9_4`.

Pablo Mesejo, Roberto Ugolotti, Ferdinando Di Cunto, Mario Giacobini, and Stefano Cagnoni. Automatic hippocampus localization in histological images using Differential Evolution-based deformable models. *Pattern Recognition Letters*, 34(3):299–307, February 2013. ISSN 0167-8655. doi: 10.1016/j.patrec.2012.10.012.

F. Meyer and S. Beucher. Morphological segmentation. *Journal of Visual Communication and Image Representation*, 1(1):21–46, September 1990. ISSN 10473203. doi: 10.1016/1047-3203(90)90014-M.

Fernand Meyer. Topographic distance and watershed lines. *Signal Processing*, 38(1):113–125, July 1994. ISSN 0165-1684. doi: 10.1016/0165-1684(94)90060-4. URL `http://www.sciencedirect.com/science/article/pii/0165168494900604`.

Fernand Meyer. Morphological segmentation revisited. In Michel Bilodeau, Fernand Meyer, and Michel Schmitt, editors, *Space, Structure and Randomness*, number 183 in Lecture Notes in Statistics, pages 315–347. Springer New York, 2005. ISBN 978-0-387-20331-7 978-0-387-29115-4. doi: 10.1007/0-387-29115-6_13. URL `http://link.springer.com/chapter/10.1007/0-387-29115-6_13`.

Peter W. Michor and David B. Mumford. Riemannian geometries on spaces of plane curves. *Journal of the European Mathematical Society*, 8(1):1–48, 2006. ISSN 1435-9855. doi: 10.4171/JEMS/37.

Krystian Mikolajczyk, Andrew Zisserman, and Cordelia Schmid. Shape recognition with edge-based features. In Richard Harvey and Andrew Bangham, editors, *Proceedings of British Machine Vision Conference (BMVC '03)*, volume 2, pages 779–788, Norwich, United Kingdom, September 2003. The British Machine Vision Association. URL https://hal.inria.fr/inria-00548226.

Michael I. Miller and Anqi Qiu. The emerging discipline of Computational Functional Anatomy. *NeuroImage*, 45(1, Supplement 1):S16–S39, March 2009. ISSN 1053-8119. doi: 10.1016/j.neuroimage.2008.10.044. URL http://www.sciencedirect.com/science/article/pii/S1053811908011622.

Rupert G. Miller. The jackknife-a review. *Biometrika*, 61(1):1–15, April 1974. ISSN 0006-3444. doi: 10.1093/biomet/61.1.1. URL https://academic.oup.com/biomet/article/61/1/1/264346/The-jackknife-a-review.

W. Mio and A. Srivastava. Elastic-string models for representation and analysis of planar shapes. In *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004.*, volume 2, pages II–10–II–15, June 2004. doi: 10.1109/CVPR.2004.1315138.

Washington Mio, Anuj Srivastava, and Shantanu Joshi. On Shape of Plane Elastic Curves. *International Journal of Computer Vision*, 73(3):307–324, July 2007. ISSN 0920-5691, 1573-1405. doi: 10.1007/s11263-006-9968-0.

Mehryar Mohri, Afshin Rostamizadeh, and Ameet Talwalkar. *Foundations of Machine Learning*. The MIT Press, Cambridge, MA, August 2012. ISBN 978-0-262-01825-8.

Raouia Mokni, Hassen Drira, and Monji Kherallah. Combining shape analysis and texture pattern for palmprint identification. *Multimedia Tools and Applications*, pages 1–28, November 2016. ISSN 1380-7501, 1573-7721. doi: 10.1007/s11042-016-4088-5. URL https://link.springer.com/article/10.1007/s11042-016-4088-5.

A. P. Moore, S. J. D. Prince, J. Warrell, U. Mohammed, and G. Jones. Superpixel lattices. In *2008 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8, June 2008. doi: 10.1109/CVPR.2008.4587471.

Andrew W. Moore and Mary S. Lee. Efficient Algorithms for Minimizing Cross Validation Error. In *Proceedings of the Eleventh International Conference on International Conference on Machine Learning*, ICML'94, pages 190–198, San Francisco, CA, USA, 1994. Morgan Kaufmann Publishers Inc. ISBN 978-1-55860-335-6.

G. Mori, Xiaofeng Ren, A. A. Efros, and J. Malik. Recovering human body configurations: combining segmentation and recognition. In *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004.*, volume 2, pages II–326–II–333, June 2004. doi: 10.1109/CVPR.2004.1315182.

A. Mottini, X. Descombes, and F. Besse. From Curves to Trees: A Tree-like Shapes Distance Using the Elastic Shape Analysis Framework. *Neuroinformatics*, 13(2):175–191, April 2015. ISSN 1539-2791, 1559-0089. doi: 10.1007/s12021-014-9255-0. URL http://link.springer.com/article/10.1007/s12021-014-9255-0.

Dibyendu Mukherjee, Q. M. Jonathan Wu, and Guanghui Wang. A comparative experimental study of image feature detectors and descriptors. *Machine Vision and Applications*, 26(4):443–466, May 2015. ISSN 0932-8092, 1432-1769. doi: 10.1007/s00138-015-0679-9.

Priyanka Mukhopadhyay and Bidyut B. Chaudhuri. A survey of Hough Transform. *Pattern Recognition*, 48:993–1010, 2015. URL http://www.sciencedirect.com/science/article/pii/S0031320314003446.

Kevin P. Murphy. *Machine Learning: A Probabilistic Perspective*. Adaptive Computation and Machine Learning. The MIT Press, Cambridge, MA, 1st edition, August 2012. ISBN 978-0-262-01802-9.

Delphine Nain, Steven Haker, Aaron Bobick, and Allen Tannenbaum. Multiscale 3-D shape representation and segmentation using spherical wavelets. *IEEE transactions on medical imaging*, 26(4):598–618, April 2007. ISSN 0278-0062. doi: 10.1109/TMI.2007.893284.

Laurent Najman and Michel Schmitt. Watershed of a Continuous Function. *Signal Process.*, 38(1):99–112, July 1994. ISSN 0165-1684. doi: 10.1016/0165-1684(94)90059-0.

F. Natterer. A Sobolev Space Analysis of Picture Reconstruction. *SIAM Journal on Applied Mathematics*, 39(3):402–411, December 1980. ISSN 0036-1399. doi: 10.1137/0139034. URL http://epubs.siam.org/doi/abs/10.1137/0139034.

P. Neubert and P. Protzel. Compact Watershed and Preemptive SLIC: On Improving Trade-offs of Superpixel Segmentation Algorithms. In *2014 22nd International Conference on Pattern Recognition (ICPR)*, pages 996–1001, August 2014. doi: 10.1109/ICPR.2014.181.

Lenhard L. Ng. Hamiltonian Decomposition of Lexicographic Products of Digraphs. *Journal of Combinatorial Theory, Series B*, 73(2):119–129, July 1998. ISSN 0095-8956. doi: 10.1006/jctb.1998.1816. URL http://www.sciencedirect.com/science/article/pii/S0095895698918160.

J. A. Noble and D. Boukerroui. Ultrasound image segmentation: a survey. *IEEE Transactions on Medical Imaging*, 25(8):987–1010, August 2006. ISSN 0278-0062. doi: 10.1109/TMI.2006.877092.

Stanley Osher and Ronald Fedkiw. *Level Set Methods and Dynamic Implicit Surfaces*. Number 153 in Applied Mathematical Sciences. Springer New York, 2003. ISBN 978-0-387-22746-7.

Stanley Osher and Ronald P. Fedkiw. Level Set Methods: An Overview and Some Recent Results. *Journal of Computational Physics*, 169(2):463–502, May 2001. ISSN 0021-9991. doi: 10.1006/jcph.2000.6636. URL http://www.sciencedirect.com/science/article/pii/S0021999100966361.

Stanley Osher and James A Sethian. Fronts propagating with curvature-dependent speed: Algorithms based on Hamilton-Jacobi formulations. *Journal of Computational Physics*, 79(1):12–49, November 1988. ISSN 0021-9991. doi: 10.1016/0021-9991(88)90002-2. URL http://www.sciencedirect.com/science/article/pii/0021999188900022.

Nikhil R. Pal and Dinabandhu Bhandari. Image thresholding: Some new techniques. *Signal Processing*, 33(2):139–158, August 1993. ISSN 0165-1684. doi: 10.1016/0165-1684(93)90107-L. URL http://www.sciencedirect.com/science/article/pii/016516849390107L.

Nikhil R Pal and Sankar K Pal. Entropic thresholding. *Signal Processing*, 16(2):97–108, February 1989. ISSN 0165-1684. doi: 10.1016/0165-1684(89)90090-X. URL http://www.sciencedirect.com/science/article/pii/016516848990090X.

F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

Xavier Pennec. Barycentric Subspaces and Affine Spans in Manifolds. In Frank Nielsen and Frédéric Barbaresco, editors, *Geometric Science of Information*, number 9389 in Lecture Notes in Computer Science, pages 12–21. Springer International Publishing, October 2015. ISBN 978-3-319-25039-7 978-3-319-25040-3. doi: 10.1007/978-3-319-25040-3_2. URL http://link.springer.com/chapter/10.1007/978-3-319-25040-3_2.

P. Perona and W. Freeman. A factorization approach to grouping. In *Computer Vision — ECCV'98*, pages 655–670. Springer, Berlin, Heidelberg, June 1998. doi: 10.1007/BFb0055696. URL https://link.springer.com/chapter/10.1007/BFb0055696.

P. Perona and J. Malik. Scale-space and edge detection using anisotropic diffusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(7):629–639, July 1990. ISSN 0162-8828. doi: 10.1109/34.56205.

Tomasz Pięciak, Mateusz Baran, and Michał Ubrańczyk. Level-set based segmentation of carotid arteries in computer tomography angiography images. *Journal of Medical Informatics & Technologies*, 17:281–286, 2011. ISSN 1642-6037.

John C. Platt. Probabilistic Outputs for Support Vector Machines and Comparisons to Regularized Likelihood Methods. In *Advances in Large Margin Classifiers*, Neural Information Processing Systems, pages 61–74. MIT Press, 1999. ISBN 0-262-19448-1.

Irwin Pollack and Louis R. Decker. Confidence Ratings, Message Reception, and the Receiver Operating Characteristic. *The Journal of the Acoustical Society of America*, 30(4):286–292, April 1958. ISSN 0001-4966. doi: 10.1121/1.1909571. URL http://asa.scitation.org/doi/abs/10.1121/1.1909571.

William K. Pratt. *Digital Image Processing*. John Wiley & Sons, Inc., New York, NY, USA, 1978. ISBN 978-0-471-01888-9.

William K. Pratt. *Digital Image Processing: PIKS Scientific Inside*. Wiley-Interscience, Hoboken, N.J, 4th edition, February 2007. ISBN 978-0-471-76777-0.

Alexander Prokhorov. Communicating Vessels. In Alexander Prokhorov, editor, *The Great Soviet Encyclopedia*. Soviet Encyclopedia Publishing House, 3rd edition, 1979. Retrieved 2017-09-21 from http://encyclopedia2.thefreedictionary.com/Communicating+Vessels.

M. H. Quenouille. Approximate Tests of Correlation in Time-Series. *Journal of the Royal Statistical Society. Series B (Methodological)*, 11(1):68–84, 1949. ISSN 0035-9246. URL http://www.jstor.org/stable/2983696.

Radhakrishna Rao. Information and the accuracy attainable in the estimation of statistical parameters. *Bull. Calcutta Math. Soc.*, 37:81–91, 1945. URL http://www.ams.org/mathscinet-getitem?mr=0015748.

X. Ren and J. Malik. Learning a classification model for segmentation. In *Proceedings Ninth IEEE International Conference on Computer Vision*, volume 1, pages 10–17, October 2003. doi: 10.1109/ICCV.2003.1238308.

Soufiane Rital, Alain Bretto, Driss Aboutajdine, and Hocine Cherifi. Application of Adaptive Hypergraph Model to Impulsive Noise Detection. In *Computer Analysis of Images and Patterns*, Lecture Notes in Computer Science, pages 555–562. Springer, Berlin, Heidelberg, September 2001. ISBN 978-3-540-42513-7 978-3-540-44692-7. doi: 10.1007/3-540-44692-3_67. URL https://link.springer.com/chapter/10.1007/3-540-44692-3_67.

Robert Rosenthal. The "File Drawer Problem" and Tolerance for Null Results. *Psychological Bulletin*, 86 (3):638–641, May 1979. doi: 10.1037/0033-2909.86.3.638.

Walter Rudin. *Real and Complex Analysis*. McGraw-Hill Education, New York, 3rd edition, May 1986. ISBN 978-0-07-054234-1.

Walter Rudin. *Functional Analysis*. McGraw-Hill Science/Engineering/Math, New York, 2nd edition, January 1991. ISBN 978-0-07-054236-5.

Stuart Russell and Peter Norvig. *Artificial Intelligence: A Modern Approach*. Pearson, Upper Saddle River, 3rd edition, December 2009. ISBN 978-0-13-604259-4.

P. K Sahoo, S Soltani, and A. K. C Wong. A survey of thresholding techniques. *Computer Vision, Graphics, and Image Processing*, 41(2):233–260, February 1988. ISSN 0734-189X. doi: 10.1016/0734-189X(88)90022-9. URL http://www.sciencedirect.com/science/article/pii/0734189X88900229.

S. Sarkar and P. Soundararajan. Supervised learning of large perceptual organization: graph spectral partitioning and learning automata. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(5):504–525, May 2000. ISSN 0162-8828. doi: 10.1109/34.857006.

J. Sauvola and M. Pietikäinen. Adaptive document image binarization. *Pattern Recognition*, 33(2):225–236, February 2000. ISSN 0031-3203. doi: 10.1016/S0031-3203(99)00055-2. URL http://www.sciencedirect.com/science/article/pii/S0031320399000552.

Robert E. Schapire, Yoav Freund, Peter Bartlett, and Wee Sun Lee. Boosting the margin: a new explanation for the effectiveness of voting methods. *The Annals of Statistics*, 26(5):1651–1686, October 1998. doi: 10.1214/aos/1024691352.

Julia A. Schnabel and Simon R. Arridge. Active Contour Models for Shape Description Using Multiscale Differential Invariants. In *Proceedings of British Machine Vision Conference 1995*, volume 1, pages 197–206, Birmingham, UK, 1995.

T. Schoenemann, F. Kahl, and D. Cremers. Curvature regularity for region-based image segmentation and inpainting: A linear programming relaxation. In *Computer Vision, 2009 IEEE 12th International Conference on*, pages 17–23, Kyoto, Japan, September 2009. IEEE. doi: 10.1109/ICCV.2009.5459209.

T. B. Sebastian, P. N. Klein, and B. B. Kimia. On aligning curves. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(1):116–125, January 2003. ISSN 0162-8828. doi: 10.1109/TPAMI.2003.1159951.

E. Sharon and D. Mumford. 2d-Shape Analysis Using Conformal Mapping. *International Journal of Computer Vision*, 70(1):55–75, October 2006. ISSN 0920-5691, 1573-1405. doi: 10.1007/s11263-006-6121-z. URL https://link.springer.com/article/10.1007/s11263-006-6121-z.

Jianbo Shi and Jitendra Malik. Normalized Cuts and Image Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):888–905, August 2000. ISSN 0162-8828. doi: 10.1109/34.868688.

J. Shukla, M. Alwani, and A. K. Tiwari. A survey on lossless image compression methods. In *2010 2nd International Conference on Computer Engineering and Technology*, volume 6, pages V6–136–V6–141, April 2010. doi: 10.1109/ICCET.2010.5486344.

Abul Hasan Siddiqi. *Applied Functional Analysis: Numerical Methods, Wavelet Methods, and Image Processing*. CRC Press, New York, 1st edition, September 2003. ISBN 978-0-8247-4097-9.

Tobias Sing, Oliver Sander, Niko Beerenwinkel, and Thomas Lengauer. ROCR: visualizing classifier performance in R. *Bioinformatics*, 21(20):3940–3941, October 2005. ISSN 1367-4803. doi: 10.1093/bioinformatics/bti623.

L. Sirovich and M. Kirby. Low-dimensional procedure for the characterization of human faces. *Journal of the Optical Society of America A*, 4(3):519, March 1987. ISSN 1084-7529, 1520-8532. doi: 10.1364/JOSAA.4.000519. URL https://www.osapublishing.org/abstract.cfm?URI=josaa-4-3-519.

Stephen P. Smith and Anil K. Jain. Chord distributions for shape matching. *Computer Graphics and Image Processing*, 20(3):259–271, November 1982. ISSN 0146-664X. doi: 10.1016/0146-664X(82)90084-3. URL http://www.sciencedirect.com/science/article/pii/0146664X82900843.

J. L. Sobrinho. Algebra and algorithms for QoS path computation and hop-by-hop routing in the Internet. *IEEE/ACM Transactions on Networking*, 10(4):541–550, August 2002. ISSN 1063-6692. doi: 10.1109/TNET.2002.801397.

J. L. Sobrinho. An algebraic theory of dynamic network routing. *IEEE/ACM Transactions on Networking*, 13(5):1160–1173, October 2005. ISSN 1063-6692. doi: 10.1109/TNET.2005.857111.

João Luis Sobrinho. Network Routing with Path Vector Protocols: Theory and Applications. In *Proceedings of the 2003 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications*, SIGCOMM '03, pages 49–60, New York, NY, USA, 2003. ACM. ISBN 978-1-58113-735-4. doi: 10.1145/863955.863963.

S. Sommer, A. Tatu, Chen Chen, D. R. Jurgensen, M. de Bruijne, M. Loog, M. Nielsen, and F. Lauze. Bicycle chain shape models. In *2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, pages 157–163, June 2009. doi: 10.1109/CVPRW.2009.5204053.

S. Sommer, F. Lauze, and M. Nielsen. Optimization over geodesics for exact principal geodesic analysis. *Advances in Computational Mathematics*, 40(2):283–313, June 2013. ISSN 1019-7168, 1572-9044. doi: 10.1007/s10444-013-9308-1. URL http://link.springer.com/article/10.1007/s10444-013-9308-1.

Stefan Sommer. Horizontal Dimensionality Reduction and Iterated Frame Bundle Development. In Frank Nielsen and Frédéric Barbaresco, editors, *Geometric Science of Information*, number 8085 in Lecture

Notes in Computer Science, pages 76–83. Springer Berlin Heidelberg, 2013. ISBN 978-3-642-40019-3 978-3-642-40020-9. doi: 10.1007/978-3-642-40020-9_7. URL http://link.springer.com/chapter/10.1007/978-3-642-40020-9_7.

Stefan Sommer, François Lauze, Søren Hauberg, and Mads Nielsen. Manifold Valued Statistics, Exact Principal Geodesic Analysis and the Effect of Linear Approximations. In Kostas Daniilidis, Petros Maragos, and Nikos Paragios, editors, *Computer Vision – ECCV 2010*, number 6316 in Lecture Notes in Computer Science, pages 43–56. Springer Berlin Heidelberg, September 2010. ISBN 978-3-642-15566-6 978-3-642-15567-3. doi: 10.1007/978-3-642-15567-3_4. URL http://link.springer.com/chapter/10.1007/978-3-642-15567-3_4.

Michael Spivak. *A Comprehensive Introduction to Differential Geometry*, volume 1. Publish or Perish, Houston, Tex, 3rd edition, January 1999. ISBN 978-0-914098-70-6.

A. Srivastava and I. H. Jermyn. Looking for Shapes in Two-Dimensional Cluttered Point Clouds. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(9):1616–1629, September 2009. ISSN 0162-8828. doi: 10.1109/TPAMI.2008.223.

A. Srivastava, E. Klassen, S. H. Joshi, and I. H. Jermyn. Shape Analysis of Elastic Curves in Euclidean Spaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(7):1415–1428, July 2011. ISSN 0162-8828. doi: 10.1109/TPAMI.2010.184.

Anuj Srivastava and Eric P. Klassen. *Functional and Shape Data Analysis*. Springer Series in Statistics. Springer New York, 2016. ISBN 978-1-4939-4018-9.

Anuj Srivastava, Aastha Jain, Shantanu Joshi, and David Kaziska. Statistical Shape Models Using Elastic-String Representations. In *Computer Vision – ACCV 2006*, Lecture Notes in Computer Science, pages 612–621. Springer, Berlin, Heidelberg, January 2006. ISBN 978-3-540-31219-2 978-3-540-32433-1. doi: 10.1007/11612032_62.

Anuj Srivastava, Pavan Turaga, and Sebastian Kurtek. On advances in differential-geometric approaches for 2d and 3d shape analyses and activity recognition. *Image and Vision Computing*, 30(6–7):398–416, June 2012. ISSN 0262-8856. doi: 10.1016/j.imavis.2012.03.006.

M. Stone. Cross-Validatory Choice and Assessment of Statistical Predictions. *Journal of the Royal Statistical Society. Series B (Methodological)*, 36(2):111–147, 1974. ISSN 0035-9246. URL http://www.jstor.org/stable/2984809.

Arthur N. Strahler. Quantitative analysis of watershed geomorphology. *Eos, Transactions American Geophysical Union*, 38(6):913–920, December 1957. ISSN 2324-9250. doi: 10.1029/TR038i006p00913. URL http://onlinelibrary.wiley.com/doi/10.1029/TR038i006p00913/abstract.

J. Su, I. L. Dryden, E. Klassen, H. Le, and A. Srivastava. Fitting smoothing splines to time-indexed, noisy points on nonlinear manifolds. *Image and Vision Computing*, 30(6):428–442, June 2012. ISSN 0262-8856. doi: 10.1016/j.imavis.2011.09.006. URL http://www.sciencedirect.com/science/article/pii/S0262885611000990.

Jingyong Su, Sebastian Kurtek, Eric Klassen, and Anuj Srivastava. Statistical analysis of trajectories on Riemannian manifolds: Bird migration, hurricane tracking and video surveillance. *The Annals of Applied*

Statistics, 8(1):530–552, March 2014. ISSN 1932-6157, 1941-7330. doi: 10.1214/13-AOAS701. URL https://projecteuclid.org/euclid.aoas/1396966297.

Zhe Su, Eric Klassen, and Martin Bauer. The Square Root Velocity Framework for Curves in a Homogeneous Space. *arXiv:1706.03095 [math]*, June 2017. URL http://arxiv.org/abs/1706.03095. arXiv: 1706.03095.

Zihua Su, Tryphon Lambrou, and Andrew Todd-Pokropek. Principal Geodesic Analysis for the Study of Nonlinear Minimum Description Length. In *Medical Imaging and Informatics*, pages 89–98. Springer Berlin Heidelberg, 2008. doi: 10.1007/978-3-540-79490-5_13. URL http://link.springer.com/chapter/10.1007/978-3-540-79490-5_13.

G. Sundaramoorthi, J. D. Jackson, A. Yezzi, and A. C. Mennucci. Tracking With Sobolev Active Contours. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, volume 1, pages 674–680, June 2006. doi: 10.1109/CVPR.2006.314.

G. Sundaramoorthi, A. Mennucci, S. Soatto, and A. Yezzi. A New Geometric Metric in the Space of Curves, and Applications to Tracking Deforming Objects by Prediction and Filtering. *SIAM Journal on Imaging Sciences*, 4(1):109–145, January 2011. doi: 10.1137/090781139.

Ganesh Sundaramoorthi, Anthony Yezzi, and Andrea C. Mennucci. Sobolev Active Contours. *International Journal of Computer Vision*, 73(3):345–366, July 2007. ISSN 0920-5691, 1573-1405. doi: 10.1007/s11263-006-0635-2. URL https://link.springer.com/article/10.1007/s11263-006-0635-2.

Najiba Tagougui, Monji Kherallah, and Adel M. Alimi. Online Arabic handwriting recognition: a survey. *International Journal on Document Analysis and Recognition (IJDAR)*, 16(3):209–226, September 2013. ISSN 1433-2833, 1433-2825. doi: 10.1007/s10032-012-0186-8.

Xue-Cheng Tai, Erlend Hodneland, Joachim Weickert, Nickolay V. Bukoreshtliev, Arvid Lundervold, and Hans-Hermann Gerdes. Level Set Methods for Watershed Image Segmentation. In Fiorella Sgallari, Almerico Murli, and Nikos Paragios, editors, *Scale Space and Variational Methods in Computer Vision*, number 4485 in Lecture Notes in Computer Science, pages 178–190. Springer Berlin Heidelberg, May 2007. ISBN 978-3-540-72822-1 978-3-540-72823-8. URL http://link.springer.com/chapter/10.1007/978-3-540-72823-8_16.

Douglas L. Theobald and Deborah S. Wuttke. Empirical Bayes hierarchical models for regularizing maximum likelihood estimation in the matrix Gaussian Procrustes problem. *Proceedings of the National Academy of Sciences of the United States of America*, 103(49):18521–18527, December 2006. ISSN 0027-8424. doi: 10.1073/pnas.0508445103.

D'Arcy Wentworth Thompson. *On Growth and Form*. Cambridge University Press, Cambridge, abridged edition, July 2014. ISBN 978-1-107-67256-7.

E.C. Titchmarsh. *The Theory of Functions*. Oxford University Press, 2nd edition, 1939.

Arkadiusz Tomczyk and Piotr S. Szczepaniak. On the Relationship Between Active Contours and Contextual Classification. In Marek Kurzyński, Edward Puchała, Michał Woźniak, and Andrzej Żołnierek, editors, *Computer Recognition Systems*, number 30 in Advances in Soft Computing, pages 303–310. Springer Berlin Heidelberg, 2005. URL http://link.springer.com/chapter/10.1007/3-540-32390-2_35.

Xin Tong, P. Ozturk, and Mingyang Gu. Dynamic feature weighting in nearest neighbor classifiers. In *Proceedings of 2004 International Conference on Machine Learning and Cybernetics (IEEE Cat. No.04EX826)*, volume 4, pages 2406–2411, August 2004. doi: 10.1109/ICMLC.2004.1382206.

Fernando De la Torre and Michael J. Black. A Framework for Robust Subspace Learning. *International Journal of Computer Vision*, 54(1-3):117–142, August 2003. ISSN 0920-5691, 1573-1405. doi: 10.1023/A:1023709501986. URL https://link.springer.com/article/10.1023/A:1023709501986.

V. Torre and T. A. Poggio. On Edge Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-8(2):147–163, March 1986. ISSN 0162-8828. doi: 10.1109/TPAMI.1986.4767769.

Alain Trouvé. Diffeomorphisms Groups and Pattern Matching in Image Analysis. *International Journal of Computer Vision*, 28(3):213–221, July 1998. ISSN 0920-5691, 1573-1405. doi: 10.1023/A:1008001603737. URL https://link.springer.com/article/10.1023/A:1008001603737.

Loring W. Tu. *An Introduction to Manifolds*. Springer Science & Business Media, October 2010. ISBN 978-1-4419-7399-3.

John W. Tukey. Bias and confidence in not-quite large samples (Abstract). In *The Annals of Mathematical Statistics*, volume 29, page 614, 1958. doi: 10.1214/aoms/1177706647.

Pavan K Turaga and Anuj Srivastava. *Riemannian Computing in Computer Vision*. Springer International Publishing, Switzerland, 2016. ISBN 978-3-319-22957-7. OCLC: 932168505.

W. T. Tutte. Convex Representations of Graphs. *Proceedings of the London Mathematical Society*, s3-10 (1):304–320, January 1960. ISSN 1460-244X. doi: 10.1112/plms/s3-10.1.304.

W. T. Tutte. How to Draw a Graph. *Proceedings of the London Mathematical Society*, s3-13(1):743–767, January 1963. ISSN 1460-244X. doi: 10.1112/plms/s3-13.1.743.

L. Valiant. The Complexity of Enumeration and Reliability Problems. *SIAM Journal on Computing*, 8(3): 410–421, August 1979. ISSN 0097-5397. doi: 10.1137/0208032. URL http://epubs.siam.org/doi/abs/10.1137/0208032.

B. van Ginneken, A. F. Frangi, J. J. Staal, B. M. ter Haar Romeny, and M. A. Viergever. Active shape model segmentation with optimal features. *IEEE Transactions on Medical Imaging*, 21(8):924–933, August 2002. ISSN 0278-0062. doi: 10.1109/TMI.2002.803121.

Gitte Vanwinckelen and Hendrik Blockeel. On estimating model accuracy with repeated cross-validation. In *BeneLearn 2012: Proceedings of the 21st Belgian-Dutch Conference on Machine Learning*, pages 39–44, Ghent, 2012. ISBN 978-94-6197-044-2. URL https://lirias.kuleuven.be/handle/123456789/346385.

Andrea Vedaldi and Stefano Soatto. Quick Shift and Kernel Methods for Mode Seeking. In *Computer Vision – ECCV 2008*, Lecture Notes in Computer Science, pages 705–718. Springer, Berlin, Heidelberg, October 2008. ISBN 978-3-540-88692-1 978-3-540-88693-8. doi: 10.1007/978-3-540-88693-8_52. URL https://link.springer.com/chapter/10.1007/978-3-540-88693-8_52.

Olga Veksler, Yuri Boykov, and Paria Mehrani. Superpixels and Supervoxels in an Energy Optimization Framework. In *Computer Vision – ECCV 2010*, Lecture Notes in Computer Science, pages 211–224.

Springer, Berlin, Heidelberg, September 2010. ISBN 978-3-642-15554-3 978-3-642-15555-0. doi: 10.1007/978-3-642-15555-0_16. URL https://link.springer.com/chapter/10.1007/978-3-642-15555-0_16.

P. W. Verbeek and B. J. H. Verwer. Shading from shape, the eikonal equation solved by grey-weighted distance transform. *Pattern Recognition Letters*, 11(10), October 1990. ISSN 0167-8655. doi: 10.1016/0167-8655(90)90102-8. URL http://www.sciencedirect.com/science/article/pii/0167865590901028.

R. Vidal, Yi Ma, and S. Sastry. Generalized principal component analysis (GPCA). *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(12):1945–1959, December 2005. ISSN 0162-8828. doi: 10.1109/TPAMI.2005.244.

L. Vincent and P. Soille. Watersheds in digital spaces: an efficient algorithm based on immersion simulations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(6):583–598, June 1991. ISSN 0162-8828. doi: 10.1109/34.87344.

Petr Švestka and Mark H. Overmars. Coordinated path planning for multiple robots. *Robotics and Autonomous Systems*, 23(3):125–152, April 1998. ISSN 0921-8890. doi: 10.1016/S0921-8890(97)00033-X.

Richard S. Wallace, Ping-Wen Ong, Benjamin B. Bederson, and Eric L. Schwartz. Space variant image processing. *International Journal of Computer Vision*, 13(1):71–90, September 1994. ISSN 0920-5691, 1573-1405. doi: 10.1007/BF01420796. URL https://link.springer.com/article/10.1007/BF01420796.

Bang Wang, Kee Chaing Chua, Wei Wang, and Vikram Srinivasan. Worst and Best Information Exposure Paths in Wireless Sensor Networks. In Xiaohua Jia, Jie Wu, and Yanxiang He, editors, *Mobile Adhoc and Sensor Networks: First International Conference, MSN 2005, Wuhan, China, December 13-15, 2005. Proceedings*, number 3794 in Lecture Notes in Computer Science, pages 52–62. Springer Berlin Heidelberg, 2005. ISBN 978-3-540-32276-4. doi: 10.1007/11599463_6.

S. Wang and J. M. Siskind. Image segmentation with ratio cut. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(6):675–690, June 2003. ISSN 0162-8828. doi: 10.1109/TPAMI.2003.1201819.

J. S. Weszka and A. Rosenfeld. Threshold Evaluation Techniques. *IEEE Transactions on Systems, Man, and Cybernetics*, 8(8):622–629, August 1978. ISSN 0018-9472. doi: 10.1109/TSMC.1978.4310038.

Robin J. Wilson. *Introduction to Graph Theory*. Addison-Wesley Longman Publishing Co., Inc., 4th edition, 1996. ISBN 0-582-24993-7.

Daniel J. Withey and Zoltan J. Koles. A review of medical image segmentation: Methods and available software. *International Journal of Bioelectromagnetism*, 10(3):125–148, 2008.

Wadim Wojciechowski, Adrian Molka, and Zbisław Tabor. Automated measurement of parameters related to the deformities of lower limbs based on x-rays images. *Computers in Biology and Medicine*, 70:1–11, March 2016. ISSN 1879-0534. doi: 10.1016/j.compbiomed.2015.12.027.

Weiyi Xie. *A Geometric Approach to Visualization of Variability in Univariate and Multivariate Functional Data*. PhD Thesis, The Ohio State University, 2017.

Weiyi Xie, Sebastian Kurtek, Karthik Bharath, and Ying Sun. A Geometric Approach to Visualization of Variability in Functional Data. *Journal of the American Statistical Association*, 112(519):979–993, July 2017. ISSN 0162-1459. doi: 10.1080/01621459.2016.1256813.

C. Yang, L. Zhang, H. Lu, X. Ruan, and M. H. Yang. Saliency Detection via Graph-Based Manifold Ranking. In *2013 IEEE Conference on Computer Vision and Pattern Recognition*, pages 3166–3173, June 2013. doi: 10.1109/CVPR.2013.407.

Le Yang. Riemannian Median and Its Estimation. *LMS Journal of Computation and Mathematics*, 13:461–479, December 2010. ISSN 1461-1570. doi: 10.1112/S1461157020090531. URL `http://arxiv.org/abs/0911.3474`.

Y. Yang, S. Hallman, D. Ramanan, and C. C. Fowlkes. Layered Object Models for Image Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(9):1731–1743, September 2012. ISSN 0162-8828. doi: 10.1109/TPAMI.2011.208.

Yaling Yang and Jun Wang. Design guidelines for routing metrics in multihop wireless networks. In *Proceedings of IEEE Annual Conference on Computer Communications (INFOCOM)*, pages 1615–1623, Phoenix, AZ, USA, April 2008. IEEE.

Zhigang Yao and Tung Pham. Principal Sub-manifolds. *arXiv:1604.04318 [stat]*, April 2016. URL `http://arxiv.org/abs/1604.04318`.

L. Younes. Computable Elastic Distances Between Shapes. *SIAM Journal on Applied Mathematics*, 58(2):565–586, April 1998. ISSN 0036-1399. doi: 10.1137/S0036139995287685.

Laurent Younes. Optimal matching between shapes via elastic deformations. *Image and Vision Computing*, 17(5–6):381–389, April 1999. ISSN 0262-8856. doi: 10.1016/S0262-8856(98)00125-5. URL `http://www.sciencedirect.com/science/article/pii/S0262885698001255`.

Laurent Younes. Spaces and manifolds of shapes in computer vision: An overview. *Image and Vision Computing*, 30(6–7):389–397, June 2012. ISSN 0262-8856. doi: 10.1016/j.imavis.2011.09.009.

Hsiang-Fu Yu, Fang-Lan Huang, and Chih-Jen Lin. Dual coordinate descent methods for logistic regression and maximum entropy models. *Machine Learning*, 85(1-2):41–75, October 2011. ISSN 0885-6125, 1573-0565. doi: 10.1007/s10994-010-5221-8.

C. T. Zahn. Graph-Theoretical Methods for Detecting and Describing Gestalt Clusters. *IEEE Transactions on Computers*, C-20(1):68–86, January 1971. ISSN 0018-9340. doi: 10.1109/T-C.1971.223083.

Nida M. Zaitoun and Musbah J. Aqel. Survey on Image Segmentation Techniques. *Procedia Computer Science*, 65:797–806, January 2015. ISSN 1877-0509. doi: 10.1016/j.procs.2015.09.027.

Wei Zeng, Yi-Jun Yang, and Muhammad Razib. Graph-Constrained Surface Registration Based on Tutte Embedding. In *2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 516–523, Las Vegas, NV, USA, June 2016. IEEE.

Hui Zhang, Jason E. Fritts, and Sally A. Goldman. Image segmentation evaluation: A survey of unsupervised methods. *Computer Vision and Image Understanding*, 110(2):260–280, May 2008. ISSN 1077-3142. doi: 10.1016/j.cviu.2007.08.003. URL `http://www.sciencedirect.com/science/article/pii/S1077314207001294`.

L. Zhang and Q. Ji. A Bayesian Network Model for Automatic and Interactive Image Segmentation. *IEEE Transactions on Image Processing*, 20(9):2582–2593, September 2011. ISSN 1057-7149. doi: 10.1109/TIP.2011.2121080.

Miaomiao Zhang and P. Thomas Fletcher. Probabilistic Principal Geodesic Analysis. In *Proceedings of the 26th International Conference on Neural Information Processing Systems*, NIPS'13, pages 1178–1186, USA, 2013. Curran Associates Inc.

Miaomiao Zhang and Polina Golland. Statistical shape analysis: From landmarks to diffeomorphisms. *Medical Image Analysis*, 33:155–158, October 2016. ISSN 1361-8415. doi: 10.1016/j.media.2016.06.025.

Y. J. Zhang. A survey on evaluation methods for image segmentation. *Pattern Recognition*, 29(8):1335–1346, August 1996. ISSN 0031-3203. doi: 10.1016/0031-3203(95)00169-7.

Márton Zubor, Attila Kőrösi, András Gulyás, and Gábor Rétvári. On the Computational Complexity of Policy Routing. In *Advances in Communication Networking*, volume 8846 of *Lecture Notes in Computer Science*, pages 202–214. Springer, Cham, September 2014. doi: 10.1007/978-3-319-13488-8_19.