

Using Atom-Like Local Image Features to Study Human Genetics and Neuroanatomy in Large Sets of 3D Medical Image Volumes

by

Laurent CHAUVIN

MANUSCRIPT-BASED THESIS PRESENTED TO ÉCOLE DE
TECHNOLOGIE SUPÉRIEURE
IN PARTIAL FULFILLMENT FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
Ph.D.

MONTREAL, JULY 26, 2022

ÉCOLE DE TECHNOLOGIE SUPÉRIEURE
UNIVERSITÉ DU QUÉBEC



Laurent Chauvin, 2022



This Creative Commons license allows readers to download this work and share it with others as long as the author is credited. The content of this work cannot be modified in any way or used commercially.

BOARD OF EXAMINERS

**THIS THESIS HAS BEEN EVALUATED
BY THE FOLLOWING BOARD OF EXAMINERS**

M. Matthew Toews, Thesis supervisor
Ecole de Technologie Supérieure, Département de Génie des Systèmes

M. Jacques de Guise, co-Supervisor
Ecole de Technologie Supérieure, Département de Génie des Systèmes

M. Rafael Menelau Oliveira Cruz, President of the board of examiners
Ecole de Technologie Supérieure, Département de Génie Logiciel

M. Ismail Ben Ayed, Member of the jury
Ecole de Technologie Supérieure, Département de Génie des Systèmes

M. Jean-Baptiste Poline, External examiner
McGill University

**THIS THESIS WAS PRESENTED AND DEFENDED
IN THE PRESENCE OF A BOARD OF EXAMINERS AND THE PUBLIC
ON JULY 14, 2022
AT ÉCOLE DE TECHNOLOGIE SUPÉRIEURE**

FOREWORD

This thesis was the result of Laurent Chauvin's career path from the Surgical Planning Laboratory of the Harvard Medical School, where he assisted in a wide range of cutting edge research in medical robotics and image acquisition protocols related to magnetic resonance and ultrasound scanning projects of the human body. In this thesis, he has ushered in an era in 2022 where we increasingly view objects in the 3D world and universe in terms of volumetric 3D images, a virtual reality or "metaverse", in which 3D structures such as the human body are naturally described by elementary particles defined by the most fundamental of mathematical properties, including the notions of atom-like particles within 3D cartesian space defined by three spatial variables $(x, y, z) \in R^3$ and a scale σ defined by isotropic Einstein diffusion. A variety of arbitrary deep multi-channel neural network architectures will emerge to classify limited sets of labeled training data, however there will only ever be one scale-space (x, y, z, σ) defined by the most elementary of differential operators ∇ and ∇^2 used by the likes of Newton, Leibnitz, Laplace, Gauss, Fourier, Fresnel, Huygens, Hamilton, Maxwell, Planck, Einstein, Schrodinger and Dirac from a single scalar field $I(x, y, z, \sigma) \in R^1$. Laurent Chauvin has expanded human knowledge in this theoretical domain by introducing discrete sign-parity invariance (SP-invariance) into the analysis of the human brain, equivalent to CP-invariance in the transition of elementary charged particles to antiparticles such as the proton to the antiproton, and in numerous discoveries regarding the nature of the human brain, of the potential human errors and the promise for equitable, unbiased healthcare of human subjects in large medical imaging datasets from fundamental image particles.

- Matthew Toews

*«You learn the most from things
you enjoy doing so much that
you don't even notice time passing»*

- Albert Einstein

ACKNOWLEDGEMENTS

First, I want to express my most sincere gratitude to my supervisor Pr. Matthew Toews for his incredible pedagogy, his unwavering support and more importantly for sharing his infinite motivation, passion and wisdom with me. Thank you so much, it was such a great pleasure to work and learn from you. I couldn't have hoped for a better mentor.

I also want to express my gratitude to my lab mates and professors at ETS: Etienne, Mohsen, Herve, Christian, Kuldeep, Karthik, Jerome and Jose.

I also want to thank my friends for always being there, not only in good times (D&D, Fiasco, Camping trips, etc...), but also, and especially in the bad ones to cheer me up. You have no idea how much that meant to me, so thank you Alex (my shooting star), Marmich, Nick, Laure, Romain, Clarisse, Gael, Joe, AnneMa, Eve, Baptiste, Shirley, Sam, Joel, Margaux, Valentin, Boris, Ted, Caroline, Lisa, Matthieu, Marine, David, Juliette, Florian, Sophie, Bastien, Celine, Pierre, Asli, Kristin, Sean, Alexis, Frank, Pierre-Yves, Thomas, Clement, Junichi, Sonia, Hatsuo, Mathieu, Fabien, Alan, Mik, Tonio, Tete, Coro, Vincent, Maxime. Please forgive me if I forgot some of you, but I am deeply thankful to each and every one of you.

A very very special thought to the Crou: Marion, Aude, Yahya, and Capucine, who have been my four pillars during this PhD. Your friendship means more to me than I can express and I'm so proud and grateful to have you in my life. I also want to thank Annick for her unconditional support, her patience, her faith in me, and whose laughs, courage and positivity inspired me.

I also want to deeply thank my parents and my brother, who always believed in me. None of this would have been possible without their trust and support during my whole life. Thank you for everything you have done for me all these years. You are a source of inspiration and pride to me.

A thought to Maya and Flora, my goddaughter and my niece. This work has been done in the hope of building a brighter future for you.

Finally, I want to dedicate my thesis to my grand-parents, that I always kept in my mind during this PhD, who gave me the strength and courage to never give up. I can only hope they are proud of what I accomplished thanks to them.

Utilisation de caractéristiques d'image inspirée par le modèle atomique pour étudier la génétique et la neuroanatomie humaines dans de grands ensembles d'images médicales 3D

Laurent CHAUVIN

RÉSUMÉ

Les contributions de cette thèse découlent de la technologie développée pour analyser de grands ensembles d'images volumétriques en termes de caractéristiques locale, inspirée du modèle atomique, extraites dans l'espace d'image 3D, suivant la célèbre transformation de caractéristique invariante d'échelle (SIFT) (Lowe, 2004) dans l'espace d'image 2D. De nouvelles propriétés de caractéristiques sont introduites, notamment un signe de caractéristique binaire $s \in \{-1, +1\}$, analogue à une charge électrique, et un ensemble discret d'états d'orientation de caractéristiques symétriques dans l'espace 3D. Ces nouvelles propriétés sont exploitées pour étendre l'invariance des caractéristiques en incluant la transformation d'inversion de signe et de parité (SP), analogue à la transformation de conjugaison de charge et de parité (CP) entre une particule et son antiparticule en mécanique quantique, ce qui permet de prendre en compte l'inversion du contraste d'intensité locale entre différentes modalités d'imagerie et les réflexions d'axe dues à la symétrie de forme. Un nouveau noyau exponentiel $\mathcal{K}(f_i, f_j) \in [0, 1]$ est proposé pour quantifier la similarité d'une paire de caractéristiques f_i et f_j extraites dans différentes images à partir de leurs propriétés, notamment la localisation, l'échelle, l'orientation, le signe et l'apparence. Ce noyau peut être calculé à l'aide d'algorithmes rapides d'indexation des caractéristiques du plus proche voisin, réduisant la complexité de la recherche d'un ensemble de N caractéristiques de $O(N)$ à $O(\log N)$, permettant à des algorithmes d'analyse d'images très efficaces de s'adapter à de larges ensembles de données. Une nouvelle mesure intitulée le Jaccard flou $J(A, B)$ est proposée pour quantifier la similarité d'une paire d'ensembles de caractéristiques $A = \{f_i\}$ et $B = \{f_j\}$ sur la base de leur chevauchement ou intersection-sur-union $|A \cap B|/|A \cup B|$, où le noyau $\mathcal{K}(f_i, f_j)$ établit une équivalence non binaire ou douce entre une paire d'éléments de caractéristiques (f_i, f_j) . Le Jaccard flou peut être utilisé pour identifier des paires d'ensembles de caractéristiques extraites de même individus ou de même familles avec une grande précision, et un simple seuil de distance a permis de découvrir de manière surprenante des erreurs d'étiquetage d'individus et de familles auparavant inconnues dans d'importants ensembles de données publiques de neuroimagerie. La distance floue de Jaccard peut également être évaluée entre un ensemble de caractéristiques et un groupe d'ensembles de caractéristiques afin de réaliser une classification en fonction des étiquettes de données par groupe. Un nouvel algorithme est proposé pour recalibrer ou aligner spatialement une paire d'ensembles de caractéristiques A et B , intitulé SIFT Coherent Point Drift (SIFT-CPD), en identifiant une transformation $T : B \rightarrow A$ qui maximise le Jaccard flou $J(A, T \circ B)$ entre un ensemble de caractéristiques fixe A et un ensemble transformé $T \circ B$. L'algorithme SIFT-CPD permet d'obtenir un recalage plus rapide et plus précis que l'algorithme CPD original basé uniquement sur les informations de localisation des caractéristiques, dans une variété de contextes difficiles, notamment la tomographie thoracique, l'IRM cérébrale, les modalités multiples, la

variabilité naturelle entre les sujets et les pathologies telles que les tumeurs et les maladies neurodégénératives.

Mots-clés: IRM, Image Médicale Volumétrique, Point-clef Invariant d'Echelle, Large Jeux de Données, Invariance, Symétrie Discrète, ADN, Neuroanatomie, Empreinte Cérébrale

Using Atom-Like Local Image Features to Study Human Genetics and Neuroanatomy in Large Sets of 3D Medical Image Volumes

Laurent CHAUVIN

ABSTRACT

The contributions of this thesis stem from technology developed to analyse large sets of volumetric images in terms of atom-like features extracted in 3D image space, following the well-known Scale Invariant Feature Transform (SIFT) (Lowe, 2004) in 2D image space. New feature properties are introduced including a binary feature sign $s \in \{-1, +1\}$, analogous to an electrical charge, and a discrete set of symmetric feature orientation states in 3D space. These new properties are leveraged to extend feature invariance to include the sign inversion and parity (SP) transform, analogous to the charge conjugation and parity (CP) transform between a particle and its antiparticle in quantum mechanics, thereby accounting for local intensity contrast inversion between imaging modalities and axis reflections due to shape symmetry. A novel exponential kernel $\mathcal{K}(f_i, f_j) \in [0, 1]$ is proposed to quantify the similarity of a pair of features f_i and f_j extracted in different images from their properties including location, scale, orientation, sign and appearance. This kernel may be computed using fast approximate k-nearest neighbor feature indexing algorithms that reduce the computational complexity of searching a set of N features from $O(N)$ to $O(\log N)$, allowing highly efficient image analysis algorithms to scale to arbitrarily large datasets. A novel measure entitled the soft Jaccard $J(A, B)$ is proposed to quantify the similarity of a pair of feature sets $A = \{f_i\}$ and $B = \{f_j\}$ based on their overlap or intersection-over-union $|A \cap B|/|A \cup B|$, where the kernel $\mathcal{K}(f_i, f_j)$ establishes non-binary or soft equivalence between a pair of feature elements (f_i, f_j) . The soft Jaccard may be used to identify pairs of feature sets extracted from the same individuals or families with high accuracy, and a simple distance threshold led to the surprising discovery of previously unknown individual and family labeling errors in major public neuroimage datasets. The soft Jaccard may also be evaluated between a feature set and a group of feature sets in order to achieve classification according to group-wise data labels. A new algorithm is proposed to register or spatially align a pair of feature sets A and B , entitled SIFT Coherent Point Drift (SIFT-CPD), by identifying a transform $T : B \rightarrow A$ that maximizes the soft Jaccard $J(A, T \circ B)$ between a fixed feature set A and a transformed set $T \circ B$. SIFT-CPD achieves faster and more accurate registration than the original CPD algorithm based on feature location information alone, in a variety of challenging contexts including chest CT, brain MRI, multiple modalities, natural inter-subject variability and pathologies such as tumors and neuro-degenerative disease.

Keywords: MRI, Volumetric Medical Images, Scale-invariant feature, Keypoints, Big data, Invariance, Discrete Symmetry, DNA, Neuroanatomy, Brain fingerprinting

TABLE OF CONTENTS

	Page
INTRODUCTION	1
0.1 Problem statement and motivation	2
0.2 Research objectives and contributions	4
0.3 Publications	5
0.4 Thesis outline	7
CHAPTER 1 LITERATURE REVIEW	9
CHAPTER 2 REGISTERING IMAGE VOLUMES USING 3D SIFT AND DISCRETE SP-SYMMETRY	17
2.1 Introduction	18
2.2 Related work	20
2.2.1 Local Image Features	21
2.2.2 Image Registration	22
2.2.3 Discrete Symmetry	24
2.3 Method	26
2.3.1 Single-Feature Properties	26
2.3.2 Two-Feature Observations	29
2.3.3 Multi-Feature Registration	31
2.4 Experiments	34
2.4.1 Synthetic Image Registration	35
2.4.2 Inter-subject Registration	35
2.4.2.1 T1w Brain Images	36
2.4.2.2 CT Chest Images	37
2.4.2.3 Abnormal Variations and Occlusion	38
2.4.2.4 Multi-Modal Images	39
2.4.3 Feature Sign and Orientation State	41
2.5 Discussion	41
CHAPTER 3 NEUROIMAGE SIGNATURE FROM SALIENT KEYPOINTS IS HIGHLY SPECIFIC TO INDIVIDUALS AND SHARED BY CLOSE RELATIVES	45
3.1 Introduction	46
3.2 Material & Methods	49
3.2.1 Data	49
3.2.2 Processing	50
3.3 Results	55
3.4 Discussion	60
3.4.1 Code and Data Availability	64
3.5 Acknowledgements	64

CHAPTER 4	EFFICIENT PAIRWISE NEUROIMAGE ANALYSIS USING THE SOFT JACCARD INDEX AND 3D KEYPOINT SETS	67
4.1	Introduction	68
4.2	Related Work	70
4.3	Method	73
4.4	Experiments	78
4.4.1	Data and Computational Details	80
4.4.2	Close genetic proximity: Siblings	81
4.4.3	Distant genetic proximity: unrelated subjects	83
4.4.4	Group Prediction: Sex	86
4.5	Discussion	88
CONCLUSION AND RECOMMENDATIONS		91
APPENDIX I	DIFFUSION ORIENTATIONS HISTOGRAMS (DOH) FOR DIFFUSION WEIGHTED IMAGE ANALYSIS	93
1	Introduction	94
2	Related Work	95
3	Diffusion Orientation Histogram Descriptors	96
4	Experiments	98
5	Discussion	100
APPENDIX II LAPLACIAN-OF-GAUSSIAN PARTICLE THEORY		103
APPENDIX III CPD REGISTRATION DETAILS		105
BIBLIOGRAPHY		106

LIST OF TABLES

	Page
Table 2.1	Comparison of five registration algorithms based on 100 synthetic transforms of a single T1w MRI brain volume. Error is listed terms of individual rotation and translation axes, overall point registration error (PRE), and the sum of squared intensity differences (SSD) following registration 36
Table 2.2	Inter-subject registration experimental data including the numbers of features per image and Hough transform inlier matches 36
Table 3.1	Dataset demographic and statistical information 49
Table 3.2	p-values for two-sample Kolmogorov-Smirnov tests between Jaccard distance distributions for (SM) Same Subject, (MZ) Monozygotic, (DZ) Dizygotic, (FS) Full-Sibling and (UR) Unrelated pairwise relationships 57
Table 3.3	Misabeled subject relationships identified across and within databases. Most cases are subjects mislabeled as UR * A pair labeled as MZ was established to be SM and one dataset was subsequently removed from the HCP dataset, and a pair of scans labeled as UR were confirmed to be SM from the OASIS 1 dataset ** OASIS 3 and OASIS 1 are known to share data from numerous subjects *** OASIS 3 contained 2 pairs of SM subjects mislabeled as UR, and 3 pairs of UR subjects labeled as SM 61
Table 4.1	Area Under the Curve (AUC) values for sibling classification, comparing kernels and numbers of nearest neighbors (k). $\epsilon = 0.0017$ is the maximum AUC standard deviation observed from 21 different subject-to-atlas registration trials. Note $SSE_{App+Geo}$ kernels significantly outperform SSE_{App} and HSE for the more difficult DZ and FS pairs 83

LIST OF FIGURES

		Page
Figure 0.1	Illustrating atom-like image features extracted in a 3D magnetic resonance image (MRI) of the human head. Properties are shown for an individual feature, including location $\bar{\mathbf{x}}$, scale σ , orientation axes Θ and binary sign $s = \{-, +\}$	1
Figure 2.1	Illustrating the geometrical properties of 3D features (colored spheres) extracted from an MRI image of the human brain. Properties include location $\mathbf{x}$, orientation Θ , scale σ , we introduce a binary feature sign $s = \{-, +\}$ based on the sign of the Laplacian-of-Gaussian	19
Figure 2.2	Illustrating an appearance descriptor $a(r, \phi)$ and four discrete orientation states given principal axes $\hat{\theta}_1, \hat{\theta}_2$ (left to right) and feature sign s (upper and lower). The descriptor content is shown for an example octant (r_{+++}) and gradient orientation (ϕ_{++-}). Axis orientation is defined independently of sign (left to right), and negative sign $s = -1$ inverts the descriptor gradient (lower row) achieving invariance to contrast inversion	29
Figure 2.3	Inter-subject registration performance for healthy brain T1w MRI	37
Figure 2.4	Inter-subject registration performance for chest CT images	38
Figure 2.5	Inter-subject registration performance for brains in T1w MRI, with occlusions due to tumors (Menze et al., 2015)	39
Figure 2.6	Inter-subject registration performance for healthy brain images between T1w and T2w MRI modalities (Van Essen et al., 2012), exhibiting intensity contrast inversion	40
Figure 2.7	Examples of features matched between T1w and T2w images, showing feature signs, note the sign changes do not occur everywhere	40
Figure 2.8	Orientation state transition histograms for principal gradient orientation estimation (see (Rister et al., 2017)) and max gradient orientation estimation (see (Toews et al., 2013a)) for: a) T1w brain, b) T1w brain tumor (Menze et al., 2015), c) CT chest (Regan et al., 2011), d) Multi-modal T1w/T2w brain. States described in Figure 4.3, are combinations of discrete axis reflections $\{\pm\hat{\theta}_1, \pm\hat{\theta}_2\}$ as 0: $\{++\}$, 1: $\{+-\}$, 2: $\{-+\}$, 3: $\{--\}$. Bins along the	

	diagonal indicate no state change, off-diagonals indicate single-axis reflections and inverse diagonal indicates parity transform	42
Figure 2.9	Examples of feature orientation state transitions observed between T1w and T2w modalities. Circles indicate feature location $\mathbf{x}$ and scale σ in the plane, sign $s \in \{-1, +1\}$ is indicated by white-black or black-white color. Red and blue arrows indicate the primary $\hat{\theta}_1$ and secondary $\hat{\theta}_2$ orientation axes projected onto the (x,y) image plane	43
Figure 3.1	The workflow for computing the Jaccard overlap $J(A, B)$ similarity score between two images A and B. Step I. SIFT-Rank keypoints are extracted from skull-stripped MRI data. Step II. Similar keypoints are identified between images using a K-nearest neighbor search. Step III. The Jaccard overlap is computed as ratio of the intersection vs. the union of keypoint sets	48
Figure 3.2	Distributions of the pairwise Jaccard distances conditional on relationship labels including: unrelated (UR, blue), same (SM, red), full siblings (FS, purple), dizygotic (DZ, green) and monozygotic (MZ, yellow) twin subject relationship labels, where black dashed lines indicate distribution means. Note the high degree of separation between SM and UR scores. Dots indicate data labeling inconsistencies automatically flagged by unexpected Jaccard distance. The correct relationship labels are evident upon visual inspection of cortical patterns (highlight), which are virtually identical in SM pairs mislabeled as (a) UR and (b) MZ, and highly different in a UR pair mislabelled as (c) SM	57
Figure 3.3	The spatial layout of keypoint matches for five pairwise relationship labels: a) SM, b) MZ, c) DZ, d) FS, e) UR. Heatmaps represent distributions of matching keypoints within the standard MNI305 neuroanatomic reference space and accumulated over 71 image pairs per label from HCP dataset	58
Figure 3.4	Keypoints matched between MRIs of the same individual acquired at 51 and 62 years of age. Matching keypoints (spheres, color indicating scale) represent patterns of local cortical structure that are highly unique to this individual, with notable concentrations in a) Broca's area, b) the primary somatosensory cortex and c) Wernicke's area. Note that keypoints have been slightly extruded from within the cortex for improved visualization. The visualization was generated using the 3D Slicer software (Fedorov et al., 2012)	58

Figure 3.5	Jaccard distance as a function of the age difference Δt between scans for SM (N=9885) and FS (N=607) pairs. Lines and boxes represent the means and standard deviations of the Jaccard distance. a) and b) show examples of image pairs successfully identified as SM, despite a relatively high Jaccard distance due to aging, noticeable image artefacts in a) and neurodegenerative atrophy in b) 59
Figure 4.1	Illustrating two sets of 3D keypoints $A = \{f_i\}$ and $B = \{f_j\}$ extracted from MRI brain volumes of monozygotic twins. Colored circles represent the 3D locations $\bar{x} \in R^3$ and scales (radii) $\sigma \in R^1$ of keypoints in 2D axial slices, and links are correspondences identified via the SIFT-Rank (Toews & Wells, 2009) algorithm, note the degree of similarity between keypoint pairs 69
Figure 4.2	Graph of spatial localization error $\ \bar{x}_i - \bar{x}_j\ _2$ vs. geometric mean scale $\sqrt{\sigma_i \sigma_j}$. Lines represent the linear regression based on each point set. Note how location error increases with scale (black) due to uncertainty, however this effect is reduced by scale normalization (green) as in Equation (4.9). This visualization is based on keypoint correspondences between 10 MRI aligned volumes 78
Figure 4.3	Visualizing the keypoint correspondences driving classification between a pair of identical twin brains, showing the 100 correspondences with the highest SSE kernel values. The appearance kernel alone (upper, Equation (4.6)) may include geometrically inconsistent matches (diagonal lines), however these are down weighted when combined with the geometry kernel (lower, Equations (4.9) and (4.10)), leading to geometrically consistent matches (horizontal lines) and improved prediction performance 79
Figure 4.4	Distributions of Jaccard distances $-\log J_{SSE}(A, B)$ conditioned on pairwise labels $L = \{MZ, DZ, FS, HS, UR\}$, comparing kernels with and without geometry (green vs. orange). Note that the distance generally increases with genetic separation, and that geometry kernels increase separation between sibling relationship labels. Outliers, i.e. pairs outside of their expected relationship distance distributions, can be easily identified and inspected for irregularity, e.g. (a) unusually high distances may indicate spatial misalignment (red circles) and (b) unusually low UR distances may indicate cases of incorrect family labels (blue circles) 82
Figure 4.5	ROC curves for sibling identification based on Jaccard distance for (a) MZ, (b) DZ, (c) FS and (d) HS pairs, comparing three kernels for k=200 NNs. Note that the prediction AUC is always highest for our

	proposed kernel combining appearance and geometry (green), and decreases as expected with the degree genetic and developmental difference, e.g. in the order MZ, DZ, FS, HS	84
Figure 4.6	Jaccard distance – $\log J_{SSE}(A, B)$ distributions between unrelated pairs conditioned on shared demographic information. (a) same race, same sex (R, S); different race, same sex ($\bar{R}, S$); same race, different sex ($R, \bar{S}$), and different race, different sex ($\bar{R}, \bar{S}$) and (b) Female-Female (F, F), Male-Male (M, M), Male-Female (M, F). Image pairs corresponding to minimum and maximum distances for each distribution are shown for visualization	85
Figure 4.7	Distributions of Jaccard distance – $\log J_{SSE}(A, B)$ conditioned on age difference Δt between unrelated pairs (A, B) conditioned on age difference Δt . The mean (solid red line) and standard deviation (dashed red lines) are plotted for each Δt . Age difference has no significant impact on the distance for unrelated healthy young adult brains (age 22-36 years), eliminating a possible confounding factor	86
Figure 4.8	ROC curves and AUC for sex prediction based on individual-to-group Jaccard distance, comparing binary, appearance only, and combined appearance and geometry kernels.	87
Figure 5.1	Illustration of first nearest neighbor graph between features in the scale-space based on the geometrical kernel $\mathcal{K}(g_i, g_j)$. Feature scales are represented by colors, from red for smaller ones to blue for larger ones.	92

LIST OF ALGORITHMS

	Page
Algorithm 2.1 The probabilistic SIFT-CPD registration algorithm adapted from (Myronenko & Song, 2010) to include kernel $\mathcal{K}(g_n, g'_m)$	33

LIST OF ABBREVIATIONS

AUC	Area Under the Curve
CNN	Convolutional Neural Network
CPD	Coherent Point Drift
DOH	Diffusion Orientation Histogram
ICP	Iterative Closest Point
ML	Machine Learning
MRI	Magnetic Resonance Imaging
PACS	Picture Archiving and Communication System
SIFT	Scale-Invariant Feature Transform
VBM	Voxel-based Morphometry
HCP	Human Connectome Project
ADNI	Alzheimer's Disease Neuroimaging Initiative
OASIS	Open Access Series of Imaging Studies
SM	Same Subject
MZ	Monozygotic Twin
DZ	Dizygotic Twin
HS	Half Sibling
FS	Full Sibling
UR	Unrelated Individual

INTRODUCTION

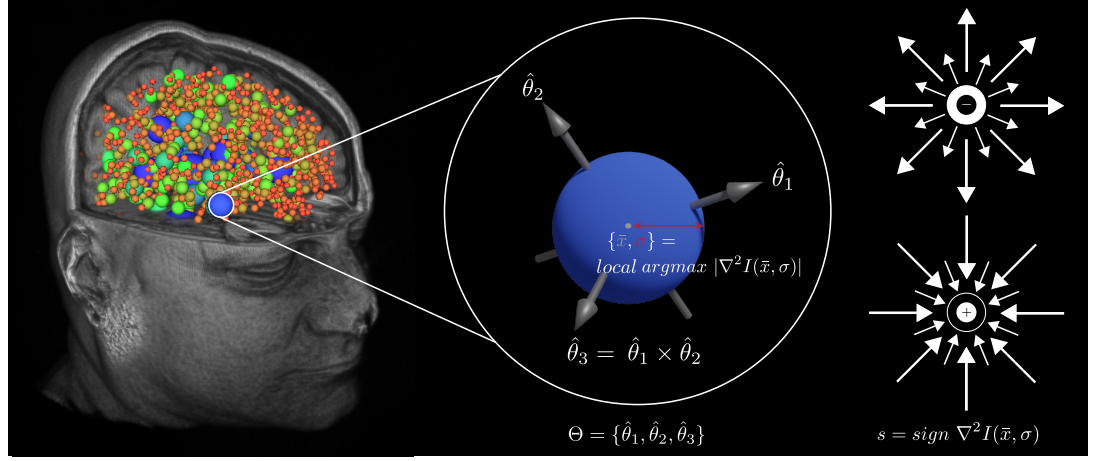


Figure 0.1 Illustrating atom-like image features extracted in a 3D magnetic resonance image (MRI) of the human head. Properties are shown for an individual feature, including location $\bar{x}$, scale σ , orientation axes Θ and binary sign $s = \{-, +\}$

The structure of matter in the 3D world is often described in terms of localized phenomena. In the 5th century BCE, Leucippus of Miletus and his disciple Democritus proposed the term 'atomos' to describe what they thought to be indivisible and fundamental particles of nature. Newton proposed that light might be composed of corpuscular particles ([Newton, 1704](#)). Einstein described the motion of particles as diffusing passively according to random Brownian motion ([Einstein, 1905](#)), equivalent to the heat diffusion equation $\partial u / \partial t \propto \nabla^2 u$. The modern atomic model of quantum chemistry in its most basic form is defined by Schrodinger's wave equation ([Schrödinger, 1926](#)), equivalent to the heat equation and with complex-valued wave solutions permitting momentum, relating a potential field $\psi(x, t)$ over space and time coordinates (x, t) to fundamental operators like the gradient ∇ with respect to time and the Laplacian ∇^2 :

$$i\hbar \frac{\partial \psi(x, t)}{\partial t} = -\frac{\hbar^2}{2m} \nabla^2 \psi(x, t), \quad (1)$$

where m is the mass of a particle and $\hbar$ is the reduced Planck constant.

The contributions of this thesis focuses on keypoint methods, where 3D image volumes such as MRI scans of the human brain are represented and analyzed as sets of highly informative atom-like image features, as shown in Figure 0.1. Features are identified and encoded via fundamental Gaussian, Laplacian and gradient operators (Toews *et al.*, 2010, 2013b; Rister *et al.*, 2017), following the highly successfully scale-invariant feature transform (SIFT) (Lowe, 2004; Lindeberg, 1998), and used in highly efficient and robust imaging applications with no explicit training procedures, including registration (Luo *et al.*, 2018b; Frisken *et al.*, 2019; Machado *et al.*, 2018a), segmentation (Wachinger *et al.*, 2018; Gill *et al.*, 2014) and indexing (Toews *et al.*, 2015).

0.1 Problem statement and motivation

Large medical imaging datasets and machine learning algorithms provide new opportunities to increase scientific understanding of anatomy, function and connectivity from subject groups defined by factors such as genetics, demographics or disease status. However, leveraging information in large sets of volumetric image data is a challenging task, for example, maintaining and curating imaging data for individual human patients in a large, hospital PACS (Picture Archiving and Communication System) database. In the age of personalized medicine promising to tailor healthcare solutions to the unique characteristics of individuals (Hamburg & Collins, 2010), can patients be assured of accurate diagnosis and treatment from their medical imaging data?

Currently, even for the most basic tasks of dataset curation and ensuring data integrity, there are few options for detecting potential errors at the level of individual subjects. For example, images with incorrect subject IDs may lead to bias in analyzing or training from large public datasets such as OASIS (Marcus *et al.*, 2007), ADNI (Jack Jr *et al.*, 2008) or HCP (Van Essen *et al.*, 2013) and potentially lead to catastrophic misdiagnosis of individual patients. Storing and curating a set of N images entails a $O(N^2)$ computational complexity for comparing all

image pairs, and is generally intractable for naive algorithms and large datasets N . Computing pairwise whole-image (dis)similarity is a challenging task in and of itself, computational complexity challenges aside. Standard feature spaces, including atlases or parcellations, tend to simplify computational complexity while masking individual diversity. For example, the dominant computational paradigm and the *de facto* state-of-the-art solution to generic tasks such as classification is the deep neural network architecture (LeCun *et al.*, 1989) coupled with gradient-based backpropagation training on graphics processing units (GPUs) (Krizhevsky *et al.*, 2012). Of the numerous challenges to this paradigm, including training set bias (Torralba & Efros, 2011; Hermann *et al.*, 2020; Buolamwini & Gebru, 2018), length training procedures (Bender *et al.*, 2021), improperly calibrated outputs (Guo *et al.*, 2017), lack of intrinsic invariance (Bardes *et al.*, 2022), susceptibility to spoofing (Nguyen *et al.*, 2015) and attacks (Su *et al.*, 2019), the most relevant is training data sparsity, i.e. generalizing from few examples. In highly sparse contexts such as specific image retrieval, deep machine learning has not significantly improved upon classic methods for keypoint-based matching in 2D computer vision (Balntas *et al.*, 2017; Bellavia & Colombo, 2020). Intuitively, machine learning seeks to generalize from training datasets, however in the process they tend to obscure or bias unique or idiosyncratic information required for fine-grained inferences at the level of individual images.

This thesis adopts an approach particularly well-suited to individual subject indexing. We adopt a memory-based (or instanced-based) model (Cover & Hart, 1967; Brin, 1995; Toussaint, 2002) specifically to index all patient-specific information stored in memory, and thereby preserve patient-specific anatomical features such as cortical folding in brain MRI that may be unique to individuals or shared by close relatives. Each image is encoded as set of generic, salient localized characteristics or features, defined from nothing other than elementary mathematical operators or filters: the Gaussian $G = \{+1, +1\}$ the gradient $\nabla = \{-1, +1\}$ and the Laplacian $\nabla^2 = \{-1, +1\} * \{-1, +1\} = \{-1, +2, -1\}$. These elementary filters are fundamentally symmetric, rotation-invariant and unbiased (i.e. the Gaussian and Laplacian) or may be so via uniform

sampling (i.e. the gradient). These may be used, without training, to identify generic image features and reliably recognised individual-specific structure when present in two or more images.

0.2 Research objectives and contributions

This thesis seeks to advance the theory of fundamental 3D image features by investigating the following research questions: What are the geometrical properties of local features extracted in 3D images, and how do these relate to fundamental properties of physical particles such as atoms, including symmetry, invariance and charge? What can local features tell us about the human brain and the genetic code that defines aspects such as family structure or biological sex?

The primary technological contributions of this thesis stem from characterizing the geometrical properties of 3D local features, in the context of the local invariant feature paradigm, where $O(N^2)$ data pairs may be evaluated in $O(N \log N)$ computational complexity via highly robust and efficient local feature indexing, where inference scales gracefully to arbitrarily large datasets of generic image data with no explicit training procedures. These contributions may be listed as follows:

1. We provide the first image feature description including a binary sign $s = \{-1, +1\}$ and discrete orientation state. This leads to feature descriptors that are invariant discrete symmetry, including charge conjugation and parity transforms (CP-symmetry), currently a topic of high interest in studying the transition of matter to anti-matter ([Cronin, 1981](#); [Pascoli & Turner, 2020](#); [Georgescu, 2020](#)). These are shown in Figure 0.1.
2. We propose a kernel function to quantify the similarity of a pair features observed in different images, based on the variability of their properties including feature geometry and appearance. This leads to higher accuracy in tasks like probabilistic image registration and prediction.

3. We propose a measure of (dis)similarity between a pair of feature sets based on their overlap or intersection-over union, quantified by a soft Jaccard-like distance. This provides a solution to a number of problems relating to massive datasets and personalized medicine, including curation and prediction.

These technical contributions lead to a number of novel experimental discoveries:

1. Individuals and family members including siblings may be rapidly identified via low Jaccard distance between MRI pairs, allowing personalized medicine to flag rare but potentially catastrophic errors in patient labels. Our approach was the first to identify labeling errors in large widely used medical imaging datasets, including images of different people labeled as the same person, images of the same person labeled as different people, images of family members labeled as unrelated due to inconclusive genotyping.
2. The Jaccard distance between brain MRIs reflects the genetic separation between subjects, increasing between pairs of monozygotic twins, siblings and half-siblings, and unrelated individuals.
3. Pairs of individuals of different race and sex may often have higher anatomical brain similarity than individuals of the same race and sex. This indicates that race and sex do not significantly factor in neuroanatomical similarity, just as they do not significantly impact overall genetic similarity in terms of shared single nucleotide polymorphism (SNP).

0.3 Publications

The work presented here resulted in several publications in journal papers and conferences, detailed in the following sections.

Journal Articles:

1. **Laurent Chauvin**, Kuldeep Kumar, Christian Wachinger, Marc Vangel, Jacques de Guise, Christian Desrosiers, William Wells III, Matthew Toews, and Alzheimer’s Disease Neu-

roimaging Initiative. "Neuroimage signature from salient keypoints is highly specific to individuals and shared by close relatives." *NeuroImage* 204: 116208, 2020. - **Accepted** (*Figure selected as cover by the journal*)

2. **Laurent Chauvin**, Kuldeep Kumar, Christian Desrosiers, William Wells III, and Matthew Toews. "Efficient Pairwise Neuroimage Analysis using the Soft Jaccard Index and 3D Keypoint Sets." *IEEE Transactions on Medical Imaging* 2021 - **Accepted**
3. **Laurent Chauvin**, William Wells III, and Matthew Toews. "A Model of Atom-like Image Structure using 3D SIFT and Discrete CP-Symmetry." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 2022 - **Submitted**

Conference Articles:

1. **Laurent Chauvin** and Matthew Toews. "Curating Subject ID Labels using Keypoint Signatures." *MICCAI Workshop on Large-scale Annotation of Biomedical data and Expert Label Synthesis (LABELS)*, 2020.
2. **Laurent Chauvin**, Kuldeep Kumar, Christian Desrosiers, Jacques De Guise, William Wells, and Matthew Toews. "Analyzing brain morphology on the bag-of-features manifold." *International Conference on Information Processing in Medical Imaging*, pp. 45-56. Springer, Cham, 2019. **(Oral Presentation)**
3. **Laurent Chauvin**, Kuldeep Kumar, Christian Desrosiers, Jacques De Guise, and Matthew Toews. "Diffusion Orientation Histograms (DOH) for diffusion weighted image analysis." *Computational Diffusion MRI*, pp. 91-99. Springer, Cham, 2018.

Abstracts:

1. **Laurent Chauvin**, Sukesh Adiga V, Jose Dolz, Herve Lombaert, Matthew Toews. "A Large-scale Neuroimage Analysis using Keypoint Signatures : UK Biobank". *26th annual meeting of the Organization for Human Brain Mapping (OHBM)*, 2020

0.4 Thesis outline

The work presented in this thesis is organized in 4 chapters, followed by a conclusion and recommendations for future work. In **Chapter 1** local features and their applications will be described as well as key concepts necessary for a full understanding of the work presented here. **Chapter 2** will lay the foundations of this work, presenting the theoretical model of extended 3D SIFT features, as well as a measure of the uncertainty of pairwise features arising from the same underlying distribution based on kernel density estimations over features appearance, location, scale and orientation. A practical application will be demonstrated by integrating this model in the Coherent Point Drift (CPD) algorithm, resulting in the capability of registering images in multiple difficult tasks (e.g. multi-modal images, tumors), more accurately and faster than the original implementation. In **Chapter 3**, an application of our model for quantitatively measuring shared neuroanatomical structures across images, which led to multiple discoveries in widely used public neuroimaging datasets, will be presented. **Chapter 4** is an extension of **Chapter 3**, introducing a new geometrical kernel to reduce the impact of false keypoint matches, with similar appearance, but different geometry. A constraint induced by the geometrical kernel is the necessity for images to be aligned. However, this limitation can be lifted by registering images with the algorithm presented in **Chapter 2**. Finally, the main contributions of this thesis, the possible limitations and directions for future developments will be discussed.

CHAPTER 1

LITERATURE REVIEW

Image-based machine learning may be broadly defined in terms of artificial neural networks, where technical advances often entail lowering the difficulty of neural network training for large parameter spaces and enhancing the capacity to generalise from smaller training data sets. Artificial neural networks were designed to model and reproduce biological neural network processes, such as the Rosenblatt perceptron model, in which the output of a neuron is a weighted linear mixture of its inputs (Rosenblatt, 1958). Soon after, the multi-layer perceptron (MLP) (Rosenblatt, 1961) arose, which arranged perceptrons into numerous, feed-forward layers, assessed sequentially. Incorporation of mathematical invariance has been a frequent focus, enabling networks to learn using fewer data points or to generalise to previously unknown material. In the 1980s, LeCun designed the translation-invariant convolutional neural network (CNN) architecture (LeCun *et al.*, 1989), which, for a number of N pixels, reduced the memory requirements of the weights from $O(N^2)$ for fully-connected MLP network layers to $O(N)$ for sets of weights computed locally and shared across the image in a way invariant to image translation, i.e. convolution filters. Despite its efficiency, CNN training using the backpropagation algorithm was too computationally complex for practical applications on CPU-based architecture computers until the development of highly parallel GPU-based convolution, allowing deep CNNs to be trained from 100000s of images, such as the ImageNet dataset (Krizhevsky *et al.*, 2012), containing 1000 objects with 1000 training examples per category. However, the size of the dataset the network has been trained on has a significant impact on its performances (Luo *et al.*, 2018a), especially for complex tasks, with many samples per class to be able properly generalize classes. Another possible issue is the biasing of the model toward its training set (Geirhos *et al.*, 2019) while obscuring idiosyncrasies of samples to preserve only their shared characteristics. Bias is a particularly pernicious, unsolved phenomenon in machine learning, largely as a result of random initialization and stochastic backpropagation, leading to unintended bias and asymmetry in trained filters, negatively affecting interpretation of both low-level of image textures (Torralba & Efros, 2011; Hermann *et al.*, 2020) and high-level

discrimination at the state-of-the-art deep models for face detection (Buolamwini & Gebru, 2018; Serna *et al.*, 2021) and language (Bender *et al.*, 2021). Although deep learning played a significant role in the development of medical imaging (e.g. automatic segmentation (Lai, 2015), tumor detection (Baur *et al.*, 2021), classification (Song *et al.*, 2015)), personalized medicine requires the preservation of individual-specific information. By looking back in the early 2000s, we can find such technology in the field of computer vision and largely popularized by David Lowe (Lowe, 2004): the local handcrafted features (or keypoints).

Keypoint algorithms transform image data $I(\bar{x})$ sampled on a regular $\bar{x} = \{x, y, z\} \in R^3$ pixel lattice into a sparse set of salient image observations or keypoints within a higher-dimensional geometrical space, i.e. $I(\bar{x}, \Theta, \sigma)$ including local orientation $\Theta \in SO(3)$ and scale $\sigma \in R^+$, in addition to simple 3D location $\bar{x}$. Convolutional filters with special mathematical forms are used to discover Laplacian-of-Gaussian keypoints, ensuring invariance to image translation (Harris *et al.*, 1988), scaling (Lindeberg, 1998), similarity (Lowe, 1999), and affine deformations (Mikolajczyk & Schmid, 2004). Until the development of programmable graphics processing units (GPUs), such as the Compute Unified Device Architecture (CUDA) development system, that could leverage its massive parallelization of the backpropagation learning algorithm (Krizhevsky *et al.*, 2012) to train full image-based CNNs, keypoint methods remained the de facto standard for image-based learning. The SIFT algorithm's backbone, developed by David Lowe in 1999 (Lowe, 1999), is the Gaussian scale-space (Lindeberg, 1994) formed by recursive Gaussian convolution filtering, which may be effectively constructed using separable and even uniform 1D filters. The SIFT algorithm has the following advantages: keypoints are identified in a way that is independent of image resolution, object pose, or intensity variations; there is no need for a training procedure or data; Gaussian and Gaussian derivative filters are rotationally symmetric and uniformly sampled; and the SIFT algorithm is unbiased by the specific training data used. Because of these factors, the SIFT keypoint algorithm and descriptor remain an effective and competitive solution in situations when few training examples are available, such as matching photographs of the same exact 3D objects (Bellavia & Colombo, 2020) or scenes (Mishchuk *et al.*, 2017). The SIFT method bears some resemblance, on certain

aspects, with the mammalian vision system, such as the retinal center-surround processing in the lateral geniculate nucleus and the directed gradient operators organised into hypercolumns as discovered by Hubel & Wiesel in 1968 (Hubel & Wiesel, 1968). The scale-space resulting from Gaussian convolution operations $I(x, y, z, \sigma) = I(x, y, z) * N(\sigma)$ representing the process of isotropic diffusion of particles (such as the Brownian motion) or heat. The Laplacian-of-Gaussian operator $\nabla^2 I(x, y, z, \sigma)$ is a scale-normalized operator that is equivalent to the Laplacian operator that forms the kinetic energy component of the Hamiltonian, such as in the Schrodinger equation.

With the popularization of volumetric images, it became necessary to generalize the SIFT algorithm to 3D image data, used in a variety of applications including video processing (Scovanner *et al.*, 2007), 3D object detection (Flitton *et al.*, 2010), and medical image analysis (Cheung & Hamarneh, 2009; Allaire *et al.*, 2008; Toews *et al.*, 2013a; Rister *et al.*, 2017). Our work here is based on the 3D SIFT-Rank approach (Toews *et al.*, 2013a), which is unique in the way local keypoint orientation and descriptors are computed. Keypoint orientations may be parameterized via a suitable representation such as quaternions or 3x3 rotation matrices, generally determined from dominant gradients of the scale-space $\nabla I(x, y, z, \sigma)$ in a scale-normalized neighborhood around the keypoint. The 3D SIFT-Rank method identifies dominant peaks in a discrete spherical histogram of gradient, rather than solid angle histograms (Scovanner *et al.*, 2007; Allaire *et al.*, 2008), partial 3D orientation information (Cheung & Hamarneh, 2009) or principal components of the local gradient (Rister *et al.*, 2017). This allows identification of multiple dominant orientations (3x3 rotation matrices) at each keypoint (x, y, z, σ) , leading to multiple descriptors per keypoint and providing robustness to noise, rather than single orientations per keypoint (Rister *et al.*, 2017). Descriptors generally encode the scale-space gradient $\nabla I(x, y, z, \sigma)$ in the neighborhood about the keypoint, following reorientation. The 3D SIFT-Rank method adopts a compact 64-element gradient orientation descriptor, where the local orientation space is sampled according to $x \times y \times z = 2 \times 2 \times 2 = 8$ spatial bins and 8 orientation bins. Other approaches adopt descriptors that are orders of magnitude larger, e.g. 768-elements from 4^3 spatial bins x 12 orientation bins, leading to similar image matching performance at a much greater memory footprint (Rister *et al.*, 2017). Finally, the SIFT-Rank descriptor is

normalized by ranking (Toews & Wells, 2009), offering invariance to monotonic variations in image gradient.

The 3D SIFT-Rank method has been used in a variety of keypoint applications analyzing 3D images of the human body. These include modeling the development of the infant human brain over time (Toews *et al.*, 2012), keypoint matching between different image modalities having multi-modal, non-linear intensity relationships (Toews *et al.*, 2013b), robust alignment of lung scans (Gill *et al.*, 2014), kernel density formulation for efficient memory-based indexing from large image datasets (Toews *et al.*, 2015), efficient whole-body medical image segmentation via keypoint matching and label transfer (Wachinger *et al.*, 2015b, 2018), alignment of 4D cardiac ultrasound sequences (Bersvendsen *et al.*, 2016), alignment of 3D ultrasound volumes for panoramic stitching (Ni *et al.*, 2008), identifying family members from brain MRI (Toews & Wells III, 2016), and large-scale population studies using multiple neurological MRI modalities (Kumar *et al.*, 2018b). The 3D SIFT-Rank method was also used for robust image alignment in the context of image-guided neurosurgery (Luo *et al.*, 2018c), including non-rigid registration (Machado *et al.*, 2018b) based on regularization of deformation fields via thin-plane splines and finite element models (Frissen *et al.*, 2019, 2020), and robust filtering of image-to-image correspondences using the variogram (Luo *et al.*, 2018b).

Alternative keypoint descriptors have been proposed based on vector data, e.g. diffusion MRI histograms of the human brain (Chauvin *et al.*, 2018). In terms of keypoint descriptors, a number of 2D descriptors have been proposed, generally based on local image gradient orientation information, including gradient orientation (Lowe, 1999), ORB (Rublee *et al.*, 2011), BRIEF (Calonder *et al.*, 2011), typically these have not been extended to 3D. Rank-order normalization has been shown to improve upon standard 2D gradient descriptors (Toews & Wells, 2009). Deep learning has been used to extract keypoints and descriptors, learned invariant feature transform (LIFT) (Yi *et al.*, 2016), DISK (Tyszkiewicz *et al.*, 2020), LF-Net (Ono *et al.*, 2018), SuperPoint (Detone *et al.*, 2018), Hardnet (Mishchuk *et al.*, 2017). Surprisingly, variants of the original SIFT histogram descriptor including SIFT-Rank (Toews & Wells, 2009), DSP-SIFT (Dong & Soatto, 2015) or RootSIFT (Arandjelović & Zisserman, 2012) are still

competitive in terms of keypoint matching performance (Balntas *et al.*, 2017; Schönberger *et al.*, 2017), particularly for non-planar objects and image retrieval (Bellavia & Colombo, 2020). SIFT keypoint extraction is based solely on fundamental, symmetric and uniform mathematical operators, e.g. the Gaussian, Laplacian, and uniformly sampled gradient operators. The resulting keypoints thus represent a class of highly informative patterns that are invariant to image scaling and rotation in addition to translation, and may be identified in any context with no explicit training procedure and nor bias towards specific training datasets used.

However, most of the descriptors previously presented here are limited to single imaging modality. In the case of multi-modal images, local features have to be adapted to account for the possible contrast inversion between images. The most popular approach consists in computing contrast-invariant descriptors from different image modalities, by using gradient and region reversal (also including the image intensity inversion). A first attempt has been made in (Kelman *et al.*, 2007), where authors proposed a modification of the 2D SIFT descriptor, SIFT-GM, by mirroring gradients with an orientation larger than π , in order to reduce the range of the orientation histogram from $[0, 2\pi)$ to $[0, \pi)$. This method is known as the gradient reversal. However, authors failed to consider that a gradient reversal might also induce a change of direction of the dominant gradient, hence resulting in a region reversal. This issue has been fixed in the Symmetric-SIFT descriptor (Chen & Tian, 2009), that accounts for gradient and region reversal. This work has been the base of multiple variants, such as (Hossain, 2012) and (Teng *et al.*, 2015) that proposed to use the Symmetric-SIFT descriptor as a first step, before computing a global orientation difference between local features, and recomputing descriptors according to this new orientation. In (Hossain *et al.*, 2010) authors developed a variant by using the average of squared difference of gradient magnitude in the Symmetry-SIFT descriptor orientation histogram to overcome some limitation that may create similar descriptors for visually different regions. (Lv & G, 2019) propose to evaluate the angles between keypoints main orientation using pairs of triplets instead of pair of single features as proposed in the Symmetric-SIFT descriptor. However, to limit the complexity explosion of all triplets combinations, a subset of keypoints has to be selected first. (Bingjian *et al.*, 2011) proposed to extract SIFT features to identify

salient regions in the image and compute the shape-context descriptor for each keypoint. After a RANSAC to eliminate outliers, a transform is evaluated using the euclidean distance between shape-context descriptors and a least mean square (LMS) approach. In (Toews *et al.*, 2013b), authors propose a 3D implementation of the SIFT-Rank (Toews & Wells, 2009) with a gradient and region reversal, where point sets are registered through an iterative process based on the Hough transform for transform estimation. However, except for (Toews *et al.*, 2013b; Rister *et al.*, 2017), all these methods focus on 2D descriptors.

In terms of neuroimage analysis, local features have been widely used for a ranged of applications. In the context of inference and classification, feature-based morphometry has been used in identify group-related anatomical features relating to disease (Toews *et al.*, 2010), to predict infant developmental age (Toews *et al.*, 2012), to classify neurodegenerative diseases (Chen *et al.*, 2014) or mental disorders (Castellani *et al.*, 2012). In the context of image registration, local features have been combined with various point cloud registration algorithms such as the Iterative Closest Point (ICP) (Sharp *et al.*, 2002; Jiao *et al.*, 2019) or the Coherent Point Drift (CPD) (Xia *et al.*, 2013) methods. Local features have also been used to estimate non-rigid or deformable registration solutions (Franz *et al.*, 2006; Urschler *et al.*, 2006; Luo *et al.*, 2018c; Machado *et al.*, 2018b), and in the context of multiple image modality registration (Hossain *et al.*, 2011). The majority of the applications in this thesis focus on brain anatomy from structural MRI, however descriptor analysis has also been extended to diffusion MRI (dMRI) (Chauvin *et al.*, 2018; Kumar *et al.*, 2018b; Cheung & Hamarneh, 2009) and functional MRI (fMRI) (Hardoon *et al.*, 2007; Shahamat & Pouyan, 2015).

Scientific breakthroughs in understanding the human brain include the cortical homunculus (Penfield & Boldrey, 1937) mapping sensory inputs and motor outputs to the human body surface, the retinotopic neural structure of the visual cortex (Hubel & Wiesel, 1968), the invention of the magnetic resonance imaging (MRI) (Lauterbur, 1973), the functional MRI (fMRI) measuring the blood oxygenation level or BOLD signal in brain regions (Belliveau *et al.*, 1991), the diffusion MRI (dMRI) measuring water diffusion and axonal connectivity (Basser *et al.*, 1994), the development of group analysis software including based on voxel-based morphometry

(VBM) (Ashburner & Friston, 2000) and functional and structural libraries (FSL) (Smith *et al.*, 2004), and finally large publicly available datasets including the Open Access Series of Imaging Studies (OASIS) (416 subjects for the cross-sectional study, 1098 for the longitudinal) (Marcus *et al.*, 2007), the Alzheimer's Disease Neuroimaging Initiative (ADNI) (1400 subjects, Alzheimer's disease) (Jack Jr *et al.*, 2008), the Human Connectome Project (1200 young adults, 405 families) (Van Essen *et al.*, 2013) or the UK Biobank (100,000s of subjects) (Sudlow *et al.*, 2015).

CHAPTER 2

REGISTERING IMAGE VOLUMES USING 3D SIFT AND DISCRETE SP-SYMMETRY

Laurent Chauvin^a, William Wells III^{b,c}, and Matthew Toews^a

^a Department of Systems Engineering, École de Technologie Supérieure, Québec, Canada

^b Brigham and Women's Hospital, Harvard Medical School, USA

^c Computer Science and Artificial Intelligence Lab, Massachusetts Institute of Technology, USA

Paper submitted in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, June 2022

Abstract

This paper proposes to extend local image features in 3D to include invariance to discrete symmetry including inversion of spatial axes and image contrast. A binary feature sign $s \in \{-1, +1\}$ is defined as the sign of the Laplacian operator ∇^2 , and used to obtain a descriptor that is invariant to image sign inversion $s \rightarrow -s$ and 3D parity transforms $(x, y, z) \rightarrow (-x, -y, -z)$, i.e. SP-invariant or SP-symmetric. SP-symmetry applies to arbitrary scalar image fields $I : R^3 \rightarrow R^1$ mapping 3D coordinates $(x, y, z) \in R^3$ to scalar intensity $I(x, y, z) \in R^1$, generalizing the well-known charge conjugation and parity symmetry (CP-symmetry) applying to elementary charged particles. Feature orientation is modeled as a set of discrete states corresponding to potential axis reflections, independently of image contrast inversion. Feature descriptors augmented to include a binary sign are invariant to discrete coordinate reflections and intensity contrast inversions, in addition continuous 3D similarity transforms. Feature properties are factored in to probabilistic point-based registration as symmetric kernels, based on a model of binary feature correspondence. Experiments using the well-known coherent point drift (CPD) algorithm demonstrate that SIFT-CPD kernels achieve the most accurate and rapid registration of the human brain and CT chest, including multiple MRI modalities and abnormal local variations such as tumors or occlusions.

2.1 Introduction

3D structure is often described in terms of localized phenomena or features. In classical physics, fine-scale matter in the atomic model is characterized via the Schrodinger equation (Schrödinger, 1926), where the symmetry of the Laplacian operator ensures that atomic properties remain invariant to changes in the viewpoint of the observer. Similarly, elementary 3D image features may be identified in generic image volumes via fundamental mathematical operators, e.g. the 3D Laplacian-of-Gaussian operator (Toews *et al.*, 2013a; Rister *et al.*, 2017), following from scale-space theory and the 2D scale-invariant feature transform (SIFT) (Lindeberg, 1998; Lowe, 2004). These are used in highly robust and efficient memory-based applications, e.g. detection (Flitton *et al.*, 2010), registration (Machado *et al.*, 2018b; Luo *et al.*, 2018c), segmentation (Wachinger *et al.*, 2018; Gill *et al.*, 2014), indexing (Chauvin *et al.*, 2021; Toews *et al.*, 2015) in a manner invariant to transformations including translation, rotation and scaling, and are applicable in generic contexts with no training or calibration procedures.

Our work here seeks to advance the understanding of 3D keypoints observed in image volumes from the notion of symmetry, specifically discrete symmetry in feature orientation. A geometrical object may be said to exhibit a symmetry if its observed properties remain unaffected by (or are invariant to) a group of transforms, i.e., a symmetry group. In particle physics, discrete symmetries include axis reflections and also simultaneous charge conjugation (inversion), 3D parity transformation and time reversal (CPT-symmetry). The CPT transform models the transition of a charge-associated particle to its anti-particle, e.g. a proton to an anti-proton, and is currently a topic of high interest (Borchert *et al.*, 2022; Lehnert, 2022) as symmetry violations have been observed in special cases including charge conjugation and parity (CP) transforms (Christenson *et al.*, 1964; Sozzi, 2008). Nevertheless in image analysis, discrete shape symmetries including reflections (Liu *et al.*, 2010; Cohen *et al.*, 2019) and image contrast inversion (Chen & Tian, 2009; Hossain, 2012; Teng *et al.*, 2015; Alexander *et al.*, 2001; Lv G, 2019) are typically represented as separate, unrelated phenomena, typically in 2D coordinate space where 3D parity does not apply. Indeed, image processing and graphics research focuses largely on representations and challenges stemming from 2D data, including photographs,

surface meshes or projections, and achieving invariance to viewpoint changes. Despite the compelling analogy between local features and the atomic model, the notion of a local feature sign, analogous to a particle charge, has not been proposed nor linked to the 3D parity transform in the context of image processing.

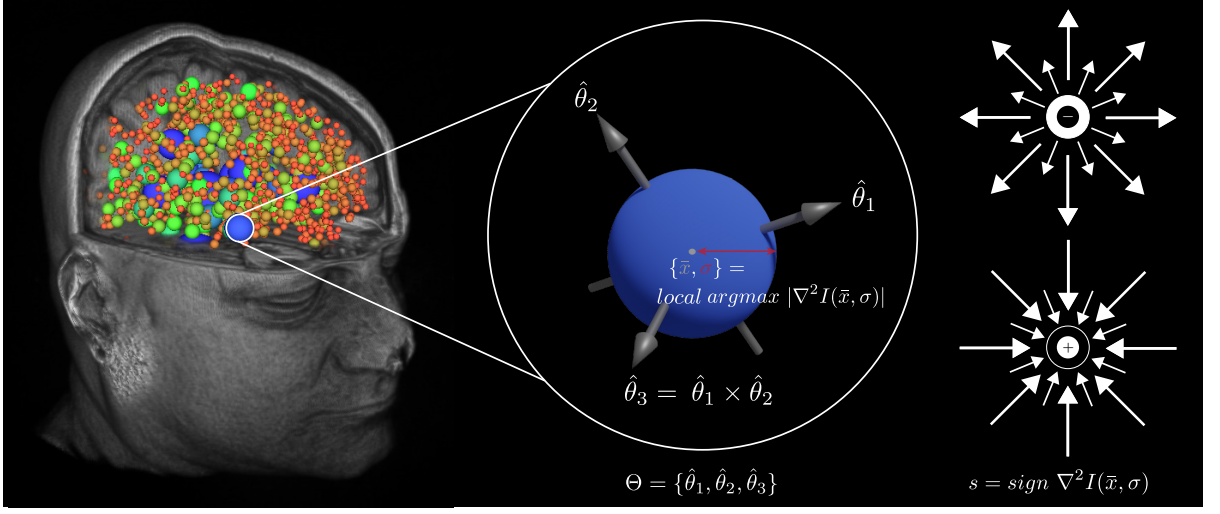


Figure 2.1 Illustrating the geometrical properties of 3D features (colored spheres) extracted from an MRI image of the human brain. Properties include location $\mathbf{x}$, orientation Θ , scale σ , we introduce a binary feature sign $s = \{-, +\}$ based on the sign of the Laplacian-of-Gaussian

Our primary contribution is the first model of discrete symmetry including intensity inversion in the context of local feature processing as shown in Figure 2.1. We note that a CP-like transform occurs in the context of image processing, where a local feature observed in different imaging modalities may exhibit simultaneous contrast inversion and gradient reversal. We refer to this as a sign inversion and parity transform (SP-transform), which applies generally to arbitrary scalar image fields $I : R^3 \rightarrow R^1$ mapping 3D coordinates $(x, y, z) \in R^3$ to a scalar image intensity $I(x, y, z) \in R^1$, and we seek SP-invariance. Local features are endowed with a binary sign $s \in \{-1, +1\}$ and a set of discrete orientation states, allowing feature descriptors and image registration to be computed in a manner invariant to reflections and intensity contrast inversions, thus achieving SP-invariance. Volumetric data allows physical matter and local image features to be observed in situ within isotropic 3D space and free from projective distortion, thus facilitating

the analogy between local image features and atoms in terms of charge. Local feature orientation is defined by a 3D rotation matrix or coordinate reference frame, where two primary axes are estimated from dominant image gradient directions and the third vector is defined by the cross product, ensuring feature reference frames are restricted to $SO(3)$ despite potential parity transform due to contrast inversion. Four discrete orientation states are defined, corresponding to inversions of the two primary axis vectors, due to a combination of symmetric image gradient patterns and intensity contrast inversion. Gradient-based descriptors may be inverted via sign change, independently of orientation state.

Our secondary contribution is to incorporate invariant feature properties into a probabilistic point-based registration framework. The variability of feature scale, orientation and location between images is modeled via an exponential kernel function, approximating a zero-mean noise model, invariant to feature reflection and intensity contrast inversion. Experiments adapt our model to the well-known Coherent Point Drift (CPD) ([Myronenko & Song, 2010](#)) algorithm, and demonstrate that our SIFT-CPD model of enhanced feature geometry leads to faster and more accurate inter-subject registration, in a diverse variety of image data including multiple T1 and T2 weighted MRI modalities of the human brain, tumors and CT images of the human torso. We also demonstrate that our model may be used to achieve SP-invariance independently of the method used to estimate feature orientation, e.g. via principal component analysis or maximum gradient directions.

The remainder of this paper describes related work, method and experiments. Throughout the paper, we aim to provide an accessible yet complete presentation of related concepts in 3D geometry, image processing and particle physics.

2.2 Related work

Our work focuses on characterizing 3D features extracted from volumetric images, and extending invariance to account for discrete symmetry due to both axis reflection and intensity contrast inversion.

2.2.1 Local Image Features

In computer vision and image processing, informative image structures are typically identified via mathematically-defined filters or convolution operators. First order gradient ∇I operators (Harris *et al.*, 1988; Shi & Tomasi, 1994) may be used to identify corner-like structures, in a manner invariant to translation and rotation. Second-order derivative operators such as the Laplacian $\nabla^2 I$ may be used to identify blob-like structures, and to extend feature invariance to scale or size based on Gaussian scale-space theory (Lindeberg, 1998), as in the seminal scale-invariant feature transform (SIFT) (Lowe, 2004). The gradient and Laplacian operators may be regarded as minimal filter representations required to localize singular patterns in a scalar image $I : \mathcal{R}^D \rightarrow \mathcal{R}^1$ defined over a D -dimensional coordinate space. Extended representations may be used to achieve affine invariance (Mikolajczyk & Schmid, 2005), or make use of alternative representations, eg. non-linear scale spaces (Alcantarilla *et al.*, 2012), information theory (Kadir & Brady, 2001; Toews & Wells, 2010), integral images for efficiency (Bay *et al.*, 2008). Over-complete filters banks, ie. convolutional neural networks (CNNs), may be trained for the purpose of identifying point correspondences (Jiang *et al.*, 2021; Ono *et al.*, 2018; Noh *et al.*, 2017), possibly from training labels generated from fundamental operators (Yi *et al.*, 2016; Detone *et al.*, 2018).

Local features incorporate descriptors of the image content in a region or receptive field. The Gaussian scale-space (Lindeberg, 1998) may be viewed as a CNN with a single Gaussian channel per layer (Carlier *et al.*, 2021), where the scale or receptive field size is defined explicitly by the Gaussian standard deviation σ , and is sampled in 3 logarithmic increments between $[\sigma, 2\sigma]$ prior to subsampling. The receptive fields of deep network activations are a function of the filter sizes and depth (Luo *et al.*, 2016), and may be relatively large and dependent on the image contents due to non-linear operations including max pooling and ReLu (Noh *et al.*, 2017). Traditional descriptors have made use of gradients ∇I , including gradient orientation histograms SIFT (Lowe, 2004), binary comparisons (Calonder *et al.*, 2011) and variants such as rank ordering (Toews & Wells, 2009), RootSIFT (Arandjelović & Zisserman, 2012). Descriptors may be learned via triplet loss and gradient analysis (Ono *et al.*, 2018; Tyszkiewicz *et al.*, 2020). While GPU-based deep learning is the defacto state-of-the-art for image classification (Krizhevsky

et al., 2012), variants of the traditional gradient representations remain competitive with learned variants (Balntas *et al.*, 2017; Schönberger *et al.*, 2017), specifically for matching images of non-planar objects (Moreels & Perona, 2007) and image retrieval (Bellavia & Colombo, 2020).

The SIFT algorithm has been generalized to 3D volumetric images. A body of work investigates SIFT-like detectors in 3D video coordinates (x, y, t) defined by 2D space x, y and 1D time t (Laptev, 2005; Scovanner *et al.*, 2007). In the context of 3D space x, y, z , applications include object detection (Flitton *et al.*, 2010), medical image analysis (Cheung & Hamarneh, 2009; Allaire *et al.*, 2008; Toews *et al.*, 2013a; Rister *et al.*, 2017), segmentation (Gill *et al.*, 2014; Wachinger *et al.*, 2018), alignment (Bersvendsen *et al.*, 2016), image stitching (Ni *et al.*, 2008), large-scale indexing (Toews *et al.*, 2015), population studies (Toews & Wells III, 2016; Kumar *et al.*, 2018a). Most recently, the Jaccard distance between feature sets was introduced to automatically flag errors in large public training MRI datasets (Chauvin *et al.*, 2019, 2020), and to identify family members from brain MRI (Chauvin *et al.*, 2021), capabilities facilitated by highly efficient feature indexing. We adopt the 3D SIFT-Rank method (Toews *et al.*, 2013a) using a GPU-optimized implementation (Carlier *et al.*, 2021) first used in the context of brain MRI analysis (Pepin *et al.*, 2020).

2.2.2 Image Registration

Registration is a fundamental image processing task, and seeks to identify an optimal transform $T : \Omega_1 \rightarrow \Omega_2$ between the coordinate systems $\Omega_1 \in R^3$, $\Omega_2 \in R^3$ of the same object or scene, often in three spatial dimensions, based on a pair of observed images (I_1, I_2) . Registration generally makes use of intensity and geometry information (Saiti & Theoharis, 2020), our work focuses on properties of local feature geometry, specifically location, orientation, scale and sign.

Point-based registration approaches are most generally applicable to local features, where the loss function minimizes the distance between pairs of points or points and a model. Examples of point cloud registration algorithms (Pomerleau *et al.*, 2015) include iterative algorithms such as the Iterative Closest Point (ICP) algorithm (Besl & McKay, 1992) where points contribute uniformly

to a solution, or the Coherent Point Drift (CPD) algorithm (Myronenko & Song, 2010) where points contribute according to a probabilistic weighting. 3D SIFT keypoint correspondences have been used to achieve point-based registration in the context of image-guided neurosurgery (Luo *et al.*, 2018c), including non-rigid image registration via thin plate splines (Machado *et al.*, 2018b), finite element methods (Frissen *et al.*, 2019) and the CPD algorithm (Jiao *et al.*, 2019), however these have made use of keypoint locations and not orientation and scale properties as we propose.

Feature-based registration methods generally consider points with properties beyond simple location. Points on a surface model may be endowed with properties such as mass (Golyanik *et al.*, 2016) or charge (Jauer *et al.*, 2019), however these properties tend to be assigned algorithmically, eg. uniformly assigned across points, and not derived from the image content itself. Local properties including orientation, scale and affine deformation may be used to enhance point-based registration (Riggi *et al.*, 2006; Ma *et al.*, 2017). Smooth deformations may be computed in a manner consistent with local rigid or affine reference frames (Arsigny *et al.*, 2009). The variability of geometrical misalignment is related to feature scale σ (Chauvin *et al.*, 2019; Toews & Arbel, 2007), consistent with uncertainty due to Gaussian blur. The localization accuracy of SIFT correspondences has been compared to that of manual human labeling in 2D and 3D (Toews & Arbel, 2007; Machado *et al.*, 2018b), showing similar accuracy, where experts preferred automatic SIFT correspondences in 80% cases (Machado *et al.*, 2018b) in the case of brain imaging. Our 3D SIFT-CPD algorithm proposed in the following section accounts for feature properties of orientation and scale, in addition to point locations.

Multi-modal image registration is a major challenge, where intensities may vary locally in a non-linear manner between modalities. Training may be used to approximate an intensity mapping for a specific domain, e.g. MRI and CT (Hu *et al.*, 2021), however this mapping may not be functional or stationary throughout the image. In the general case with no specific training domain or data, intensity-based registration must adopt statistical similarity measures such as mutual information (Viola & Wells, 1997). In the case of 2D image keypoints, invariance to local contrast inversion may be achieved by transforming the image intensity into a contrast-invariant

format, including phase congruency (Xia *et al.*, 2013) or Laplacian (Wachinger & Navab, 2012) images. As contrast inversion leads to image gradient reversal, attempts have been made to reverse aspects linked to the gradient, including descriptor orientation (Kelman *et al.*, 2007; Chen & Tian, 2009; Hossain, 2012; Bingjian *et al.*, 2011). Combinations of multiple keypoints (Teng *et al.*, 2015) including self-similarity (Lv G, 2019) may be adopted. In preliminary work, we proposed to account for 3D gradient reversal (Toews *et al.*, 2013b), here we extend this to include reflections and SP-symmetry via the notion of a binary sign $s \in \{-1, +1\}$.

2.2.3 Discrete Symmetry

In a physical system, a symmetry refers to a property that remains unchanged under set of transforms, i.e. an invariant property. The symmetry group is the Lie group of transforms under which a geometrical object such as a particle is invariant (Liu *et al.*, 2010). A symmetry may be described as continuous or discrete. Continuous symmetry pertains to continuous parameters of pose and scale, eg. translation and orientation relative to the reference frame of an image acquisition device, and are analogous to a consequence of Noether’s theorem for Lagrangian mechanical systems (Noether, 1971; Tanaka & Kunin, 2021). Continuous symmetry has been investigated in the computer vision literature in terms of differential invariance (Florack *et al.*, 1992; Olver *et al.*, 1994; Lindeberg, 1998). Classical neural networks such as CNNs are generally invariant to translations but not to reflections or scale changes (LeCun *et al.*, 1989; Krizhevsky *et al.*, 2012). Robustness to certain transforms may be improved via regularization (Bardes *et al.*, 2022). Recently models of invariance or equivariance in trainable neural networks have emerged, including SO(3) (Esteves *et al.*, 2018; Spezialetti *et al.*, 2019) rotations, 3D rigid transforms (Moyer *et al.*, 2021) non-linear transforms (Spezialetti *et al.*, 2019), probabilistic symmetries (Bloem-Reddy & Teh, 2020) and gauge theory (Cohen *et al.*, 2019). Much of this work has been limited to 2D image modalities, i.e. surfaces or projective images, and has not considered the 3D notion of SP-symmetry including a binary sign $s \in \{-1, +1\}$.

A discrete symmetry generally involves discrete displacements in location and/or orientation, e.g. translations, rotations, reflections, glide reflections or repetitive crystal structure. Charge

conjugation and parity symmetry (CP-symmetry) (Sozzi, 2008), and generally space-time reversal in CPT-symmetry including time relating to the transition of a particle to its anti-particle (Borchert *et al.*, 2022; Lehnert, 2022) are important aspects of discrete symmetry relating to 3D reflection of elementary physical particles. CP-symmetry is closely related to reflection symmetry, which exists in a function exhibiting one or more planes of symmetry, i.e. a reflection about a single axis $x = 0$ such that $f(x) = f(-x)$, or in a 3D system, the parity coordinate transform $f(x, y, z) = f(-x, -y, -z)$, both characterized in 3D by a rotation matrix R with determinant $\det(R) = -1$. The effect of the parity transform may be seen by first representing $f(x)$ in terms of $f(x) = f_s(x) + f_a(x)$, the sum purely symmetric and anti-symmetric image components $f_s(x)$ and $f_a(x)$ defined by $f_s(-x) = f_s(x)$ and $f_a(-x) = -f_a(x)$. Symmetric or even functions $f_s(x)$ include the Gaussian and Laplacian-of-Gaussian operators, wave functions associated with bosonic particles such as the photon, the cosine function, etc. Anti-symmetric or odd functions $f_a(x)$ include the gradient operator ∇I , wave functions associated with fermionic particles such as electrons, and the sine function, etc. As the gradient is an anti-symmetric operator, an image contrast inversion $-\nabla I(x) = \nabla I(-x)$ is equivalent to an axis reflection. This phenomenon is analogous to charge conjugation and parity (CP) symmetry, which was thought to be an inviolable in nature until the discovery of special cases beginning with the composite meson particle (Christenson *et al.*, 1964). In our work here, charge is analogous to image sign or contrast, which may be inverted, and we propose SP-invariance to register images of differing modality.

Image processing methods specifically relating to our work have analyzed discrete shape symmetry (ie. reflections) (Liu *et al.*, 2010) and image contrast inversion (Chen & Tian, 2009; Hossain, 2012; Teng *et al.*, 2015; Alexander *et al.*, 2001; Lv G, 2019) as separate, unrelated phenomena. Our work is the first to consider these within the unified framework of discrete SP-symmetry, including a binary sign and discrete states of local reference frame orientation. Our method generalizes to various contexts where orientation is estimated from dominant local image gradients ∇I . For example, Toews *et al.* identify maxima in a 3D spherical gradient histogram (Toews *et al.*, 2013a) such that $\hat{\theta}_1 = \underset{\hat{\theta}}{\operatorname{argmax}} |\nabla I \cdot \hat{\theta}|$, $\hat{\theta}_2 = \underset{\hat{\theta}}{\operatorname{argmax}} |\hat{\theta} \times (\nabla I \times \hat{\theta}_1)|$,

with a third axis defined by the cross product $\hat{\theta}_3 = \hat{\theta}_1 \times \hat{\theta}_2$. Rister et al. compute the eigenvectors of the local gradient structure tensor, ie. the 3×3 gradient correlation matrix (Rister *et al.*, 2017). We demonstrate that both such representations may be modeled in terms of discrete SP-symmetry.

2.3 Method

Our method provides a means of characterizing 3D image features in a manner invariant to discrete symmetry transforms, including symmetric image patterns and sign inversion and parity (SP-symmetry) transforms due to image contrast inversion. We begin by describing the properties of a single local image feature, which we augment to include a binary sign and a set of discrete axis reflections. This allows us to compute a descriptor that is invariant to reflections and changes in sign due inversion of intensity contrast. We then define a kernel function $\mathcal{K}(f_n, f_m)$ to quantify the similarity of a pair of features (f_n, f_m) potentially arising from the same underlying structure in two different images. Finally, this kernel is incorporated into probabilistic image registration in order to estimate a transform $T : f_m \rightarrow f_n$ aligning the coordinate systems of two images.

2.3.1 Single-Feature Properties

A feature is a distinctive spherical region localized in 3D image space, as illustrated in Figure 2.1. An individual feature $f = \{g, \mathbf{a}, s\}$ is characterized within the image by its geometry g , a descriptor of the image appearance $\mathbf{a}$ surrounding the feature and a binary sign s . The feature geometry $g = \{\Theta, \sigma, \mathbf{x}\} \in SO(3) \times \mathbb{R}^+ \times \mathbb{R}^3$ is a scaled coordinate reference frame in 3D defined by 7 parameters. These include a scale $\sigma \in \mathbb{R}^+$, a coordinate location $\mathbf{x} \in \mathbb{R}^3$ and a reference frame $\Theta \in SO(3)$. The reference frame may be defined as a set $\Theta = \{\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3\}$ of orthonormal unit vectors $\hat{\theta}_i, i \in [1, 2, 3]$ where $\hat{\theta}_i \cdot \hat{\theta}_j = 0, i \neq j$. Note that alternative representations for rotation could be used, eg. $SU(2)$ Pauli matrices or unit quaternions. Feature appearance $\mathbf{a} \in \mathbb{R}^M$ is an M-dimensional descriptor of the image, resampled according to geometry g . The sign $s \in \{-1, 1\}$ is used to achieve invariance to image contrast inversion.

Let $I : \mathbb{R}^3 \rightarrow \mathbb{R}^1$ be a scalar image sampled over 3D coordinates where $I(\mathbf{x})$ represents a voxel measurement at location $\mathbf{x}$. Let $I(\mathbf{x}, \sigma) = I(\mathbf{x}) * G(\mathbf{x}, \sigma)$ be a scale-space defined by convolution of the image with a Gaussian filter $G(\mathbf{x}, \sigma)$ of parameter σ (Lindeberg, 1998). Features are detected as points in scale-space $\{\mathbf{x}_m, \sigma_m, s_m\}$ at which the magnitude of the scale-normalized Laplacian operator is maximized, similarly to the original SIFT method (Lowe, 2004), here m is a feature index. Furthermore, they are endowed with a binary sign as follows.

$$\begin{aligned} \{\mathbf{x}_m, \sigma_m\} &= \underset{\mathbf{x}, \sigma}{\text{local argmax}} |\nabla^2 I(\mathbf{x}, \sigma)|, \\ s_m &= \text{sign} \left(\nabla^2 I(\mathbf{x}_m, \sigma_m) \right). \end{aligned} \quad (2.1)$$

Scale-space loci $\{\mathbf{x}_m, \sigma_m, s_m\}$ represent generic spherical regions in which the divergence of the local gradient field $\nabla I(\mathbf{x}, \sigma)$ is maximized. Feature sign s_m is novel to our characterization, and is defined as the sign of the Laplacian-of-Gaussian.

Feature orientation Θ is derived from the image gradient ∇I surrounding the feature origin $\mathbf{x}$. Orientation axis vectors $\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3$ may be determined according dominant gradient orientations within a window $w(\mathbf{x}, \sigma) \in [0, 1]$ centered upon $\mathbf{x}$ with spatial extent proportional to scale σ . As $\Theta \in SO(3)$, the determinant $\det(\Theta) = 1$ is positive, ie. no reflections. Axes are thus defined by two primary axis vectors $\hat{\theta}_1$ and $\hat{\theta}_2$, with the third constrained according to a handedness convention, ie. the right-hand rule via the cross product $\hat{\theta}_3 = \hat{\theta}_1 \times \hat{\theta}_2$. Furthermore, we assume that axis vectors may be ordered uniquely according to the gradient magnitude along their directions, i.e. $[|\nabla I \cdot \hat{\theta}_1|] > [|\nabla I \cdot \hat{\theta}_2|]$, where $[|\nabla I \cdot \hat{\theta}_i|] = \int w(\mathbf{x}, \sigma) |\nabla I(\mathbf{u}) \cdot \hat{\theta}_i| d\mathbf{u}$ is the expected value of the gradient magnitude along axis $\hat{\theta}_i$.

Our primary contribution is a descriptor that is invariant to discrete SP-transforms, i.e. is SP-symmetric, based on binary sign s . We first note that the sign of the orientation axes $\hat{\theta}_1, \hat{\theta}_2$ may generally be ambiguous due to a variety of factors including bilateral symmetry of the image pattern or its local gradient distribution (e.g. an ellipsoid or rectangular box), an image contrast inversion (e.g. different imaging modalities), or the algorithm used to estimate orientation (e.g. principal component analysis has an inherent eigenvector sign ambiguity). Sign ambiguity may

be represented by maintaining a discrete set of possible orientation reference frames Θ . We consider four potential reference frame states $\{\pm\hat{\theta}_1, \pm\hat{\theta}_2\}$ (with $\hat{\theta}_3 = \pm\hat{\theta}_1 \times \pm\hat{\theta}_2$) resulting from axis inversion, i.e. due to either intensity contrast inversion or pattern symmetry, which are equivalent to the identity in addition to rotations of π about each of three axes. By defining the third axis vector $\hat{\theta}_3$ using the cross product, we ensure feature orientation follows the right-hand rule. Figure 4.3 illustrates the four possible orientation states, including an initial reference frame and rotations of π about each of the 3 axes.

Feature sign s is used to generate a novel contrast inversion-invariant descriptor $\mathbf{a}_m$ for each orientation state, based on the image content surrounding location $\mathbf{x}_m$ at scale σ_m and orientation Θ_m . Note that the geometry g defines a 7-parameter similarity transform from the local feature reference frame to a canonical reference frame, including rotation, translation and scaling. This is used to resample the image according to a characteristic reference frame $\hat{I}(\mathbf{u}) = I(\sigma_m \Theta_m \mathbf{u} + \mathbf{x}_m)$ that is invariant to global similarity transforms of the image. A variety of descriptors may be computed, eg. spherical harmonics (DiMaio *et al.*, 2009), we adopt a computationally efficient variant of the gradient orientation histogram (Lowe, 2004; Toews *et al.*, 2013a; Rister *et al.*, 2017). As in (Toews *et al.*, 2013a), 3D space surrounding the feature origin $\mathbf{x}_m$ is quantized uniformly into eight discrete octants $\mathbf{r} = [\pm 1, \pm 1, \pm 1]$, and the image gradient at each point $\mathbf{u}$ into eight discrete symmetric directions $\phi = [\pm 1, \pm 1, \pm 1]$. These are combined into an $8 \times 8 = 64$ -element appearance descriptor, by accumulating the gradient magnitude into a histogram over spatial location and orientation

$$a(\mathbf{r}, \phi) = \sum_{\mathbf{u} \in \mathbf{r}} \|\nabla \hat{I}(\mathbf{u}) \cdot \phi\| [\phi = s\phi_u], \quad (2.2)$$

where $\phi_u = \underset{\phi}{\operatorname{argmax}} \{\nabla \hat{I}(\mathbf{u}) \cdot s\phi\}$ is the direction of maximum gradient at image location $\mathbf{u}$, which is inverted in the case of $s = -1$ by the Iverson bracket $[\phi = s\phi_u]$ evaluating to 1 upon equality and 0 otherwise. Thus descriptor orientation elements ϕ are invariant to reversal due to sign change $a(\mathbf{r}, \phi) = a(\mathbf{r}, s\phi)$, and parity transforms of orthants $\mathbf{r} \rightarrow -\mathbf{r}$ are accounted for by feature orientation states $\{\pm\hat{\theta}_1, \pm\hat{\theta}_2\}$, as illustrated in Figure 4.3.

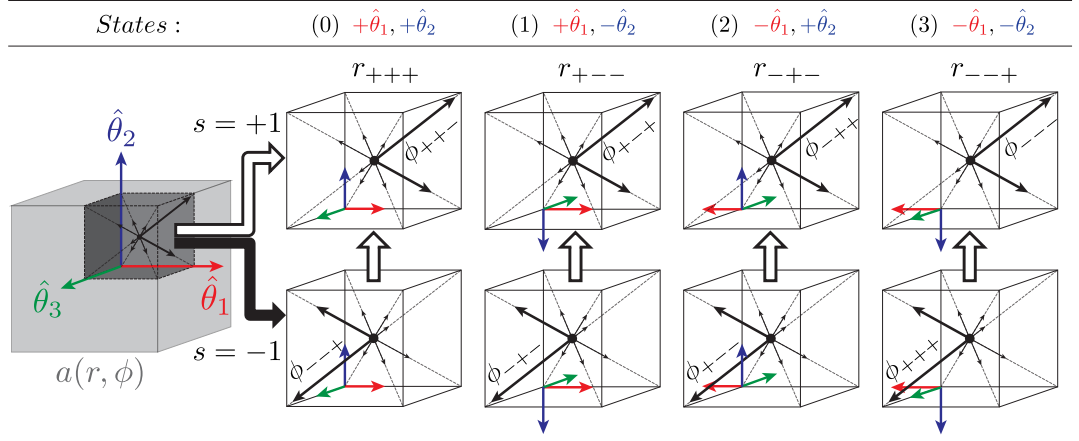


Figure 2.2 Illustrating an appearance descriptor $a(r, \phi)$ and four discrete orientation states given principal axes $\hat{\theta}_1, \hat{\theta}_2$ (left to right) and feature sign s (upper and lower). The descriptor content is shown for an example octant (r_{+++}) and gradient orientation (ϕ_{+++}). Axis orientation is defined independently of sign (left to right), and negative sign $s = -1$ inverts the descriptor gradient (lower row) achieving invariance to contrast inversion

2.3.2 Two-Feature Observations

Image registration requires identifying pairs of features (f_n, f_m) arising from the same underlying structure in different images. To do this, we propose a kernel function $K(f_n, f_m) \in [0, 1]$ to quantify the similarity of two features f_n and f_m potentially sampled from the same distribution. We construct $\mathcal{K}(f_n, f_m)$ as a product of squared exponential kernels, ensuring that it is positive, symmetric $\mathcal{K}(f_m, f_n) = \mathcal{K}(f_n, f_m)$, and proportional to a Gaussian density. It may be expressed as the product of three factors

$$\mathcal{K}(g_n, g_m) = \mathcal{K}_\sigma(\sigma_n, \sigma_m) \mathcal{K}_\Theta(\Theta_n, \Theta_m) \mathcal{K}_x(\mathbf{x}_n, \mathbf{x}_m), \quad (2.3)$$

quantifying the variability in scale, orientation and location, respectively. As in the notions of linear and angular momentum from classical physics, deviations of location and orientation are modeled as orthogonal and independent.

The first factor in Equation (2.3), $\mathcal{K}_\sigma(\sigma_n, \sigma_m)$, is a kernel penalizing the log scale difference:

$$\mathcal{K}_\sigma(\sigma_n, \sigma_m) = \exp\left(-\|\log \sigma_n - \log \sigma_m\|^2\right), \quad (2.4)$$

and is proportional to a log normal density about mean $\log 1 = 0$ with unit variance 1. The log maps multiplicative variations in feature scale on the range $\sigma \in [0, \infty]$ to additive variations on the range $\log \sigma \in [-\infty, \infty]$, which may be modelled as symmetric and Gaussian.

The second factor $\mathcal{K}_\Theta(\Theta_n, \Theta_m)$ quantifies the variability of feature orientations based on the angular displacement of the axes:

$$\mathcal{K}_\Theta(\Theta_n, \Theta_m) = \exp\left(-3 + \sum_{i=1,2,3} \hat{\theta}_{in} \cdot \hat{\theta}_{im}\right). \quad (2.5)$$

In Equation (2.5), the scalar product $\hat{\theta}_{in} \cdot \hat{\theta}_{im} = \cos(\angle(\hat{\theta}_{in}, \hat{\theta}_{im}))$ quantifies the angular separation between axis unit vectors $\hat{\theta}_{in}$ and $\hat{\theta}_{im}$ on the range $[-1, 1]$. This is equivalent to modeling the angular difference between features independently in each axis i as proportional to a von Mises density over the unit circle.

The third factor $\mathcal{K}_x(\mathbf{x}_n, \mathbf{x}_m)$ is a kernel that quantifies the similarity of feature locations based on their squared displacement:

$$\mathcal{K}_x(\mathbf{x}_n, \mathbf{x}_m) = \exp\left(-\frac{\|\mathbf{x}_n - \mathbf{x}_m\|^2}{k \sigma_n \sigma_m + \sigma_T^2}\right), \quad (2.6)$$

where the denominator normalizes the spatial displacement $\|\mathbf{x}_n - \mathbf{x}_m\|^2$ via a linear function of the product of feature scales $\sigma_n \sigma_m$. The first term $k \sigma_n \sigma_m$ embodies intrinsic variance due to the scale of the observed feature, where $\sigma_n \sigma_m$ represents the square of the geometric mean of feature scales and k is a positive proportionality constant. This allows quantifying the deviation in location in a manner invariant to keypoint scale. The second term σ_T^2 represents the minimum achievable variance in mapping T given the particular imaging context, which dominates in the case of small-scaled features, i.e. $k \sigma_n \sigma_m < \sigma_T^2$. Note that in previous work ([Chauvin et al.](#),

2021) $k = 1$ and $\sigma_T = 0$, here we use k and σ_T to estimate a more accurate linear relationship between scale and observed variability. These are empirically set to provide a generally useful weighting, e.g. here we use $k = 12$, $\sigma_T^2 = 200$.

2.3.3 Multi-Feature Registration

Registration seeks a transform $\mathcal{T}$ mapping the coordinate system of a fixed image I_f to that of a moving image I_m , here from sets of features extracted in the fixed $\mathcal{F} = \{f_n\}$ and moving $\mathcal{M} = \{f_m\}$ images. The transform maps the properties of features from one coordinate space to the next $\mathcal{T} : \mathbf{x}_m \rightarrow \mathbf{x}_n$, and here is taken to be a global similarity transform, followed by independent zero-mean deviations for individual features. Note that the similarity transform is a 7-parameter representation $\mathcal{T} = \{d\Theta, d\sigma, d\mathbf{x}\} \in SO(3) \times R^+ \times R^3$ equivalent that of an individual feature.

Let g'_m represent the geometry of feature f_m as transformed by $\mathcal{T}$, ie. $g'_m = T \circ g_m$, note that T may act on location, scale and orientation parameters in a consistent manner. We propose using our kernel function $\mathcal{K}(g_n, g'_m)$ to account for feature properties including scale and orientation in a standard point-based registration framework. Our kernel lends itself naturally to a probabilistic algorithm such as Coherent Point Drift (CPD) in Algorithm 2.1, which seeks a maximum likelihood solution based on the Expectation-Maximization (EM) algorithm (Dempster *et al.*, 1977). The expectation step (E) intend estimates the log-likelihood based on the current parameters, while the maximization step (M) maximizes the log-likelihood over the parameters using the result of the E step. CPD registration is based on a probability map p_{mn} between each pair of points $(\mathbf{x}_n, \mathbf{x}'_m)$, and takes the form of a softmax function. We propose to bias this probability map based on a kernel function $\mathcal{K}(g_n, g'_m)$ as follows:

$$p_{mn} = \frac{\exp\left(-\frac{\|\mathbf{x}_n - \mathbf{x}'_m\|^2}{2\lambda^2}\right) \mathcal{K}(g_n, g'_m)}{\sum_{k=1}^M \exp\left(-\frac{\|\mathbf{x}_n - \mathbf{x}'_k\|^2}{2\lambda^2}\right) \mathcal{K}(g_n, g'_k) + \eta}, \quad (2.7)$$

where $\eta = (2\pi\lambda^2)^{D/2} \frac{w}{1-w} \frac{M}{N}$ is a constant ensuring the denominator p_{mn} is non-zero, where w is parameter accounting for the relative probability of a uniform background density not associated with any specific feature. Equation (2.7) includes an exponential expression with data variance parameter λ^2 and our kernel function $\mathcal{K}(g_n, g'_m)$, note that with $\mathcal{K}(g_n, g'_m) = 1$ it is equivalent to the original CPD algorithm (Myronenko & Song, 2010). Thus in our proposed SIFT-CPD, two exponential factors operate on point coordinates $\mathbf{x}$ including Equation 2.6.

Iterative algorithms such as EM must be initialized within a neighborhood of the solution in order to converge correctly. We initialize registration via a global 3D Hough transform between feature sets $\mathcal{F}$ and $\mathcal{M}$ in 3D (Toews *et al.*, 2013a) analogously to the classic 2D SIFT algorithm (Lowe, 2004), where all features vote independently as to the predicted transform, after which a most likely transform is identified¹. Here, each feature $f_n \in \mathcal{F}$ votes for a transform $\mathcal{T}_{nm} : g_n \rightarrow g_m$ mapping the geometry of feature f_n to that of feature f_m , where $f_m = \underset{f}{\operatorname{argmin}} \|\mathbf{a}_n - \mathbf{a}\|$ is the nearest neighbor (NN) of f_n in terms of the Euclidean distance $\|\mathbf{a}_n - \mathbf{a}_m\|$ between appearance descriptors. Matches are identified between all features and discrete orientation states $\{\pm\hat{\theta}_1, \pm\hat{\theta}_2\}$ including parity transforms, and in a manner invariant to intensity contrast inversions due to the binary sign $s \in \{-1, +1\}$ in Equation (2.2). A dominant Hough transform $\mathcal{T}^* \in \{\mathcal{T}_{nm}\}$ is then identified as the transform consistent with the largest number of matching features or inlier correspondences. A matching feature pair (f_n, f_m) is an inlier of the transform $\mathcal{T}^*$ if their transform $\mathcal{T}_{nm} : g_n \rightarrow g_m$ differs by less than a threshold $\|\mathcal{T}_{nm} - \mathcal{T}^*\| < Thres$. $\mathcal{T}^*$ is thus defined as:

$$\mathcal{T}^* = \underset{T}{\operatorname{argmax}} |\{T : \|\mathcal{T}_{nm} - T\| < Thres\}|, \quad (2.8)$$

¹ The Hough transform method was originally proposed to track particles in bubble chamber photographs (Hough, 1959)

where the thresholding operation in Equation (2.8) may be defined as the logical conjunction of thresholds independently applied in rotation, scaling and squared displacement

$$\begin{aligned}
 \|\mathcal{T}_{nm} - \mathcal{T}\| < Thres &\rightarrow \cos(\hat{\theta}_i - \hat{\theta}_{inn}) < \epsilon_{cos\theta} \\
 &\wedge |\log \sigma - \log \sigma_{nm}| < \epsilon_{\log \sigma} \\
 &\wedge \|\mathbf{x} - \mathbf{x}_{nm}\|^2 / \sigma \sigma_{nm} < \epsilon_{x/\sigma},
 \end{aligned} \tag{2.9}$$

using thresholds $(\epsilon_{cos\theta}, \epsilon_{\log \sigma}, \epsilon_{x/\sigma})$. These may be set generously to apply in a wide variety contexts, here we use $(\epsilon_{cos\theta}, \epsilon_{\log \sigma}, \epsilon_{x/\sigma}) = (0.7, \log 1.5, 0.25)$. The Hough transform may be implemented efficiently using approximate NN search (Muja & Lowe, 2014) $\mathcal{T}_{nm}$ similar to the mean-shift algorithm (Comaniciu & Meer, 2002). Our modified CPD algorithm is provided in Algorithm 2.1, and the *solve* functions for similarity transforms is found in Algorithm 9 of Appendix III. The algorithm iterates until the variance λ^2 converges.

Algorithm 2.1 The probabilistic SIFT-CPD registration algorithm adapted from (Myronenko & Song, 2010) to include kernel $\mathcal{K}(g_n, g'_m)$

Inputs	$\mathcal{F} = \{f_n\}$	Fixed Features
	$\mathcal{M} = \{f_m\}$	Moving Features
Outputs	$\mathcal{T} : \mathcal{M} \rightarrow \mathcal{F}$	Transform
Initialization	$0 \leq w \leq 1$	
	$\lambda^2 = \frac{1}{DNM} \sum_{n=1}^N \sum_{m=1}^M \ \mathbf{x}_n - \mathbf{x}_m\ ^2$	
	$\mathcal{T} \leftarrow \mathcal{T}^*$	

- 1 **EM Optimization:** repeat until convergence:
- 2 **E-Step:** Compute matrix P :

$$p_{mn} = \frac{\exp^{-\frac{1}{2\lambda^2} \|\mathbf{x}_n - \mathbf{x}'_m\|^2} \mathcal{K}(g_n, g'_m)}{\sum_{k=1}^M \exp^{-\frac{1}{2\lambda^2} \|\mathbf{x}_n - \mathbf{x}'_k\|^2} \mathcal{K}(g_n, g'_m) + \eta}$$
- 3 **M-Step:** Solve for optimal transformation $\mathcal{T}$
- 4 $\{\mathcal{T}, \lambda^2\} = solve(\mathcal{F}, \mathcal{M}, P)$
- 5 The aligned point set is $\mathcal{T}(\mathcal{M})$
- 6 The correspondence probability is P .

2.4 Experiments

We hypothesize that our kernel function $\mathcal{K}(g_n, g_m)$ incorporating feature scale and orientation will improve the accuracy of registration beyond point information alone, and that feature orientation state and binary sign will provide invariance to intensity contrast inversion and reflections, independently of the method used to identify feature orientation Θ . Experiments validate these hypotheses in the context of 3D medical imaging data, which provide the opportunity to investigate features arising from realistic, diverse anatomical structure observed across subjects and 3D imaging modalities.

Three sets of experiments are performed. The first is based on synthetic similarity transforms, and establishes baseline accuracy of methods in the case of known ground truth. The second consists of inter-subject image registration trials, first between brain MRI and then between chest CT volumes of different people, with confounds including brain tumors and contrast variations due to multiple T1 and T2-weighted MRI modalities, demonstrating superior algorithm performance in a diverse set of contexts. Variability is estimated by repeating trials following synthetic transforms of images and multiple subjects. The third investigates orientation state changes in registration trials for different feature orientation estimation methods, i.e., principal components (Rister *et al.*, 2017) or maximum orientation directions (Toews *et al.*, 2013a).

Note that the true mapping T between the anatomies of different subjects is generally non-linear and may not exist throughout the image, due to aspects of anatomy specific to individuals or in the case of occlusion or missing structure. As mentioned, features are assumed to follow a transform T that is globally linear followed by random feature-specific deviations as specified by our kernel $\mathcal{K}(g_n, g_m)$. For inter-subject registration, ground truth is established from inlier correspondences identified via the Hough transform, which are visually validated for correctness and rejected or manually adjusted if necessary. This follows the work of (Toews & Arbel, 2007; Machado *et al.*, 2018b), where automatic and manually labeled correspondences exhibit the same error range. The average Point Registration Error (PRE) measure is used to quantify

registration performance, based on the sum of 3D point differences between the registration solution and ground truth.

In all experiments, feature sets are extracted using a GPU implementation of the SIFT-Rank algorithm for computational efficiency (Carluer *et al.*, 2021), and registered using five point-cloud registration methods: ICP with 20 and 100 iterations (respectively ICP20 and ICP100), the original CPD algorithm using feature centers $\mathbf{x}$, SIFT-CPD using full feature geometry and our kernel $\mathcal{K}(g_n, g_m)$, and SIFT-CPD* using only inlier features from the Hough transform. All algorithms are initialized to the Hough transform solution prior to iterative registration in order to ensure a fair comparison.

2.4.1 Synthetic Image Registration

A preliminary experiment was first performed establish the baseline performances of algorithms against known ground truth, here synthetic similarity transformations of a single image. A T1w brain image from the Human Connectome Project (HCP) (Van Essen *et al.*, 2012) dataset was selected (see Table 2.2 for more details), and 100 different synthetic transforms were applied, where each was generated by a random rotation about each axis on an angle range of $\rho \in \pm[10^\circ, 30^\circ]$ followed by a translation $\delta \in \pm[0 \text{ mm}, 10 \text{ mm}]$. The registration error in rotation and translation was evaluated relative to the known transform applied, and the results are presented in Table 2.1. As expected, the lowest error is achieved for SIFT-CPD* (inliers alone), and this serves as a minimum error baseline. SIFT-CPD and CPD perform similarly, with slightly less error for SIFT-CPD. ICP results are generally poor, failing to converge in 8 and 21 cases for 20 and 100 iterations.

2.4.2 Inter-subject Registration

Four sets of experiments quantified the accuracy of inter-subject registration, i.e. registration of images of different individuals. The goal was to evaluate the accuracy and speed of registration in a diverse set of contexts, including different aspects of anatomy (brain, chest), abnormalities

Table 2.1 Comparison of five registration algorithms based on 100 synthetic transforms of a single T1w MRI brain volume. Error is listed terms of individual rotation and translation axes, overall point registration error (PRE), and the sum of squared intensity differences (SSD) following registration

Error Algorithm	Rotation Error (deg)			Translation Error (mm)			PRE (mm)	SSD (e^{11})
	X-Axis	Y-Axis	Z-Axis	X-Axis	Y-Axis	Z-Axis	Mean $\pm$ Std.Dev.	Mean $\pm$ Std.Dev.
ICP20	25.7 ± 22.0	16.8 ± 14.8	24.8 ± 17.2	14.2 ± 19.6	12.8 ± 29.1	17.4 ± 29.0	16.37 ± 8.41	7.62 ± 1.03
ICP100	8.1 ± 18.9	6.3 ± 18.8	8.5 ± 18.5	9.0 ± 24.0	9.2 ± 32.8	8.5 ± 23.2	8.16 ± 17.15	4.89 ± 2.31
CPD	0.06 ± 0.03	0.06 ± 0.03	0.06 ± 0.03	0.8 ± 0.6	1.3 ± 1.1	0.7 ± 0.5	1.81 ± 0.88	1.38 ± 0.39
SIFT-CPD	0.02 ± 0.01	0.01 ± 0.01	0.02 ± 0.01	0.9 ± 0.5	0.8 ± 0.6	0.5 ± 0.4	1.05 ± 0.32	1.43 ± 0.33
SIFT-CPD*	0.02 ± 0.01	0.01 ± 0.01	0.02 ± 0.01	0.9 ± 0.4	0.9 ± 0.4	0.4 ± 0.3	0.68 ± 0.15	0.96 ± 0.39

(tumor) and modalities (T1w and T2w MRI, CT). Information regarding datasets is provided in Table 2.2. Note that minimum registration error is generally non-zero due to differences in anatomy, and the goal is to identify a robust transform minimizing the PRE for sets of inlier correspondences.

Table 2.2 Inter-subject registration experimental data including the numbers of features per image and Hough transform inlier matches

Dataset	Anatomy	Modality	Resolution	Voxel Size (mm)	Features	Inliers
(Van Essen <i>et al.</i> , 2012)	Brain	T1w MRI	$260 \times 260 \times 311$	$0.7 \times 0.7 \times 0.7$	6000	350
(Regan <i>et al.</i> , 2011)	Chest	CT	$280 \times 280 \times 235$	$1.2 \times 1.2 \times 1.2$	3000	70
(Menze <i>et al.</i> , 2015)	Brain Tumor	T1w MRI	$240 \times 240 \times 155$	$1.0 \times 1.0 \times 1.0$	1200	70
(Van Essen <i>et al.</i> , 2012)	Brain	T1w-T2w MRI	$260 \times 260 \times 311$	$0.7 \times 0.7 \times 0.7$	5000	115

2.4.2.1 T1w Brain Images

The following experiment quantified inter-subject registration accuracy between brain images of a pair of different subjects. As in the previous experiment, one image was fixed, while the other was subjected to a random transform from the previously mentioned ranges ρ and δ . A pair of monozygotic twin subjects was selected, which generally share more inlier correspondences in

comparison to unrelated subjects due to higher neuroanatomical similarity (Chauvin *et al.*, 2020). The results for 100 random transforms are shown in Figure 2.3, these are generally consistent across all registration experiments. ICP variants which weight all points uniformly do not correctly converge (red and orange). CPD and SIFT-CPD (green and light blue) both converge consistently to acceptable results in term of PRE, both lower than that the initial robust Hough transform (purple). SIFT-CPD is consistently lower in terms of error and computation time in comparison to CPD, demonstrating that geometrical properties improve estimation, specifically in the case where the majority of features may be outliers with no valid correspondence between images. Overall, the SIFT-CPD* (dark blue) leads to the lowest error and computation time, as it refines the original set of Hough transform inliers.

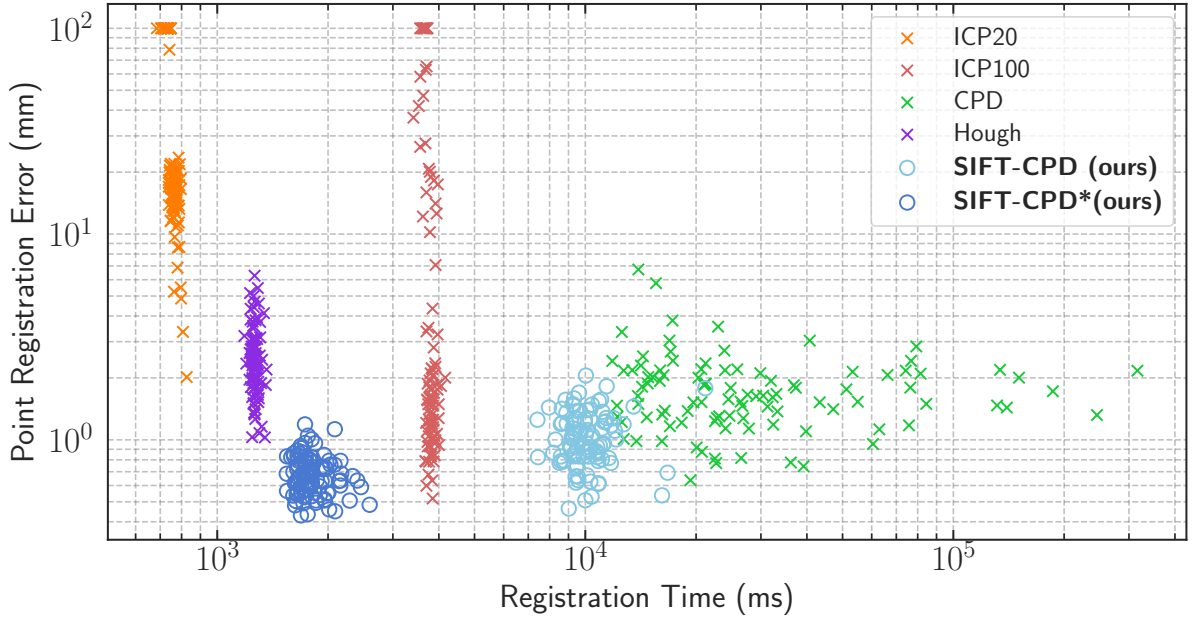


Figure 2.3 Inter-subject registration performance for healthy brain T1w MRI

2.4.2.2 CT Chest Images

In order to show that our method is not tied to specific contexts or imaging modalities, we performed inter-subject registration of CT chest images from the Chronic Obstructive Pulmonary Disorder Genetics (COPDGene) dataset (Regan *et al.*, 2011). Two different subject images in

expiration breathing state were randomly selected (see Table 2.2 for more details). As in the previous experiment, we generated 100 random transform with the same ρ and δ parameters, applied the transform to one image, and registered it using different registration approaches. The PRE for each registration is shown in Figure 2.4. Although the computational complexity of SIFT-CPD is comparable to that of CPD in Figure 2.4, SIFT-CPD* is significantly faster.

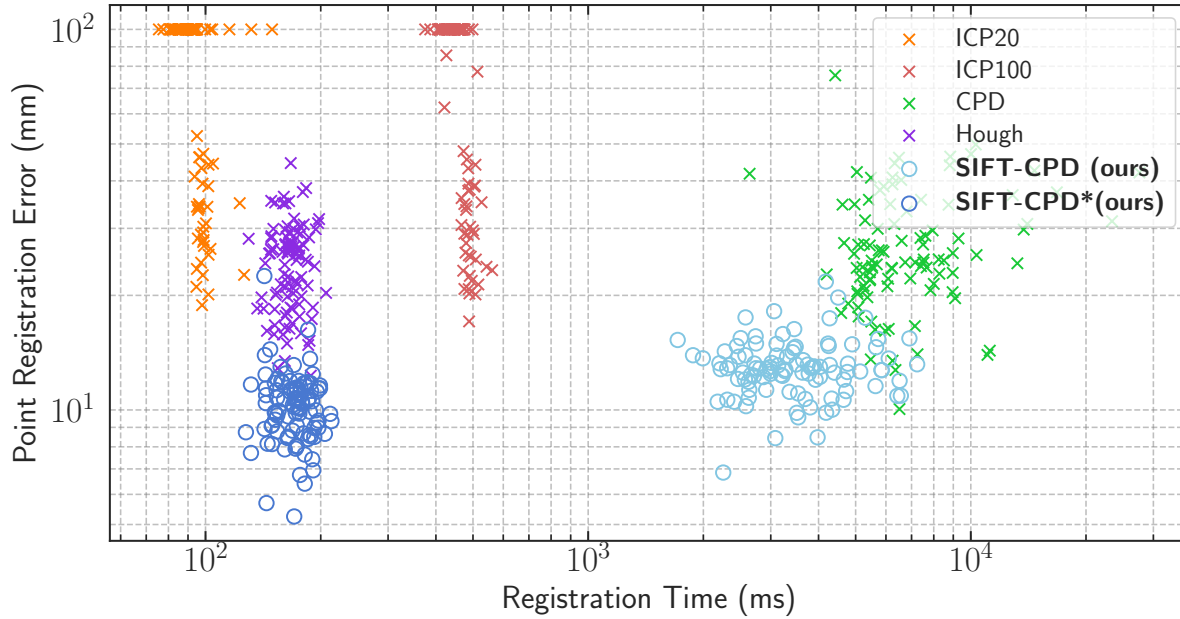


Figure 2.4 Inter-subject registration performance for chest CT images

2.4.2.3 Abnormal Variations and Occlusion

Practical image analysis requires coping robustly with occlusions and a lack of one-to-one correspondence, due to both normal inter-subject variations and abnormal pathological structure such as tumors. Our method is based on local feature properties, and is thus particularly robust to occlusion and missing structure. We used 10 T1w MRI images of different subjects with tumors from the BraTS dataset (Menze *et al.*, 2015) in order to evaluate registration in the presence of occlusion (see Table 2.2 for more details). Nine subject images are randomly selected, and transformed 5 times each, and registered to the remaining subject image. Results are presented in Figure 2.5.

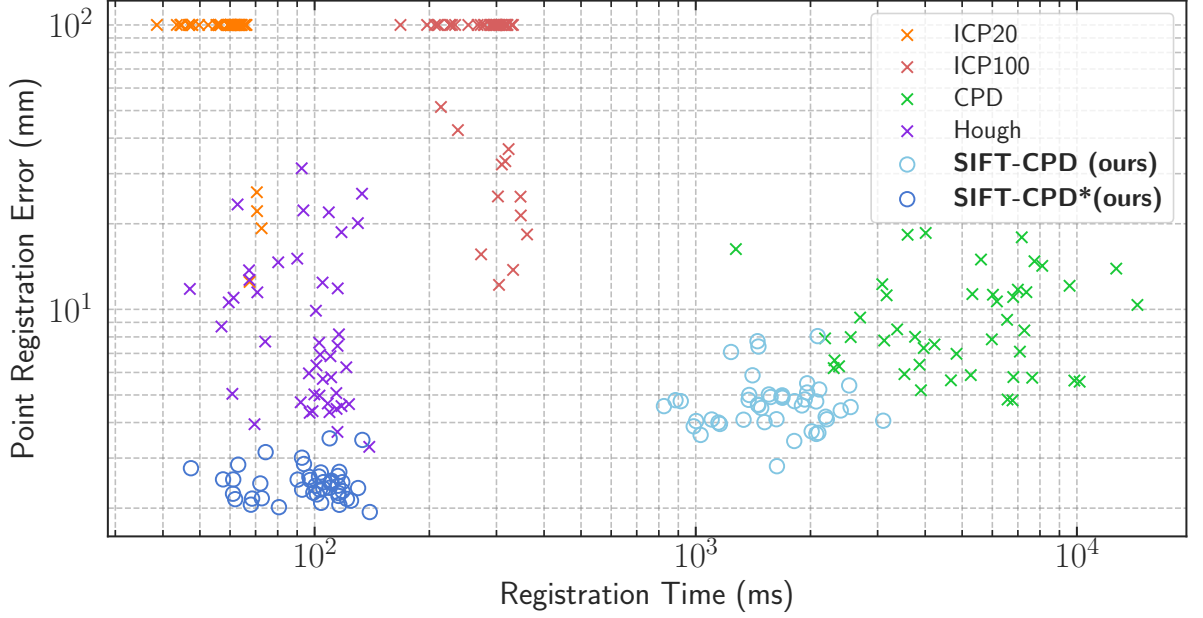


Figure 2.5 Inter-subject registration performance for brains in T1w MRI, with occlusions due to tumors (Menze *et al.*, 2015)

2.4.2.4 Multi-Modal Images

To evaluate the registration performance of our approach on multi-modal images, a random pair of brain images was selected from HCP dataset, one with a T1w image and the other one with a T2w (see Table 2.2 for more details). T1w and T2w images are acquired with different MRI parameters and highlight different tissue properties, where T2w imaging involves longer repetition time (TR) and time-to-echo (TE) parameters. The T2w image was randomly transformed and registered as in previous experiments, with 100 trials. Results are presented in Figure 2.6. Note that without SP-invariant descriptors as proposed in Equation (2.2), multi-modal registration fails completely due to a lack of correspondence. Inlier feature correspondences between different modalities are shown in Figure 2.7, these are investigated in greater detail in the following experiment section.

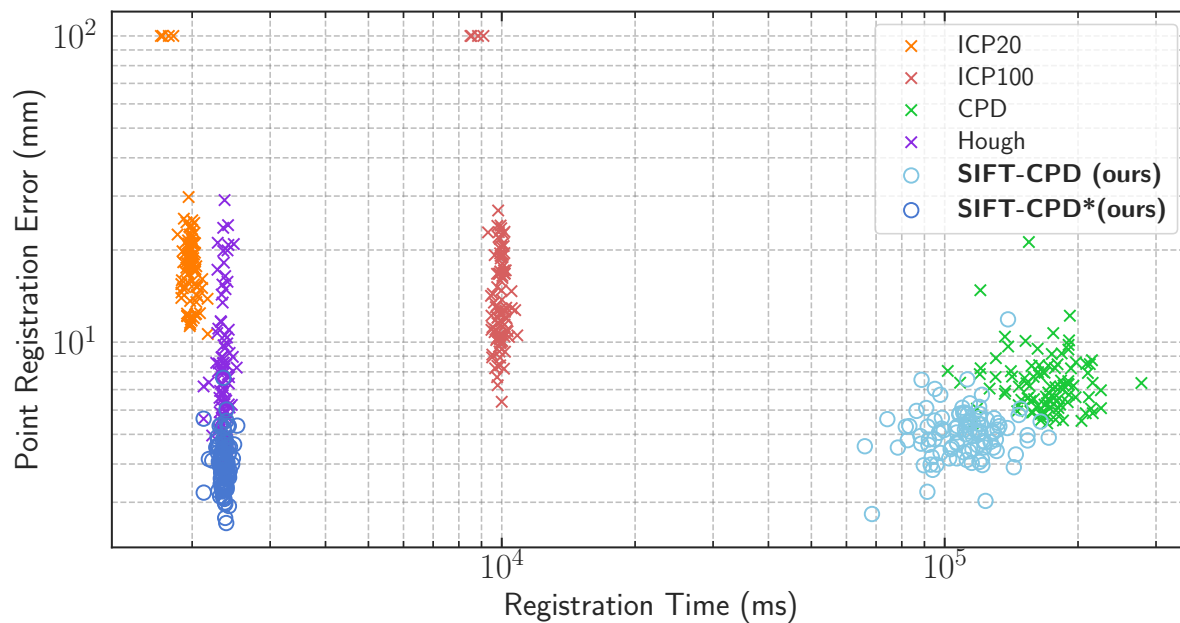


Figure 2.6 Inter-subject registration performance for healthy brain images between T1w and T2w MRI modalities (Van Essen *et al.*, 2012), exhibiting intensity contrast inversion

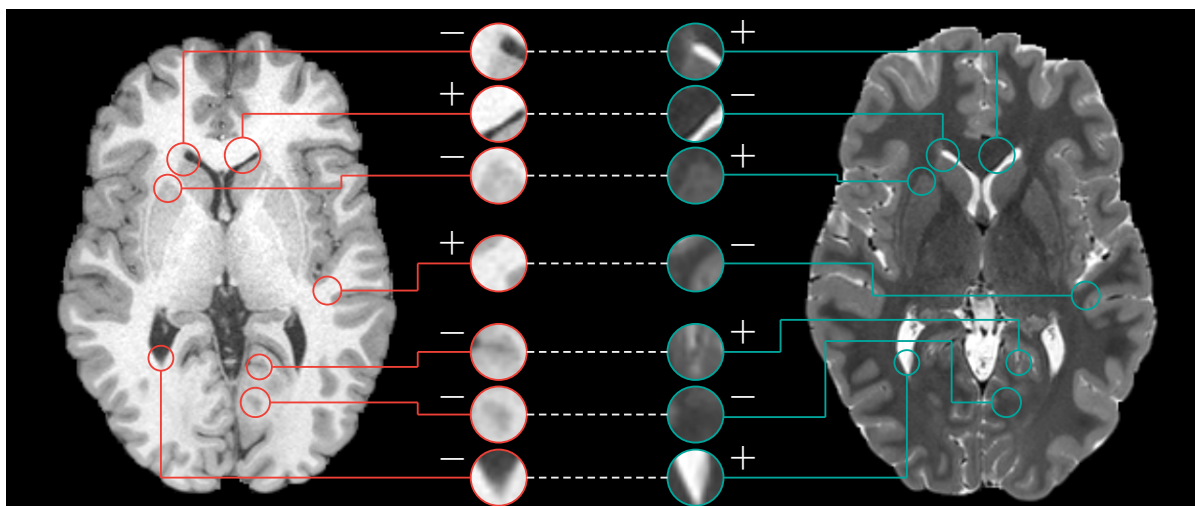


Figure 2.7 Examples of features matched between T1w and T2w images, showing feature signs, note the sign changes do not occur everywhere

2.4.3 Feature Sign and Orientation State

Here we investigated the binary sign s and orientation state (θ_1, θ_2) of local features in registration experiments, using two different methods for estimating feature orientation Θ : eigenvectors of the local gradient structure tensor matrix (Rister *et al.*, 2017) or maximum orthogonal gradient directions (Toews *et al.*, 2013a). Both methods have been demonstrated to achieve effective local feature correspondences between images of the same modality, but not different modalities where discrete orientation state changes may occur.

Figures 2.8 show distributions of orientation state transitions of inlier correspondences between images. For same modality registration (a, b, c) there are relatively few state changes (main diagonal), however in the case of multi-modal registration (d) noticeable differences exist. In the case of orientation determined via principal component analysis (PCA) of gradients (Rister *et al.*, 2017), approximately 40% of all correspondences exhibit single-axis reflections (Figure 2.8 upper d) rather than SP transforms, this is due to eigenvector sign ambiguity, which is accounted for in our algorithm. In the case of orientation determined by maximum gradient directions, correspondences generally exhibit full SP-transforms, i.e. sign inversion and parity (Figure 2.8 lower d). Figure 2.9 shows examples of corresponding features and state transitions in the case of sign inversion, including identity (0-0), single axis reflections of major (0-1) and minor axes (0-2), and parity transforms (0-3).

2.5 Discussion

This paper proposed to extend local feature methods to account for SP-symmetry and to achieved 3D multi-modal image registration. We introduced the notion of discrete SP-symmetry in the of case features extracted from a scalar intensity image $I(\mathbf{x}) \in \mathbb{R}^1$ over $\mathbf{x} \in \mathbb{R}^3$ space. We proposed to model the geometrical properties of 3D image features in a manner invariant to SP-transforms observed in multi-modal image registration. We proposed to include a feature sign $s \in \{-, +\} = \text{sign } \nabla^2 I(\mathbf{x}, \sigma)$ analogous to an electric charge, and maintain a discrete set of four orientation states including reflections of primary axes and parity transforms. Sign is

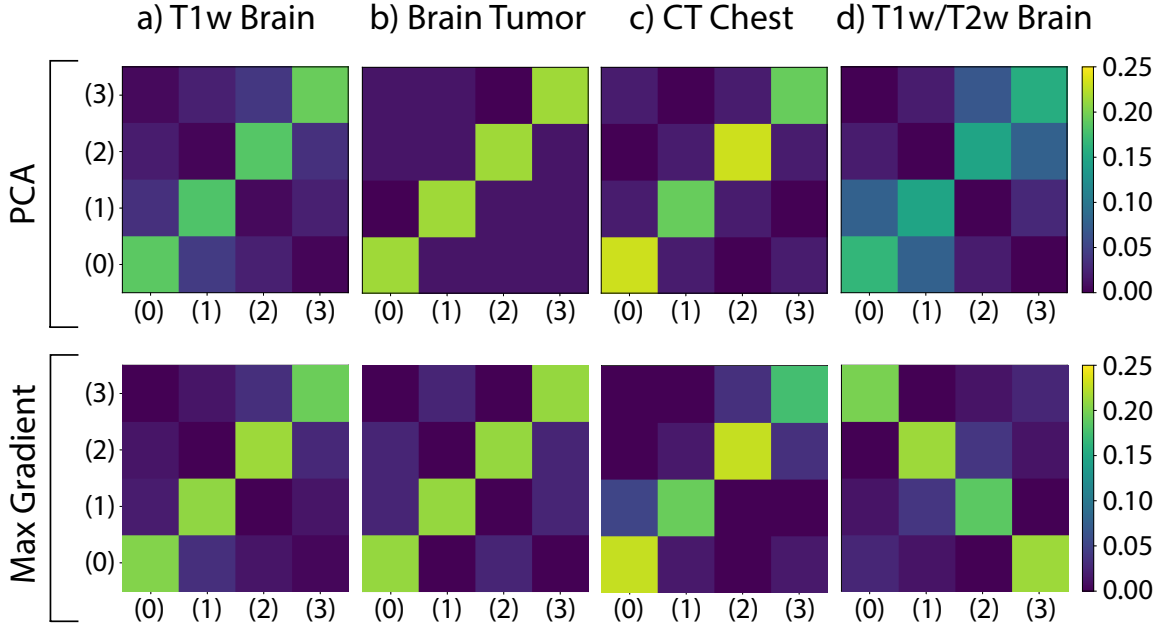


Figure 2.8 Orientation state transition histograms for principal gradient orientation estimation (see (Rister *et al.*, 2017)) and max gradient orientation estimation (see (Toews *et al.*, 2013a)) for: a) T1w brain, b) T1w brain tumor (Menze *et al.*, 2015), c) CT chest (Regan *et al.*, 2011), d) Multi-modal T1w/T2w brain. States described in Figure 4.3, are combinations of discrete axis reflections $\{\pm\hat{\theta}_1, \pm\hat{\theta}_2\}$ as 0: $\{++\}$, 1: $\{+-\}$, 2: $\{-+\}$, 3: $\{--\}$. Bins along the diagonal indicate no state change, off-diagonals indicate single-axis reflections and inverse diagonal indicates parity transform

used to invert gradient descriptor elements in the case of intensity contrast inversion, i.e. charge conjugation. Feature geometry and appearance descriptors are thus invariant to SP transforms, in addition to continuous rotations, scalings and translations as in the 3D SIFT method (Toews *et al.*, 2013a; Rister *et al.*, 2017).

We integrated feature properties into a well-known probabilistic point-based image registration framework, the CPD algorithm (Myronenko & Song, 2010), via a kernel function $\mathcal{K}(f_n, f_m)$, leading to a highly robust and stable SIFT-CPD registration algorithm. Image registration experiments are performed using a range of volumetric image modalities of the human brain and chest, showing that additional properties consistently improve the accuracy of registration in comparison to point locations alone. Registration error and computation time is consistently lower for our proposed SIFT-CPD vs. standard point-based CPD frameworks, indicating that

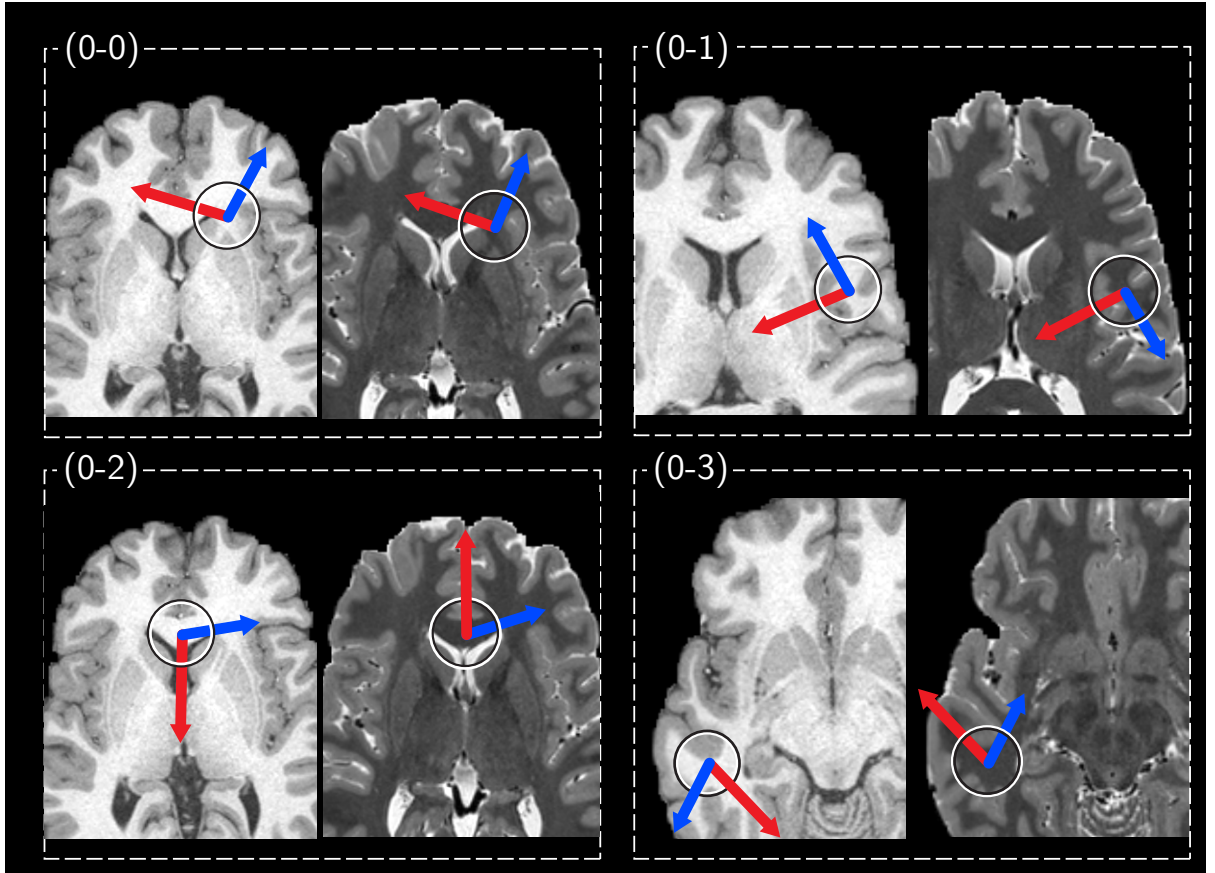


Figure 2.9 Examples of feature orientation state transitions observed between T1w and T2w modalities. Circles indicate feature location $\mathbf{x}$ and scale σ in the plane, sign $s \in \{-1, +1\}$ is indicated by white-black or black-white color. Red and blue arrows indicate the primary $\hat{\theta}_1$ and secondary $\hat{\theta}_2$ orientation axes projected onto the (x, y) image plane

feature scale and orientation improve the estimation of image-to-image transforms. An optimized version of SIFT-CPD functioning solely on sparse inlier feature correspondences exhibits the lowest error, refining the solution achieved via a robust but coarse initial Hough transform. Experiments also demonstrated how SP-symmetry may be applied despite the method used to estimate feature orientation estimation, here principal gradient orientations (Rister *et al.*, 2017) and maximum gradient orientations (Toews *et al.*, 2013a). Invariant feature appearance descriptors are used to identify inliers and to initialize registration, there were not explicitly incorporated into the SIFT-CPD kernel function based solely on geometrical properties here, however this possibility is left for future work.

We note that the theory of SP-symmetry we present applies generally to individual channels of convolutional neural networks, and could be used within the general neural network learning framework to achieve SP-invariance, in addition to continuous invariance (Bardes *et al.*, 2022; Esteves *et al.*, 2018; Spezialetti *et al.*, 2019; Cohen *et al.*, 2019). Here, 3D invariant features have been identified via unbiased, symmetric operators with no explicit training procedure, which could be used as is in general imaging contexts, or to train or validate domain-specific detectors via neural networks (Yi *et al.*, 2016; Detone *et al.*, 2018). All code required to reproduce our results will be available at <https://github.com/3dsift-rank/SIFT-CPD>.

CHAPTER 3

NEUROIMAGE SIGNATURE FROM SALIENT KEYPOINTS IS HIGHLY SPECIFIC TO INDIVIDUALS AND SHARED BY CLOSE RELATIVES

Laurent Chauvin^a, Kuldeep Kumar^b, Christian Wachinger^{c,f}, Marc Vangel^d, Jacques de Guise^a, Christian Desrosiers^a, William Wells III^{e,f}, and Matthew Toews^a for the Alzheimer's Disease Neuroimaging Initiative

^a Department of Systems Engineering, École de Technologie Supérieure, Canada

^b CHU Sainte-Justine Research Center, University of Montreal, Canada

^c Laboratory for Artificial Intelligence in Medical Imaging, University of Munich, Germany

^d Massachusetts General Hospital, Harvard Medical School, USA

^e Brigham and Women's Hospital, Harvard Medical School, USA

^f Computer Science and Artificial Intelligence Lab, Massachusetts Institute of Technology, USA

Paper published in *NeuroImage*, September 2019

Our figure has been selected to make the cover of the journal issue

Abstract

Neuroimaging studies typically adopt a common feature space for all data, which may obscure aspects of neuroanatomy only observable in subsets of a population, e.g. cortical folding patterns unique to individuals or shared by close relatives. Here, we propose to model individual variability using a distinctive keypoint signature: a set of unique, localized patterns, detected automatically in each image by a generic saliency operator. The similarity of an image pair is then quantified by the proportion of keypoints they share using a novel Jaccard-like measure of set overlap. Experiments demonstrate the keypoint method to be highly efficient and accurate, using a set of 7536 T1-weighted MRIs pooled from four public neuroimaging repositories, including twins, non-twin siblings, and 3334 unique subjects. All same-subject image pairs are identified by a similarity threshold despite confounds including aging and neurodegenerative disease progression. Outliers reveal previously unknown data labeling inconsistencies, demonstrating the usefulness of the keypoint signature as a computational tool for curating large neuroimage datasets.

3.1 Introduction

The human brain is a highly complex organ in terms of both structure and function, that is widely studied in vivo through magnetic resonance imaging (MRI) (Lerch *et al.*, 2017). To what degree is neuroanatomy, as observed in MRI, unique to individuals? Can individuals be reliably distinguished from close relatives, i.e. siblings or monozygotic twins sharing 50-100% of their polymorphic genes, despite natural aging, neurodegenerative disease, or noise due to the data measurement process? To what degree are unique aspects of neuroanatomy shared by close relatives? These questions are motivated by increasingly personalized modern medical practices and the need to accurately curate growing sets of clinical and research neuroimaging data. We address them in this paper using a unique computer vision method.

A number of studies have investigated the variability of individuals rather than populations (Valizadeh *et al.*, 2018; Finn *et al.*, 2015; Miranda-Dominguez *et al.*, 2014), where a common theme has been to encode data in terms of a unique neuroimage “fingerprint” or brainprint. Although the specific encodings used are data-dependent, the accuracy with which individuals can be identified based on their brainprint may indicate the degree to which inferences may be drawn from unique, subject-specific observations (Finn *et al.*, 2015). Brainprint investigations have been performed from a variety of MRI modalities, including structural (Wachinger *et al.*, 2015a; Takao *et al.*, 2015), diffusion (Valizadeh *et al.*, 2018; Kumar *et al.*, 2017; Yeh *et al.*, 2016) and functional (Colclough *et al.*, 2017; Finn *et al.*, 2015; Miranda-Dominguez *et al.*, 2014; Chen & Hu, 2018; Amico & Goñi, 2018) MRI, and in non-MRI data such as EEG (Armstrong *et al.*, 2015). Our work here is the first to investigate individual identification from multiple, large-scale public MRI datasets used by the neuroimaging community, including the Human Connectome Project (HCP) (Van Essen *et al.*, 2013), the Alzheimer’s Disease Neuroimaging Initiative (ADNI) (Jack Jr *et al.*, 2008) and the Open Access Series of Imaging Studies (OASIS) (Marcus *et al.*, 2007), and first to report perfect accuracy in individual identification experiments.

The challenge in establishing a distinctive brainprint can perhaps be best illustrated by the convoluted neocortex, arguably the most distinguishing aspect of human neuroanatomy. Cortical folding patterns are highly unique to individuals and generally exhibit higher correlation between twins than unrelated individuals (Van Essen *et al.*, 2016; Thompson *et al.*, 2001), suggesting a link between subtle neuroanatomic structure and shared genetics. A pair-wise image correlation analysis could potentially distinguish individuals and relatives, however such an approach is generally impractical for large datasets as the number of pair-wise operations including image registration is quadratic $N(N - 1)/2$ in the number of images N . Most studies of individual variability have interpreted all data in terms of a standard feature set, e.g. spatially registered voxel-wise measurements (Takao *et al.*, 2015), neuroanatomic segmentations (Wachinger *et al.*, 2015a), cortical parcellations (Colclough *et al.*, 2017; Finn *et al.*, 2015; Miranda-Dominguez *et al.*, 2014; Fischl, 2012) and related volume or thickness measurements (Sabuncu *et al.*, 2016). While standard measurements are invaluable in interpreting data and reducing computational complexity, a number of authors have noted that a one-size-fits-all representation may obscure or average out aspects of anatomy unique to individuals (Finn *et al.*, 2015; Gordon *et al.*, 2017; Chen & Hu, 2018) or close relatives, for example in the case where a one-to-one mapping between images may be ill-defined or nonexistent due to individual variability.

An alternative is to encode the image as a unique set of informative localized features or keypoints, that can be detected efficiently from generic image content and identified robustly when present in different images. For example, the highly successful scale-invariant feature transform (SIFT) (Lowe, 2004) from the field of computer vision uses highly efficient K-nearest neighbor (KNN) keypoint indexing to identify correspondences between generic intensity patterns in large image sets. Inspired by this work, we developed the 3D SIFT-Rank keypoint method (Toews *et al.*, 2013a) for analyzing volumetric image data as illustrated in Figure 3.1. Toews *et al.* (Toews & Wells III, 2016) showed that the proportion of detected keypoints common to a pair of brain MRIs, as quantified by the Jaccard measure of set overlap (Levandowsky & Winter, 1971), could be used to identify MRI pairs of siblings with high reliability. The Jaccard measure defines a similarity matrix $J(A, B)$ between all keypoints and image pairs (A, B)

of a dataset, that can be used in learning-based MRI analysis. Toews et. al. (Toews *et al.*, 2010) identified class-informative keypoint clusters in the similarity matrix $J(A, B)$ for MRI-based disease classification. Kumar et al. (Kumar *et al.*, 2018b) combined similarity matrices derived from multiple MRI modalities in a low-rank manifold embedding in order to study correlations between MRIs of siblings, and found that keypoints outperformed a number of baseline representations including MRI intensities and FreeSurfer derived measures (Volume + Area + Cortical Thickness). While learning procedures may be useful in analyzing a fixed dataset or group analysis, they are difficult to adapt to previously unseen data or classes, i.e. to account for images of previously unobserved individuals. This paper is the first to investigate the task of individual identification using the keypoint representation.

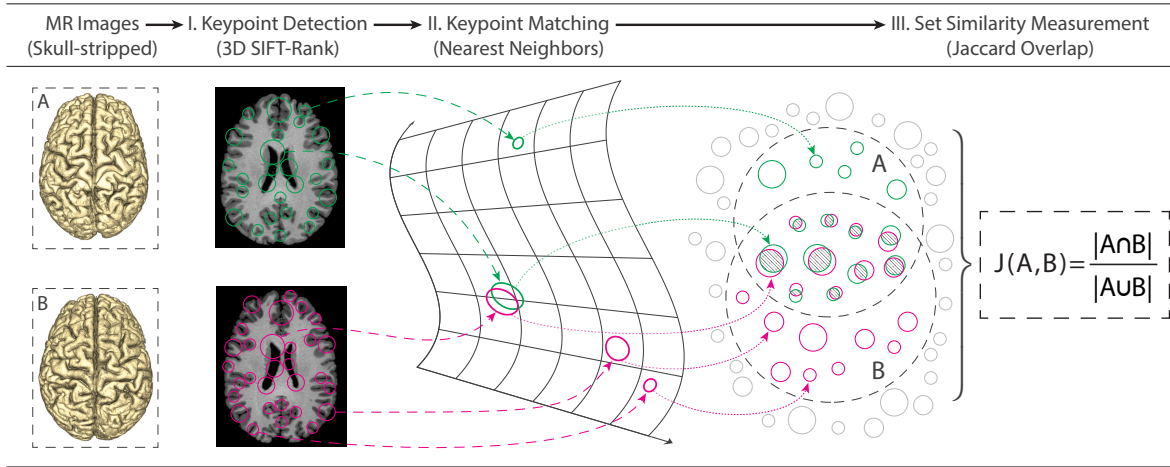


Figure 3.1 The workflow for computing the Jaccard overlap $J(A, B)$ similarity score between two images A and B. **Step I.** SIFT-Rank keypoints are extracted from skull-stripped MRI data. **Step II.** Similar keypoints are identified between images using a K-nearest neighbor search. **Step III.** The Jaccard overlap is computed as ratio of the intersection vs. the union of keypoint sets

These results lead us to hypothesize that SIFT-Rank keypoint sets serve as a highly specific encoding of unique neuroanatomic structure. Here, we propose a novel generalisation of the Jaccard score $J(A, B)$ to account for probabilistic rather than hard set equivalence between keypoints. This leads to a highly efficient instance-based inference model, allowing new MRI data to be compared to a large dataset on-the-fly in order to identify all previous scans of the same

individual despite myriad potential confounds including noise, atrophy due to neurodegeneration. This paper reports the first comprehensive investigation of keypoint signatures for individual identification from MRI data. Our results based on several large-scale, public neuroimaging datasets demonstrated that all same-subject image pairs can be identified by a simple threshold on Jaccard overlap. A visual analysis of outlier cases revealed all were due to data labeling inconsistencies previously unknown to the neuroimaging community, demonstrating the practical importance of the approach in curating increasingly large data collections.

3.2 Material & Methods

3.2.1 Data

Experiments are based on a large, multi-site dataset pooled from 4 public neuroimaging repositories HCP Q4, ADNI 1, OASIS 1 and OASIS 3. The FreeSurfer v6.0 pre-processing pipeline was used to remove non-brain content such as skull from MRIs while preserving cortical and subcortical structures. Out of 8152 images, 616 failed the pre-processing pipeline (FreeSurfer error code, no output image generated) resulting in a final dataset of $N=7536$ MRIs of 3334 unique subjects. Pre-processing failures typically occurred in Talairach alignment or in skull stripping steps, and visual inspection revealed that all cases were due to noticeable image artifacts or noise. Table 3.1 lists demographic and statistical information for this dataset.

Table 3.1 Dataset demographic and statistical information

Dataset	Subjects	Gender (M / F)	Age (Min/Avg/Max)	Images	Voxel Size (mm)	keypoints (Avg/Image)
HCP Q4	1011	469 / 542	22 / 29 / 36	1053	0.7	3633
ADNI 1	844	488 / 356	55 / 75 / 91	3291	1.0	1879
OASIS 1	416	160 / 256	18 / 53 / 96	416	1.0	2143
OASIS 3	1063	470 / 593	42 / 70 / 97	2776	1.0	1856
Total	3334	1587 / 1747	18 / 56 / 97	7536	-	2130

Our analysis is based on a pair-wise comparison of $N(N - 1)/2 = 28,391,880$ possible image pairs. Each pair is assigned a relationship label based on database metadata, for five possible

sibling relationships: same subject (SM), monozygotic twins (MZ), dizygotic twins (DZ), full-sibling (FS) or unrelated subjects (UR). Subjects from different data sets are naïvely assumed to be unrelated, due to a lack of information across databases. It is important to note that these datasets were acquired under different protocols and over different periods of time, resulting in potential bias due to within-dataset similarity. For example, the time interval between scans is under a year for HCP, in comparison to 11 years for OASIS 3, 3 years for ADNI and 2 years for OASIS 1, and pairs of HCP scans may thus exhibit higher similarity than others. Our method is nevertheless robust the ranges of inter- and intra-dataset variability of these data, as we mention later in the discussion.

3.2.2 Processing

Assessing the pairwise similarity of images in a large data set generally requires comparing measurements at spatially homologous locations throughout the images. Since image data are noisy and the precise spatial mapping between images may be unknown or nonexistent, an effective comparison requires a combination of image registration and/or feature extraction methods. Naïve similarity assessment for all image pairs is generally intractable for large datasets as the number of pairwise operations including image registration is quadratic $N(N - 1)/2$ in the number of images N , incurring a computational complexity of $O(N^2)$. Assessment based on a standard feature set such as spatially aligned cortical parcellations can reduce computational complexity, but may be insufficiently specific to capture subtle neuroanatomic patterns only observable in small subsets of a population, e.g. family members.

To address these challenges, we developed a method based on keypoint indexing (Lowe, 2004; Toews *et al.*, 2013a), where the image is represented as a set of generic features detected throughout the image via a saliency operator. Keypoints arise from generic neuroanatomical structure, and can be detected repeatably in a manner invariant to locally linear intensity shifts and global similarity transforms (i.e. 3D rigid transform + isotropic scaling) of image geometry (Toews *et al.*, 2013a). In practice, keypoint extraction is highly robust to variations in MRI intensity

and geometry, e.g. in the case of images acquired from different devices or sites, exhibiting artifacts such as low frequency MRI inhomogeneity effects or changes in patient position. It is also robust to partial occlusions, where locally missing or deformed image content will not significantly impact keypoints identified in other unaffected image regions. Once detected, image content associated with keypoints is encoded into informative descriptors that can be used to identify similar keypoints via highly efficient indexing methods. Specifically, divide-and-conquer algorithms can achieve $O(N \log N)$ computational complexity using search trees to identify sets of similar keypoints, thus sidestepping the need for pairwise image comparisons and scaling gracefully to large datasets. Our method consists of image keypoint detection (Figure 3.1, Step I), keypoint matching (Step II) and finally computation of the Jaccard overlap similarity score (Step III). These three steps are described in greater detail below.

Keypoint Detection seeks to transform each image into a set of salient image keypoints, where keypoints are encoded as informative descriptors for efficient image content indexing. A keypoint is as a spherical image region defined by a centroid $\bar{x} = [x, y, z]$ and a scale (or size) σ within the image, and associated with a descriptor $\bar{f}$ encoding local image appearance. A deterministic two-step detection procedure is used, including 1) salient keypoint localization and 2) keypoint encoding with the so-called SIFT-Rank approach (Toews & Wells, 2009; Toews *et al.*, 2013a). Keypoints are first localized by searching the image for regions that maximize an image saliency operator, signifying informative, local image patterns. A variety of such operators exist, here we use the 3D Laplacian-of-Gaussian (LoG) operator (Marr & Hildreth, 1980) that responds to generic blob-like image patterns, reminiscent of center-surround simple cell retinal processing units in the mammalian visual system (Hubel & Wiesel, 1962). The LoG operator can be approximated efficiently by the difference-of-Gaussian (DoG) operator popularized by the SIFT algorithm in 2004 (Lowe, 2004) (see Equation (3.1)). For each image I , a set of keypoints $\{ (\bar{x}_i, \sigma_i) \}$ is identified as:

$$\{ (\bar{x}_i, \sigma_i) \} = \underset{\bar{x}, \sigma}{\operatorname{argmax}} | I * G(\bar{x}, \sigma) - I * G(\bar{x}, \kappa\sigma) | \quad (3.1)$$

where in Equation (3.1) I is the 3D image, $G(\bar{x}, \sigma)$ is the Gaussian function with isotropic variance σ^2 , and κ is a constant representing the multiplicative difference in scale. Note that the keypoint scale σ is defined by the size of the Gaussian filters for which the DoG saliency operator in Equation (3.1) is maximized.

Once keypoint regions are localized, the image content within each region is rescaled and rotated to a characteristic coordinate system, then encoded as a descriptor $\tilde{f}$ representing the local image content as a fixed-length vector. The rescaling factor is defined by the keypoint scale σ and the rotation by the 3D orientation of local image gradients, thus the descriptor is invariant to global scaled rigid transforms (i.e. similarity transforms) of image geometry, e.g. in the cases of variable patient scanning position or unregistered images. Local image gradient analysis is also used to reject keypoints arising from image patterns that cannot be reliably localized in 3D, e.g. smooth surfaces.

Distinctive image patterns generally arise from boundaries between regions of differing intensity contrast, e.g. white and grey matter. Descriptors here encode image content as histograms of local image gradient information, estimated via finite difference operators and quantized into discrete bins over local 3D location and orientation, an approach known as the histogram-of-oriented gradient (HoG) descriptor (Lowe, 2004; Dalal & Triggs, 2005). Robustness to minor shifts or deformation of image geometry is achieved by the use of relatively coarse spatial bins, where small variations in gradient location or direction do not significantly impact descriptor values. The HoG descriptor is currently among the most effective and widely used descriptors for image keypoint matching, and is analogous of so-called orientation hypercolumns of complex cells that encode image structure in the mammalian visual cortex (Hubel & Wiesel, 1962). In our method, local image patches are cropped and rescaled to 11^3 voxels, after which image gradient histograms are computed to encode the image content. For encoding, local 3D space and orientation are quantized uniformly into $2 \times 2 \times 2 = 8$ spatial regions and 8 orientation histogram bins, thereby producing an $8 \times 8 = 64$ element HoG image descriptor for each keypoint. Finally, this descriptor is rank-ordered to provide invariance to monotonic variations in image gradients (Toews & Wells, 2009), e.g. due to variations in image contrast.

Keypoint Matching seeks to identify pairs of keypoints arising from similar anatomical structure in different images, based on the similarity (or dissimilarity) of their descriptors. Descriptors from the same structure in different images are rarely identical, but differ by varying amounts due to imaging variations, noise, etc. The dissimilarity of a pair of descriptors $(\bar{f}_i, \bar{f}_j)$ is quantified by the Euclidean distance or $L2$ -norm between their elements $d(\bar{f}_i, \bar{f}_j) = \|\bar{f}_i - \bar{f}_j\|_2$. Assuming an additive Gaussian noise model and independent and identically distributed (IID) descriptor elements, the smaller the distance $d(\bar{f}_i, \bar{f}_j)$, the higher the likelihood that the descriptors arise from the same underlying anatomical structure.

A K-nearest neighbor (K-NN) search is used to identify sets of similar descriptors as follows. For each descriptor $\bar{f}_i$, a set of the K^{th} closest or most similar descriptors $NN_k(\bar{f}_i)$ is identified as

$$NN_k(\bar{f}_i) = \{\bar{f}_j : d(\bar{f}_i, \bar{f}_j) \leq d_{i,K}\} \quad (3.2)$$

where $d_{i,K} = d(\bar{f}_i, \bar{f}_K)$ is the distance between $\bar{f}_i$ and the K^{th} closest descriptor $\bar{f}_K$. Enumerating $NN_k(\bar{f}_i)$ for each $\bar{f}_i$ in a set of N descriptors via naïve pairwise comparisons incurs a computational cost of $O(N^2)$. However rapid approximate K-NN algorithms using efficient tree-based search structures can perform this enumeration in $O(N \log N)$ computational complexity. We use a set 8 randomized search trees as proposed in (Muja & Lowe, 2014), where the descriptor search space is divided according to the descriptor elements exhibiting the highest variance. Parameter K can be set generously to at least the number of expected matches in the dataset, here we used $K = 30$. Note that a relatively large percentage of matches may be spurious or incorrect due to noise, however these tend to be distributed randomly between unrelated subjects and have negligible impact. With these parameters, matching features of one image to all other 7535 images (represented by approximately 16,000,000 features) requires approximately 0.35 seconds on an Intel Xeon Silver 4110@2.10Ghz, demonstrating the high computational efficiency of the method. Note that $O(N^2)$ operations are required to explicitly enumerate matches between all pairs in a set of N images for a complete analysis (i.e. 28,381,485 unrelated subjects in Figure 3.2), however small numbers of closely related subjects may identified in $O(N \log N)$ complexity.

Set Similarity Measurement seeks to quantify the likelihood that two independent keypoint sets arise from the same underlying object. As our data consist of sets of discrete keypoints, potentially arising from equivalent neuroanatomic regions in different subject scans, we use the Jaccard overlap measure from set theory. The Jaccard overlap score $J(A, B)$ between a pair of feature sets A and B is defined as

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|} \quad (3.3)$$

where, in Equation (3.3), $|A|$ and $|B|$ represent the cardinalities (sizes) of sets A and B , and $|A \cap B|$ represents the cardinality of their intersection, i.e. the set of elements shared by A and B . The Jaccard measure has several desirable properties, for example it ranges from $[0,1]$ for disjoint to identical sets respectively, and can be transformed into a distance metric (Levandowsky & Winter, 1971) $1 - J(A, B)$ in the abstract space of variable-sized keypoints sets.

The intersection in Equation (3.3) is defined as the subset of elements shared by both sets:

$$A \cap B = \{\bar{f}_i : \bar{f}_i \in A \wedge \bar{f}_i \in B\} \quad (3.4)$$

Evaluating the intersection requires identifying pairs of equivalent set elements, i.e. $\bar{f}_i \in A$ and $\bar{f}_j \in B$ such that $\bar{f}_i = \bar{f}_j$. These are defined by the union of all unique NN keypoint matches identified between A and B . As set equivalence is binary, each unique keypoint match thus $(\bar{f}_i, \bar{f}_j)$ contributes 1 element to the intersection $A \cap B$, and the cardinality $|A \cap B|$ may be computed as the sum of matching elements (Kumar *et al.*, 2018b; Toews & Wells III, 2016). Binary equivalence of keypoint descriptors is difficult to justify however, as statistical variations in image content generally introduce uncertainty into keypoint descriptors. We thus consider $(\bar{f}_i, \bar{f}_j)$ as contributing a soft value ranging from $[0, 1]$ to $|A \cap B|$, which is proportional to the likelihood of $(\bar{f}_i, \bar{f}_j)$ arising from a Gaussian density of variance α_i^2 and mean $\bar{f}_i$. The cardinality of the set intersection is then evaluated as:

$$|A \cap B| = \sum_{\bar{f}_i \in A} \max_{\bar{f}_j \in NN_k(\bar{f}_i) \cap B} \exp\{-d^2(\bar{f}_i, \bar{f}_j)/2\alpha_i^2\} \quad (3.5)$$

In Equation (3.5), variance parameter α_i^2 is set automatically for each keypoint descriptor $\tilde{f}_i$ as the squared distance to the closest NN keypoint within the entire training set. This allows the method to adjust to variable sampling density in the keypoint descriptor space about sample $\tilde{f}_i$ in a manner similar to adaptive or variable kernel density estimation (Terrell & Scott, 1992), thereby downweighting the contribution of unlikely matches to the cardinality $|A \cap B|$. A related benefit is the reduced sensitivity to the parameter K in the evaluation of $|A \cap B|$. Note that the cardinality for standard binary set equivalence is computed via Equation (3.5) by taking $d(\tilde{f}_i, \tilde{f}_j) = 0$ for all $\tilde{f}_j \in NN_k(\tilde{f}_i)$. Given that the nearest neighbor relationship is not strictly symmetric, the cardinality as computed via Equation (3.5) depends generally on the feature set over which sum is computed. In practice, we notice little difference in Jaccard values, and symmetry may be imposed.

3.3 Results

We performed experiments to quantify the variability of T1-weighted MRI data acquired from individuals and close relatives using keypoint signatures. A large set ($N=7536$) of T1w MRIs of 3334 unique subjects was pooled from four public neuroimaging datasets (HCP Q4, ADNI 1, OASIS 1 and OASIS 3), where each image pair bears a unique pair-wise relationship label: same subject (SM), monozygotic twin (MZ), dizygotic twin (DZ), non-twin full sibling (FS) or unrelated (UR) subjects. Relationship information for pairs of MR images were provided by individual datasets, while relationships between image pairs from different datasets were naïvely assumed to be unrelated. We evaluated the Jaccard overlap derived from the image content for all $N(N - 1)/2 = 28,391,880$ possible image pairs, and studied the distributions of Jaccard measurements conditioned on relationship labels. Since the Jaccard overlap was derived from the proportion of features shared between images, we expected it to decrease with the degree of genetic and environmental separation in the pairwise relationship, i.e. decreasing in the order of SM, MZ, DZ, FS, UR pairs.

Figure 3.1 illustrates the workflow for evaluating the Jaccard similarity score $J(A, B)$ between an image pair (A, B) . First, a one-time pre-processing step was applied to each image, where

non-brain structure was removed using the Freesurfer software (Fischl, 2012), after which keypoint features were detected using the authors' publicly available software implementation. Detection required approximately 15 seconds and identified approximately 2000 keypoints per MRI. After pre-processing, nearest-neighbor (NN) keypoint matches were enumerated across all images, establishing putative equivalence between keypoints in different images, and the Jaccard similarity score was computed for each image pair based on the proportions of keypoint matches they shared.

The Jaccard overlap $J(A, B)$ can be viewed as a whole-brain similarity measure ranging from $[0, 1]$ for lowest to highest similarity. Equivalently, a monotonic transform such as the negative logarithm can be used to map $J(A, B)$ to a Jaccard distance measure $d_J(A, B) = -\log J(A, B)$ ranging from $[0, \infty]$. Figure 3.2 shows the empirical distributions of Jaccard distances, obtained for the five pair-wise relationships (indicated by color), where lower distance indicates a higher proportion of shared image content and neuroanatomic similarity. Distributions for each relationship label are unimodal and concentrated about a central tendency. The order of the mean distances for relationships (dashed vertical lines) is generally consistent with the degree of similarity in genetic and environmental developmental factors, i.e. increasing in order of SM, MZ, DZ/FS, UR. A number of outliers were identified (red and blue dots), and have been confirmed to arise from inconsistent image labels. These outliers will be discussed later.

The Jaccard distance distribution for same-subject (SM) pairs was highly unique, with no overlap with any other distribution including monozygotic twins (MZ). All SM pairs could thus be identified via a simple score threshold, supporting our hypothesis that keypoints capture highly unique aspects of individual neuroanatomy. Distributions for other relationships exhibit a degree of overlap, and the two-sample Kolmogorov-Smirnov test was used to evaluate the null hypothesis that samples arise from the same underlying distribution (Table 3.2). Most p-values were extremely low, allowing us to reject the null hypothesis with high confidence. The one exception was the case of FS and DZ pairs, where the p-value of 0.108 indicated no significant difference between Jaccard distance distributions for FS and DZ siblings.

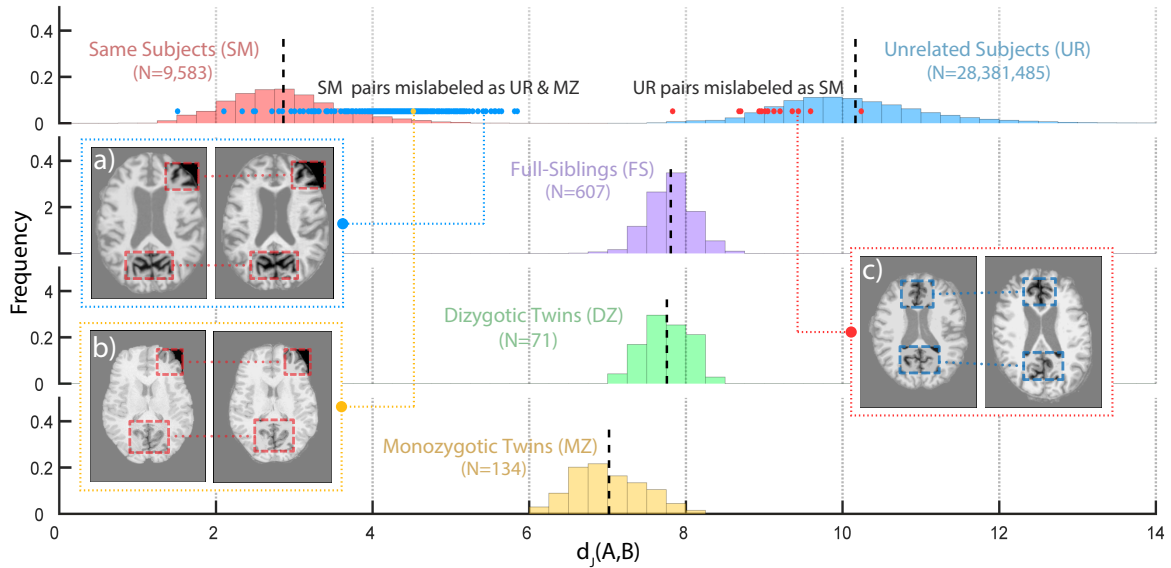


Figure 3.2 Distributions of the pairwise Jaccard distances conditional on relationship labels including: unrelated (UR, blue), same (SM, red), full siblings (FS, purple), dizygotic (DZ, green) and monozygotic (MZ, yellow) twin subject relationship labels, where black dashed lines indicate distribution means. Note the high degree of separation between SM and UR scores. Dots indicate data labeling inconsistencies automatically flagged by unexpected Jaccard distance. The correct relationship labels are evident upon visual inspection of cortical patterns (highlight), which are virtually identical in SM pairs mislabeled as (a) UR and (b) MZ, and highly different in a UR pair mislabelled as (c) SM

Table 3.2 p-values for two-sample Kolmogorov-Smirnov tests between Jaccard distance distributions for (SM) Same Subject, (MZ) Monozygotic, (DZ) Dizygotic, (FS) Full-Sibling and (UR) Unrelated pairwise relationships

	SM	MZ	DZ	FS	UR
SM	-	1.38×10^{-233}	1.55×10^{-125}	0	0
MZ	-	-	1.40×10^{-41}	6.09×10^{-97}	1.18×10^{-228}
DZ	-	-	-	0.108	2.87×10^{-120}
FS	-	-	-	-	0
UR	-	-	-	-	-

The Jaccard overlap is derived from image-to-image keypoint matches between unique neuroanatomic patterns identified in different images. Figure 3.3 illustrates the spatial distributions of matching keypoints as heatmaps within a common reference space for each relationship label.

Keypoints were distributed similarly throughout the brain for all relationships, and generally concentrated in regions with significant intensity contrast variations, i.e. the interfaces between cortical sulci or sub-cortical structures and cerebral spinal fluid (CSF). The primary quantitative difference between relationship groups was the number of keypoint matches, which was much higher for same subject images, thus reflecting a higher degree of shared anatomic structure. Figure 3.4 shows an example of keypoints matching between MRIs of the same individual acquired 11 years apart.

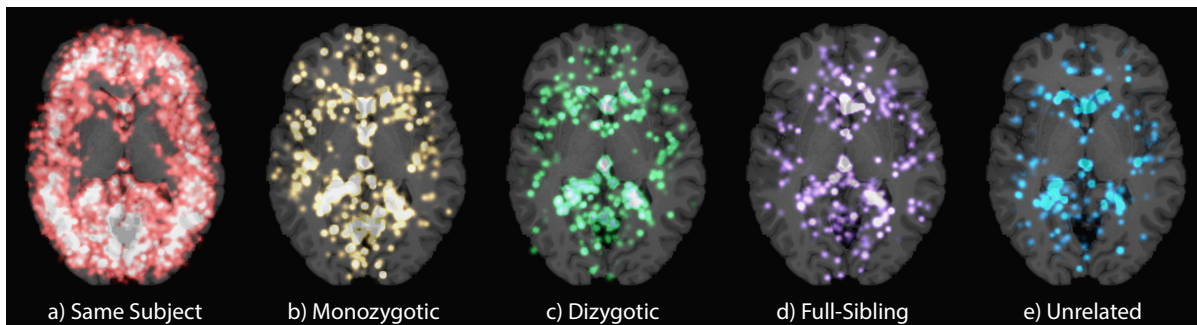


Figure 3.3 The spatial layout of keypoint matches for five pairwise relationship labels: a) SM, b) MZ, c) DZ, d) FS, e) UR. Heatmaps represent distributions of matching keypoints within the standard MNI305 neuroanatomic reference space and accumulated over 71 image pairs per label from HCP dataset

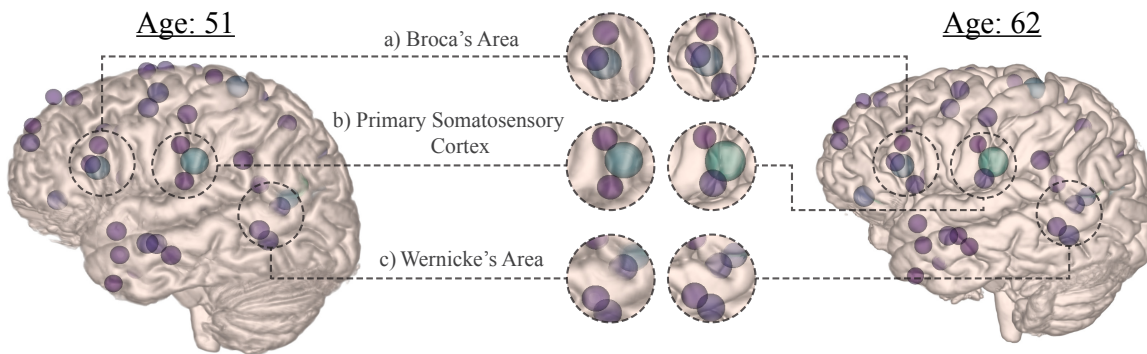


Figure 3.4 Keypoints matched between MRIs of the same individual acquired at 51 and 62 years of age. Matching keypoints (spheres, color indicating scale) represent patterns of local cortical structure that are highly unique to this individual, with notable concentrations in a) Broca's area, b) the primary somatosensory cortex and c) Wernicke's area. Note that keypoints have been slightly extruded from within the cortex for improved visualization.

The visualization was generated using the 3D Slicer software ([Fedorov et al., 2012](#))

The primary systematic confound was the age difference between image acquisitions, which was positively correlated with Jaccard distance as shown in Figure 3.5. This likely reflected changes in brain morphology due to both natural aging and disease progression, primarily in SM pairs of older adults from the ADNI and OASIS datasets designed to study Alzheimer’s disease (mean age 73 years). Data was unavailable to investigate the impact of age difference for younger SM subjects, however age separation between FS pairs (mean age 28 years) from the HCP dataset had no noticeable impact on the Jaccard distance, likely reflecting the relatively stable brain anatomy across the younger, healthy HCP cohort. By inspection, the highest Jaccard distances for SM pairs were typically associated with random confounds including image artifacts (e.g. due to MRI acquisition, pre-processing, etc.) or morphological changes in the brain (e.g. due to natural aging, neurodegenerative disease, etc.), see Figure 3.5 a) and b).

Other confounds including sex, race and age had no significant impact on Jaccard distances, and invariant keypoint matching is independent of the image (mis)alignment, an important benefit of our method.

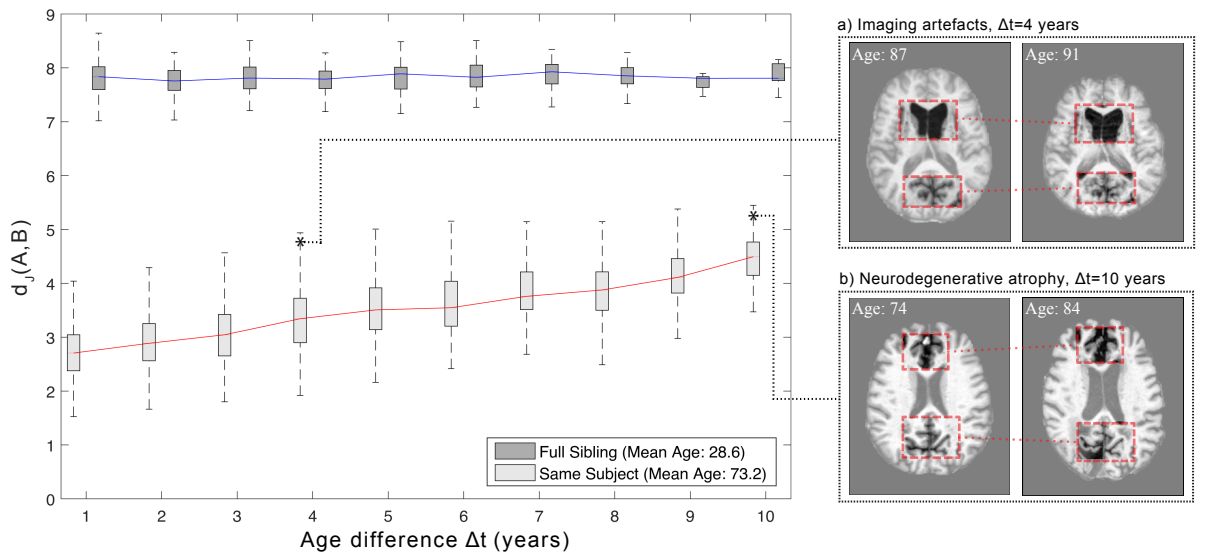


Figure 3.5 Jaccard distance as a function of the age difference Δt between scans for SM ($N=9885$) and FS ($N=607$) pairs. Lines and boxes represent the means and standard deviations of the Jaccard distance. a) and b) show examples of image pairs successfully identified as SM, despite a relatively high Jaccard distance due to aging, noticeable image artefacts in a) and neurodegenerative atrophy in b)

A surprising result was the discovery of 184 outlier image pairs with Jaccard distances that were noticeably outside of the distributions associated with their relationship labels. These included pairs labeled as UR or MZ with Jaccard distances similar to same subjects (Figure 3.2, blue dots), and pairs labeled as SM with Jaccard distances similar to those for unrelated subjects (Figure 3.2, red dots). Upon visual inspection of the cortical folding patterns (see examples in Figure 3.2), and with the help of the respective database administrators, we established that all were most likely labeled incorrectly. All individual datasets contained at least one case where images of the same subject were labelled as different subjects. Instances of the same subject were identified across the OASIS and ADNI datasets. Table 3.3 lists the numbers of subjects with inconsistent labels within and between datasets. Additionally, 11 cases of perfectly identical images were identified by unusually low Jaccard distance, these were labeled as images of the same subject acquired at different time points.

An important algorithmic parameter is the number of NN descriptor matches K used in estimating the Jaccard score. In the case of hard set equivalence (Toews & Wells III, 2016; Kumar *et al.*, 2018b), each keypoint match contributes 1 to the set union $|A \cup B|$, and an optimal K depends on the number of relevant image samples, i.e. MRIs of the same subject or group, which is generally unknown and variable. In the case of our soft weighting approach however, each match contributes a weight proportional to a Gaussian density, and thus K may be set large enough to include all relevant image samples.

3.4 Discussion

In this paper, we proposed to model neuroanatomy as a collection of distinctive image keypoints, hypothesizing that this would more accurately preserve aspects of anatomy unique to individuals or close family members, distinctive neuroanatomic signatures that might otherwise be averaged out by traditional parcellation or voxel-wise representations.

The whole-brain similarity of an image pair was assessed in terms of the proportion of keypoints they share using the Jaccard measure of set overlap, which can be computed from arbitrarily large datasets using a highly efficient keypoint matching procedure.

Table 3.3 Misabeled subject relationships identified across and within databases. Most cases are subjects mislabeled as UR

* A pair labeled as MZ was established to be SM and one dataset was subsequently removed from the HCP dataset, and a pair of scans labeled as UR were confirmed to be SM from the OASIS 1 dataset

** OASIS 3 and OASIS 1 are known to share data from numerous subjects

*** OASIS 3 contained 2 pairs of SM subjects mislabeled as UR, and 3 pairs of UR subjects labeled as SM

	HCP Q4	ADNI 1	OASIS 1	OASIS 3
HCP Q4	1*	0	0	0
ADNI 1	-	3	4	2
OASIS 1	-	-	1*	79**
OASIS 3	-	-	-	2 + 3***

Experiments validated our hypothesis in the largest study of individual identification from MRI data to date, involving 7536 T1w MRIs of 3334 unique subjects pooled from four large, public neuroimaging datasets: ADNI, OASIS1, OASIS3 and HCP. Distributions of Jaccard distances for same vs. unrelated subject MRI pairs are separated by a wide margin, and a simple threshold on the Jaccard distance was sufficient to identify all same-subject pairs with 100% accuracy¹. In contrast, the largest previous study involved almost 700 subjects from the ADNI dataset alone, required multiple scans per subject as training data, and achieved less than perfect accuracy using features derived from a standard neuroanatomic segmentation (Wachinger *et al.*, 2015a).

An important potential confound is within-dataset scan similarity; same-subject scans are typically found within the same dataset, whereas scans from the same dataset are known to generally exhibit similarity due to a number of commonalities including subject demographics, age, site, scan sequence, scanner artifacts, etc (Wachinger *et al.*, 2019). As expected, Jaccard distance distributions were generally lower within-dataset vs. across-datasets: for SM pairs (2.12 ± 0.65 vs. 4.46 ± 0.66) and for UR pairs (8.99 ± 0.76 vs. 10.01 ± 0.80), indicating a within-dataset similarity effect. However there was no overlap between distance distributions for

¹ For completeness, there is the possibility of coincidental errors in both metadata labeling and automatic identification, however the probability of this is very low, given these events are unrelated, independent and individually highly improbable.

SM and UR pairs, and we thus expect these to remain separable by a Jaccard distance threshold in new T1-weighted MRI data acquired and pre-processed (e.g. skull-stripped) with protocols similar to those used for the three datasets used here. Furthermore, the pattern of decreasing similarity in the order of SM, MZ, (DZ/FS) to UR pairs did not change when analysis was restricted to single datasets.

A surprising result was the discovery of previously unknown subject labeling inconsistencies, identified as clear outliers from their expected Jaccard distributions. These included MRIs of the same person labeled as unrelated, or MRIs of different people labelled as the same, and were identified both within and across the datasets. Such inconsistencies may lead to bias in cross-validation studies, e.g. computer-assisted prediction ([Desikan *et al.*, 2009](#); [Samper-González *et al.*, 2018](#)) where protocols assume independent training and testing data, and similar errors in a clinical context could potentially lead to errors in patient care. The ability to identify these in widely used, public datasets is noteworthy, and demonstrates the potential for the keypoint signature as a powerful tool in curating and validating large neuroimage datasets.

The Jaccard measure is driven by keypoint matches representing instances of unique neuroanatomic patterns shared between pairs of images of individual or siblings. The Jaccard overlap generally decreased with the degree of genetic and developmental separation in the pairwise relationship label, i.e. in the order of SM, MZ, DZ/FS, UR pairs, indicating decreasing proportions of unique anatomic structure shared by these groups as predicted. A notable exception was the case of dizygotic twin (DZ) and non-twin full sibling (FS) relationships, which showed no statistically significant difference in terms of their pairwise Jaccard distributions. Keypoint matches were generally distributed throughout the brain and across interfaces between adjacent tissues exhibiting high intensity contrast in MRI, e.g. cerebrospinal fluid, grey and white matter, in a manner unique to the specific image pair, rather than within regular loci defined by typical parcellation schemes. Combined with the high accuracy of identification experiments, this suggests that aspects of neuroanatomy most characteristic of individuals or close relatives may be highly idiosyncratic and not ultimately be tied to a fixed parcellation. Keypoint signatures provide a robust and efficient means of exploring these aspects across large

datasets, and a combination of keypoint signatures and traditional segmentations or parcellations may ultimately prove most effective in understanding the variability of individuals and genetic families.

In terms of technology, the keypoint signature affords the capability of rapidly comparing a new image against a large dataset, e.g. identifying all keypoint sets with high Jaccard similarity in 0.35 seconds here. This is a memory-based learning approach, which requires no explicit training procedure. New data are easily incorporated, and it is limited only by the amount of memory available, a limit that is continually reduced by technological advancement. In contrast, representations based on traditional neuroanatomic parcellations generally require extensive pre-processing, including image registration and segmentation, and alternative machine learning approaches require training procedures with multiple MRIs per subject ([Wachinger *et al.*, 2015a](#)) which may be unavailable a priori. Keypoint detection here is based on a recursive Gaussian filtering process that is analogous to a highly efficient, unbiased convolutional neural network (CNN) used in deep learning ([LeCun *et al.*, 2015](#)). Machine learning could potentially be used to optimize filters, however this would require training procedures and data that might not be readily available (i.e. multiple labeled images of an individual), and would introduce bias related to a particular training set.

Our experiments here derived keypoint signatures from the ubiquitous T1w structural MRI modality, however keypoint detection can be performed from arbitrary scalar image modalities, e.g. fractional anisotropy derived from diffusion MRI (dMRI) ([Kumar *et al.*, 2018b](#)) or statistical parameter maps derived from fMRI data, and descriptors can be used to encode vector-valued data, e.g. histograms of diffusion gradient orientations in dMRI ([Chauvin *et al.*, 2018](#)). Keypoint matching across different modalities is generally non-trivial and an avenue for future investigation ([Toews *et al.*, 2013a](#)). Our analysis focused on neuroanatomy and automatic skull stripping was used to remove image content associated with non-brain tissues. Extensive pre-processing is not generally required, as keypoints can be reliably detected despite variations in intensity or pose, and can be used to analyze non-brain image content. In fact, we found the Jaccard similarity between related subjects to be higher when non-brain structure is

included. Nevertheless, processes for normalizing or correcting intensity values, e.g. correction of MRI inhomogeneities using field maps, should generally improve the repeatability of keypoint detection, and the question of an optimal image pre-processing pipeline is left for future research. Experiments here were limited to sibling relationships and adult subjects ranging of 18-96 years of age, future investigations will consider younger age groups such as infants with the additional confound of rapid neurodevelopment and other relationships including parent-child or cousins with varying amounts of shared genetics. Finally, our work here does not investigate links between anatomical keypoints and subject abilities or behaviors. However, the keypoint representation has previously been used to interpret single anatomical scans according to group-wise clinical symptoms or labels, e.g. Alzheimer’s disease classification (Toews *et al.*, 2010) and neonatal age prediction (Toews *et al.*, 2012), and similar keypoint analysis techniques could potentially be applied to other modalities including functional MRI data in future investigations.

3.4.1 Code and Data Availability

Complete C++ source code for the method is provided by the authors at <https://github.com/3dsift-rank/3DSIFT-Rank>. Keypoint data are available at <https://central.xnat.org/data/projects/SIFTFeatures>.

3.5 Acknowledgements

OASIS. Data were provided in part by OASIS: Cross-Sectional: Principal Investigators: D. Marcus, R. Buckner, J. Csernansky J. Morris; P50 AG05681, P01 AG03991, P01 AG026276, R01 AG021910, P20 MH071616, U24 RR021382 and OASIS-3: Principal Investigators: T. Benzinger, D. Marcus, J. Morris; NIH P50AG00561, P30NS09857781, P01AG026276, P01AG003991, R01AG043434, UL1TR000448, R01EB009352. AV-45 doses were provided by Avid Radiopharmaceuticals, a wholly owned subsidiary of Eli Lilly.

HCP. Data were provided in part by the Human Connectome Project, WU-Minn Consortium (Principal Investigators: David Van Essen and Kamil Ugurbil; 1U54MH091657) funded by the

16 NIH Institutes and Centers that support the NIH Blueprint for Neuroscience Research; and by the McDonnell Center for Systems Neuroscience at Washington University.

ADNI. Data collection and sharing for this project was funded by the Alzheimer’s Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer’s Association; Alzheimer’s Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer’s Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

This work was supported by NIH grant P41EB015902, a Canadian National Sciences and Research Council (NSERC) Discovery Grant and the Canada Research Chair in 3D Imaging and Biomedical Engineering.

CHAPTER 4

EFFICIENT PAIRWISE NEUROIMAGE ANALYSIS USING THE SOFT JACCARD INDEX AND 3D KEYPOINT SETS

Laurent Chauvin^a, Kuldeep Kumar^b, Christian Desrosiers^a, William Wells III^{c,d}, and Matthew Toews^a

^a Département de Génie des Systemes, École de Technologie Supérieure, Canada

^b CHU Sainte-Justine Research Center, University of Montreal, Canada

^c Brigham and Women's Hospital, Harvard Medical School, USA

^d Computer Science and Artificial Intelligence Lab, Massachusetts Institute of Technology, USA

Paper published in *IEEE Transactions on Medical Imaging*, September 2021

Abstract

We propose a novel pairwise distance measure between image keypoint sets, for the purpose of large-scale medical image indexing. Our measure generalizes the Jaccard index to account for soft set equivalence (SSE) between keypoint elements, via an adaptive kernel framework modeling uncertainty in keypoint appearance and geometry. A new kernel is proposed to quantify the variability of keypoint geometry in location and scale. Our distance measure may be estimated between $O(N^2)$ image pairs in $O(N \log N)$ operations via keypoint indexing. Experiments report the first results for the task of predicting family relationships from medical images, using 1010 T1-weighted MRI brain volumes of 434 families including monozygotic and dizygotic twins, siblings and half-siblings sharing 100%-25% of their polymorphic genes. Soft set equivalence and the keypoint geometry kernel improve upon standard hard set equivalence (HSE) and appearance kernels alone in predicting family relationships. Monozygotic twin identification is near 100%, and three subjects with uncertain genotyping are automatically paired with their self-reported families, the first reported practical application of image-based family identification. Our distance measure can also be used to predict group categories, sex is predicted with an AUC=0.97. Software is provided for efficient fine-grained curation of large, generic image datasets.

4.1 Introduction

Health treatment is increasingly personalized, where treatment decisions are conditioned on patient-specific information in addition to knowledge regarding the general population (Hamburg & Collins, 2010). Modern genetic testing allows us, based on large libraries of human DNA samples, to cheaply predict patient-specific characteristics, including immediate family relationships, and also characteristics shared across the population including racial ancestry, sex, hereditary disease status, etc. (Annas & Elias, 2014; Auton *et al.*, 2015). The brain is the center of cognition and a complex organ, tightly coupled to the genetic evolution of animals and in particular that of the human species. To what degree is the human brain phenotype coupled to the underlying genotype? How does the brain image manifold vary locally with genotype, i.e. immediate relatives sharing 25-100% of their polymorphic genes, or between broad groups defined by subtle genetic factors such as racial ancestry or sex?

Large, publicly available databases allow us to investigate these questions from aggregated MRI and genetic data, e.g. 1000+ subjects (Van Essen *et al.*, 2012). The shape of the brain has been modeled as lying on a smooth manifold in high dimensional MRI data space (Gerber *et al.*, 2010; Brosch *et al.*, 2013), where phenotype can be described as a smooth deformation conditioned on developmental factors including the environment. However the brain is also naturally described as a rich collection of spatially localized neuroanatomical structures, including common structures such as the basal ganglia shared across the population, but also highly specific patterns such as cortical folds that may only be observable in specific individuals or close family members (Ono *et al.*, 1990).

The keypoint representation is an intuitively appealing means of modeling specific, localized phenomena, i.e. a set of descriptors automatically identified at salient image locations as shown in Figure 4.1. A keypoint set can be viewed as an element of a high-dimensional manifold, and medical imaging applications such as regression or classification can be formulated in terms of a suitable geodesic distance between sets. As keypoint sets are variable sized, typical metrics based on fixed-length vectors such as L-norms (Gerber *et al.*, 2010; Brosch *et al.*,

2013) do not readily apply. Distances defined based on set overlap measures as the Jaccard index (Levandowsky & Winter, 1971) have proven to be effective in recent studies investigating genetics and brain MRI (Toews & Wells III, 2016; Chauvin *et al.*, 2020). For example, by defining set equivalence in terms of nearest neighbor (NN) keypoint correspondences, the Jaccard distance may be used to predict pairwise relationships. Nevertheless, the assumption of binary or hard set equivalence (HSE) between keypoints is a crude approximation given probabilistic uncertainty inherent to natural image structure, and is ill-defined for variable sized datasets where the number of NN correspondences may be variable or unknown a-priori.

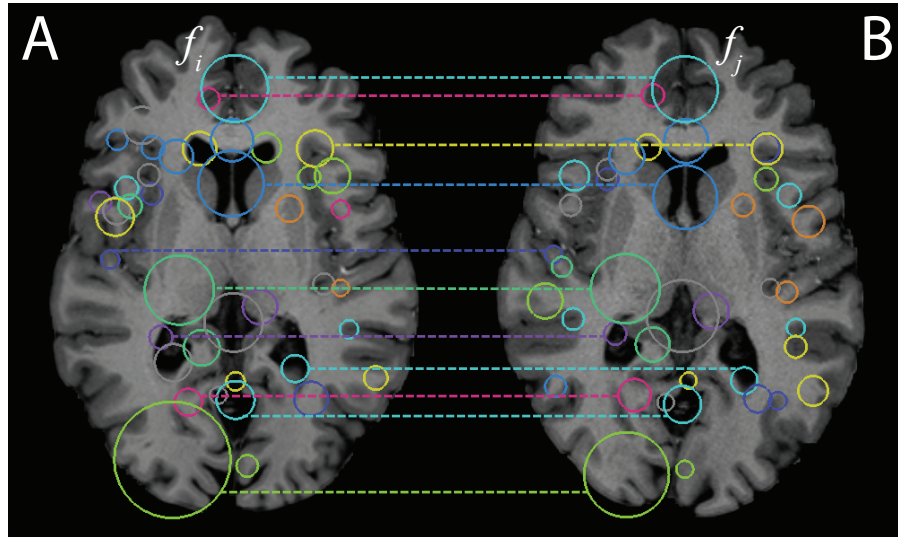


Figure 4.1 Illustrating two sets of 3D keypoints $A = \{f_i\}$ and $B = \{f_j\}$ extracted from MRI brain volumes of monozygotic twins. Colored circles represent the 3D locations $\bar{x} \in R^3$ and scales (radii) $\sigma \in R^1$ of keypoints in 2D axial slices, and links are correspondences identified via the SIFT-Rank (Toews & Wells, 2009) algorithm, note the degree of similarity between keypoint pairs

Our primary contribution is a new pairwise distance measure for 3D keypoint sets, generalizing the Jaccard index to account for soft set equivalence (SSE) between keypoint elements. SSE is achieved via a kernel framework for keypoint similarity, and we introduce a new kernel that normalizes pairwise keypoint displacement by keypoint scale, accounting for localization uncertainty in scale-space. Our work extends and compares against keypoint-based neuroimaging analysis methods (Toews & Wells III, 2016; Kumar *et al.*, 2018b; Chauvin *et al.*, 2020), specifically

the keypoint signatures approach (Chauvin *et al.*, 2020), currently the only method reporting perfect accuracy at identifying scans of the same individual in large public neuroimage MRI datasets, e.g. OASIS (Marcus *et al.*, 2007), ADNI (Jack Jr *et al.*, 2008) and HCP (Van Essen *et al.*, 2013). Here, following preliminary work (Chauvin *et al.*, 2019), we demonstrate that the soft Jaccard distance is specific enough to predict family relationships from medical images, and we present the first results for this new task using 1010 MRIs of 434 families from the HCP (Van Essen *et al.*, 2013) dataset. The entire method is based on highly efficient keypoint indexing, and scales to large datasets of volumetric image data.

4.2 Related Work

Our work is motivated both by the study of the link between genotype and the human brain phenotype from large datasets (Sabuncu *et al.*, 2016), and by practical applications such as maintaining accurate patient records in hospital Picture Archiving and Communication Systems (PACS). We adopted a memory-based learning model where all data are stored in memory and accessed via highly-specific keypoint queries. Memory-based learning requires no explicit training procedure (Boiman *et al.*, 2008), and approaches Bayes optimal performance as the number of data N becomes large (Cover & Hart, 1967). 3D SIFT keypoints (Toews *et al.*, 2013a) derived from Gaussian scale-space theory (Lowe, 2004; Lindeberg, 1994) are invariant to global similarity transforms and contrast variations, and thus account for misalignment and scanner variability between images. Efficient indexing reduces the quadratic $O(N^2)$ complexity of nearest neighbor (NN) keypoint lookup to $O(N \log N)$, e.g. via $O(\log N)$ KD-tree indexing (Muja & Lowe, 2014), ensuring that the method scales gracefully to large datasets, e.g. 7500 brain MRIs (Chauvin *et al.*, 2020) or 20000 lung CTs (Toews *et al.*, 2015).

Our method seeks a pairwise distance measure between 3D keypoint sets. As different images generally contain different numbers of elements, a natural choice for pairwise distance between variable sized sets are measures based on set intersection or overlap, e.g. the Jaccard distance metric (Levandowsky & Winter, 1971) first proposed in (Marczewski & Steinhaus, 1958) or the Tanimoto distance (Bajusz *et al.*, 2015; Rogers & Tanimoto, 1960). In medical image

analysis, the Jaccard distance is typically used to assess pixel-level segmentation accuracy (Yuan *et al.*, 2017), and has proven highly effective for assessing pairwise neuroimage similarity from 3D SIFT keypoint data (Toews & Wells III, 2016; Kumar *et al.*, 2018b), where hard set equivalence is determined by NN keypoint descriptor correspondences. Set theory is predicated on binary equivalence between set elements, which is difficult to justify in the case of keypoints extracted from image data. Soft set theory has been developed to account for non-binary equivalence (Molodtsov, 1999; Park *et al.*, 2012; Gardner *et al.*, 2014), where set operations including intersection and union are defined in terms of non-binary equivalence ranging from $[0,1]$. Jaccard-like distance was used for retrieving near-duplicate 2D photos (Chum *et al.*, 2008), where the ratio of soft intersection and union operators was used to account for document frequency. Our work here extends the Jaccard distance proposed in (Chauvin *et al.*, 2020) to include keypoint geometry in 3D Euclidean metric space coordinates, and compares to the HSE Jaccard measure used in (Toews & Wells III, 2016; Kumar *et al.*, 2018b), showing a significant improvement in predicting family relationships.

The keypoint methodology has a number of advantages, such as robustness to occlusion or missing correspondences, invariance to translation, scaling, rotation and intensity contrast variations between images. Our work adopts the 3D SIFT-Rank (Toews *et al.*, 2013a) representation, a robust, general tool used for a variety of imaging tasks, including registration (Toews *et al.*, 2013a; Machado *et al.*, 2018b), whole body segmentation (Wachinger *et al.*, 2018), classification (Toews *et al.*, 2015, 2010), regression (Toews *et al.*, 2012), without the need for context-specific training procedures or data. As in the original 2D SIFT approach, keypoint geometry is represented as a location $\bar{x}$ and scale σ , these are detected as extrema of a difference-of-Gaussian scale-space (Lowe, 2004), as $\operatorname{argmax}_{\bar{x}, \sigma} |I(\bar{x}, \sigma) - I(\bar{x}, \kappa\sigma)|$, approximating the Laplacian-of-Gaussian (Lindeberg, 1998), where $I(\bar{x}, \sigma) = I(\bar{x}) * \mathcal{N}(\bar{x}, \sigma^2)$ represents the image convolved with the Gaussian kernel $\mathcal{N}(\bar{x}, \sigma^2)$. The geometry of a keypoint may thus be characterised as an isotropic Gaussian density centered on $\bar{x}$ with variance σ^2 representing spatial extent σ in 3D Euclidean space. Rotationally symmetric Gaussian filtering and uniformly sampled derivative operators lead to scale and rotation invariance. The 3D SIFT-Rank descriptor is a 64-element

histogram of rank-ordered oriented gradients (HOG) (Toews *et al.*, 2013a), sampled about a scale-normalized reference frame centered on $\bar{x}$ and quantized uniformly into $2 \times 2 \times 2 = 8$ spatial locations $\times$ 8 orientations. Rank transformation provides invariance to arbitrary monotonic variations of descriptor element values (Toews & Wells, 2009).

The 3D SIFT-Rank keypoint method we use is a robust and well-established baseline, however our distance measure may generally be used with any keypoint representation. The search for new keypoint detectors and descriptors remains an active research focus, primarily in the context of 2D computer vision. We refer the reader to an extensive literature review of traditional hand-crafted solutions (Mukherjee *et al.*, 2015) and a recent discussion of 2D SIFT matching technology (Bellavia & Colombo, 2020). Deep learning keypoint methods have been investigated, examples include LIFT (Yi *et al.*, 2016), DISK (Tyszkiewicz *et al.*, 2020), LF-Net (Ono *et al.*, 2018), SuperPoint (Detone *et al.*, 2018). Nevertheless, variants of the original SIFT histogram descriptor, including SIFT-Rank (Toews & Wells, 2009), DSP-SIFT (Dalal & Triggs, 2005) or RootSIFT (Arandjelović & Zisserman, 2012) remain competitive with descriptors derived from training (Balntas *et al.*, 2017; Schönberger *et al.*, 2017), particularly for non-planar objects (Bellavia & Colombo, 2020), while requiring no training and few hyperparameters. Learning-based approaches often rely on hand-crafted keypoint detectors to generate training data (Yi *et al.*, 2016; Detone *et al.*, 2018). The challenges of developing deep keypoint architectures include training of individual components (i.e. keypoint detector, orientation estimator, descriptor), achieving invariance to rotation or scale changes, the need to retrain for different imaging modalities and body parts, and avoiding bias due to training data procedures (Geirhos *et al.*, 2019).

In terms of neuroimage analysis, brain fingerprinting methods have been used to investigate the variability of individuals based on neuroimage-specific preprocessing and data, e.g. brain segmentations from structural MRI (Wachinger *et al.*, 2015a; Valizadeh *et al.*, 2018), neuroanatomic parcellations and functional MRI (Finn *et al.*, 2015), fiber tracts from diffusion MRI (Kumar *et al.*, 2017). None of these methods have been sufficiently specific to achieve perfect accuracy at individual identification, and none have been used for the more difficult task

of predicting family members. Heritability studies have investigated correlations in MRIs of siblings, including structural (Thompson *et al.*, 2001; van der Lee *et al.*, 2017) and functional MRI (Colclough *et al.*, 2017), however these are not designed for prediction. In contrast, the keypoint methodology we adopt may be applied to generic imaging data with no preprocessing, representing distinctive patterns potentially present only in images of specific individuals or immediate family members. Our work here extends and compares against the keypoint-based neuroimage signatures method (Chauvin *et al.*, 2020), the only method reporting perfect accuracy at same-subject image prediction from standard neuroimage datasets (e.g. OASIS (Marcus *et al.*, 2007), ADNI (Jack Jr *et al.*, 2008), HCP (Van Essen *et al.*, 2012)).

Our results here are the first in the literature for the task of family member prediction from volumetric medical image data. Family member prediction has been investigated in the context of 2D face photographs, based on clustering (Kiley & Hossain, 2020), conditional random fields (Dai *et al.*, 2015), deep neural networks (Ahmad *et al.*, 2019), discriminant analysis (Juefei-Xu & Savvides, 2013). These solutions are not directly related to our work, as they generally address aspects and challenges specific to the context of 2D photography, e.g. facial expression and lighting conditions (Paone *et al.*, 2014), multiple images per person, group photos of relatively few families (e.g. 16) (Dai *et al.*, 2015). Accuracy is not generally available for siblings based on genotyping, and near 100% accuracy for monozygotic twin identification from face photographs has not been reported.

4.3 Method

We seek a pairwise distance measure between two sets $A = \{f_i\}$ and $B = \{f_j\}$ that can be used to estimate proximity and thus genetic relationships between subjects from image data. We first describe our method in terms of general set theory, then later include details pertaining to set elements f_i defined as 3D image keypoints. Our measure begins with the Jaccard index or intersection-over-union $J_{HSE}(A, B)$ defined as

$$J_{HSE}(A, B) = \frac{|A \cap B|}{|A| + |B| - |A \cap B|}, \quad (4.1)$$

where in Equation (4.1), $J_{HSE}(A, B)$ is defined by binary or hard set equivalence (HSE) between set elements. $A \cap B$ is the intersection operator between sets A and B , and $|A \cap B|$, $|A| = |A \cap A|$, $|B| = |B \cap B|$ represent set cardinality operators that count the numbers of elements present in each set, where $|A \cap B| \leq \min\{|A|, |B|\}$. The Jaccard index is a similarity measure, and may be used to define various distance measures including the Jaccard distance metric (Levandowsky & Winter, 1971) $1 - J_{HSE}(A, B)$ or the Tanimoto distance $-\log J_{HSE}(A, B)$.

In the case of sets of real data samples, for example keypoint descriptors, binary equivalence may be difficult to establish due to noise or uncertainty in the measurement process, and we seek to redefine $J_{HSE}(A, B)$ in Equation (4.1) such that it more accurately accounts for non-binary equivalence between set elements. This may generally be accomplished by redefining the standard set intersection cardinality operator $|A \cap B|$ in Equation (4.1) by an analogous soft cardinality operator $\mu(A \cap B)$, leading to the following expression for our proposed Jaccard index based on soft set equivalence $J_{SSE}(A, B)$

$$J_{SSE}(A, B) = \frac{\mu(A \cap B)}{\mu(A) + \mu(B) - \mu(A \cap B)}. \quad (4.2)$$

In Equation (4.2), $J_{SSE}(A, B)$ is defined by the soft cardinality operator $\mu(A \cap B)$ as described in the following section, including cardinality operators $\mu(A) = \mu(A \cap A)$, $\mu(B) = \mu(B \cap B)$. In order to ensure that J_{SSE} remains bounded to the range $[0, 1]$, it is important that $\mu(A \cap B)$ be upper bounded by the minimum of the individual soft set cardinalities $\mu(A \cap B) \leq \min\{\mu(A), \mu(B)\}$.

Defining $\mu(A \cap B)$: The cardinality of the soft set intersection $\mu(A \cap B)$ is intended to reflect the uncertainty in equivalence between set elements $f_i \in A$ and $f_j \in B$. We define soft equivalence between elements $f_i \in A$ and $f_j \in B$ as a similarity function or kernel $\mathcal{K}(f_i, f_j)$ ranging from $[0, 1]$, where $\mathcal{K}(f_i, f_j) = 1$ indicates exact equivalence and $\mathcal{K}(f_i, f_j) = 0$ represents the absence of equivalence. We then define a measure $\mu(A \rightarrow B)$ over a mapping $A \rightarrow B$ from set A to set

B as:

$$\mu(A \rightarrow B) = \sum_i^{|A|} \max_{f_j \in B} \mathcal{K}(f_i, f_j) \quad (4.3)$$

The similarity function $\mathcal{K}(f_i, f_j)$ in Equation (4.3) may generally be defined according to the specific representation of set elements f_i and f_j . The maximum operator $\max_{f_j \in B} \mathcal{K}(f_i, f_j)$ ensures the bound $\mu(A \rightarrow B) \leq \mu(A)$ via a partial injective mapping from A to B , embodying the intuition that each element $f_i \in A$ has at most one counterpart $f_j \in B$. Note also that max operator may be computed via computationally efficient NN indexing methods. The soft intersection cardinality $\mu(A \cap B)$ may be defined as a symmetrized version of $\mu(A \rightarrow B)$:

$$\begin{aligned} \mu(A \cap B) &= \min \{ \mu(A \rightarrow B), \mu(B \rightarrow A) \} \\ &= \min \left\{ \sum_i^{|A|} \max_{f_j \in B} \mathcal{K}(f_i, f_j), \sum_j^{|B|} \max_{f_i \in A} \mathcal{K}(f_j, f_i) \right\} \end{aligned} \quad (4.4)$$

Note that the Jaccard distance measure derived using the cardinality operator $\mu(A \cap B)$ as defined by Equations (4.3) and (4.4) is symmetric and follows the identity of indiscernibles, and thus may be considered to be a semi-metric. The triangle inequality holds in the case of binary equivalence (Levandowsky & Winter, 1971) but not generally for soft equivalence relationships. Equation (4.4) may be used to define the cardinality of a single set A as $\mu(A) = \mu(A \cap A) = |A|$ both for hard and soft equivalence since, since for a given element $f_i \in A$, $\max_{f_j \in A} \mathcal{K}(f_i, f_j) = \mathcal{K}(f_i, f_i) = 1$. The standard Jaccard index in Equation (4.1) is thus a special case of the soft Jaccard index in the case of a binary-valued kernel $\mathcal{K}(f_i, f_j) \in \{0, 1\}$.

Defining $\mathcal{K}(f_i, f_j)$: The soft set intersection cardinality $\mu(A \cap B)$ may be adapted to a specific task by defining $\mathcal{K}(f_i, f_j)$ according to the representation of elements f_i . In the context of this article, a set element f_i is a 3D scale-invariant keypoint $f_i = \{\bar{a}_i, \bar{g}_i\}$ as described in (Toews *et al.*, 2013a), where $\bar{g}_i$ and $\bar{a}_i$ are descriptors of local keypoint geometry and appearance, respectively. Keypoint geometry $\bar{g}_i = \{\bar{x}_i, \sigma_i\}$ consists of 3D location $\bar{x}_i$ and scale σ_i , and appearance $\bar{a}_i$

is a vector of local image information, i.e. a rank-ordered histogram of oriented gradients (HOG) (Toews & Wells, 2009).

The kernel $\mathcal{K}(f_i, f_j)$ operates on keypoint elements $f_i = \{\bar{a}_i, \bar{g}_i\} \in A$ and $f_j = \{\bar{a}_j, \bar{g}_j\} \in B$. Here we relax the assumption of hard equivalence using squared exponential kernels with non-zero support, factored into kernels operating separately on local keypoint appearance $\mathcal{K}(\bar{a}_i, \bar{a}_j)$ and geometry $\mathcal{K}(\bar{g}_i, \bar{g}_j)$ variables:

$$\mathcal{K}(f_i, f_j) = \mathcal{K}(\bar{a}_i, \bar{a}_j) \mathcal{K}(\bar{g}_i, \bar{g}_j) \quad (4.5)$$

The factorization in Equation (4.5) is due to the use of descriptors $\bar{a}_i$ that are invariant to 7-parameter similarity transforms of the 3D image coordinate system from which geometry $\bar{g}_i$ is derived. The two kernels in Equation (4.5) are defined as squared exponential functions as follows.

The appearance kernel $\mathcal{K}(\bar{a}_i, \bar{a}_j)$ is defined by the squared Euclidean distance $\|\bar{a}_i - \bar{a}_j\|_2^2$ between appearance vectors $\bar{a}_i$ and $\bar{a}_j$:

$$\mathcal{K}(\bar{a}_i, \bar{a}_j) = \exp \left(-\frac{\|\bar{a}_i - \bar{a}_j\|_2^2}{\alpha^2} \right) \quad (4.6)$$

where in Equation (4.6), α is a bandwidth parameter that may be estimated adaptively as

$$\alpha = \min_{f_j \in \Omega} \|\bar{a}_i - \bar{a}_j\|_2^2, \text{ s.t. } \|\bar{a}_i - \bar{a}_j\|_2^2 > 0 \quad (4.7)$$

the minimum Euclidean distance between appearance descriptor $\bar{a}_i \in A$ and the nearest descriptor $\bar{a}_j \in \Omega \setminus A$ within the entire available dataset Ω excluding A . Note this choice of estimator is not strictly symmetric, however it allows the kernel to adapt to arbitrary dataset sizes, shrinking the resolution of prediction as the number of data grows large, and does not affect the symmetry of Equation (4.4).

The geometry kernel $\mathcal{K}(\bar{g}_i, \bar{g}_j)$ is novel to this work, and is defined as the product of two kernels, one modeling keypoint location conditional on keypoint scale $\mathcal{K}(\bar{x}_i, \bar{x}_j; \sigma_i, \sigma_j)$ and the $\mathcal{K}(\sigma_i, \sigma_j)$ modeling scale alone

$$\mathcal{K}(\bar{g}_i, \bar{g}_j) = \mathcal{K}(\bar{x}_i, \bar{x}_j; \sigma_i, \sigma_j) \mathcal{K}(\sigma_i, \sigma_j) \quad (4.8)$$

These kernels are defined as follows

$$\mathcal{K}(\bar{x}_i, \bar{x}_j; \sigma_i, \sigma_j) = \exp \left(-\frac{\|\bar{x}_i - \bar{x}_j\|_2^2}{\sigma_i \sigma_j} \right) \quad (4.9)$$

$$\mathcal{K}(\sigma_i, \sigma_j) = \exp \left(-\log^2 \left(\frac{\sigma_i}{\sigma_j} \right) \right) \quad (4.10)$$

Kernel $\mathcal{K}(\bar{x}_i, \bar{x}_j; \sigma_i, \sigma_j)$ in Equation (4.9) penalizes the squared distance between keypoint coordinates within a local reference frame, normalized by a variance proportional to the product of keypoint scales $\sigma_i \sigma_j$. This variance embodies uncertainty in keypoint location due to scale, and has a computational form that is reminiscent of mass in Newton's law of gravitation or electric charge magnitude in Coulomb's law. Figure 4.2 demonstrates how this kernel variance normalizes higher localization error associated with keypoints of larger scale. Kernel $\mathcal{K}(\sigma_i, \sigma_j)$ in Equation (4.10) penalizes multiplicative difference between keypoint scales (σ_i, σ_j) .

Note that while the appearance kernel in Equation (4.6) is invariant to global similarity transforms due to descriptor invariance, the geometry kernel in Equation (4.8) measures zero-mean keypoint displacement and thus requires data to be aligned within a common spatial reference frame. Alignment may be established via standard subject-to-atlas registration prior to keypoint extraction or from keypoint correspondences after extraction using feature-based alignment (Toews *et al.*, 2013a). The following workflow is used to compute pairwise distances for a set of N images:

1. Extract keypoints from each image
2. Align keypoints to an atlas template (optional for pre-aligned images)

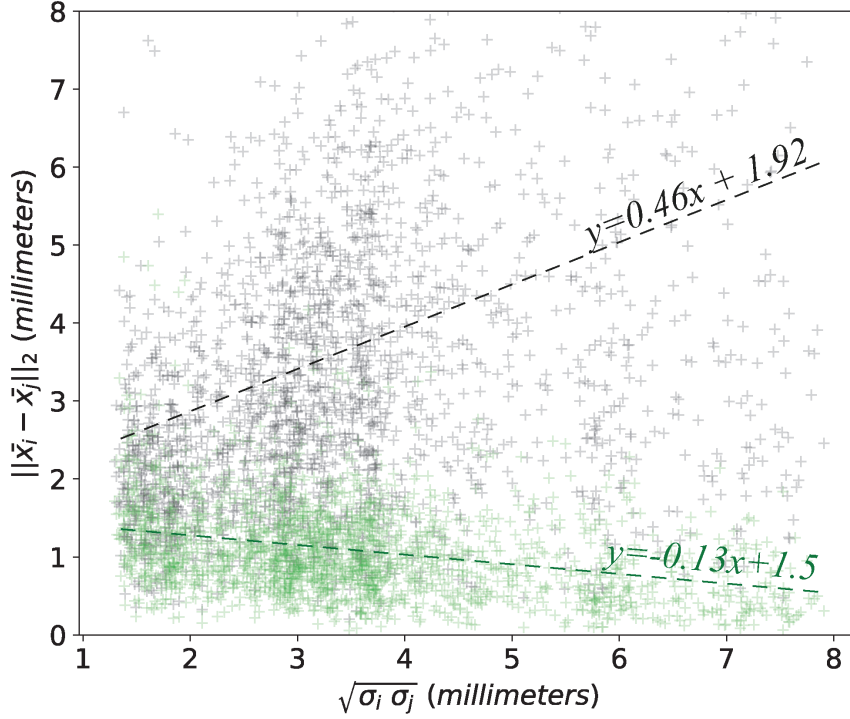


Figure 4.2 Graph of spatial localization error $\|\bar{x}_i - \bar{x}_j\|_2$ vs. geometric mean scale $\sqrt{\sigma_i \sigma_j}$. Lines represent the linear regression based on each point set. Note how location error increases with scale (black) due to uncertainty, however this effect is reduced by scale normalization (green) as in Equation (4.9). This visualization is based on keypoint correspondences between 10 MRI aligned volumes

3. Identify k-NN correspondences $(\bar{a}_i, \bar{a}_j)$ minimizing the Euclidean distance between descriptors $\|\bar{a}_i - \bar{a}_j\|$
4. Compute pairwise image distances by evaluating kernels from k-NN correspondences

Figure 4.3 shows the effect of our proposed geometry kernel in regularizing and favoring geometrically plausible correspondences for an example twin sibling pair.

4.4 Experiments

Experiments investigate the ability of the Jaccard distance to predict pairwise relationships between whole-brain MRI scans, where relationships include 1) close genetic links between

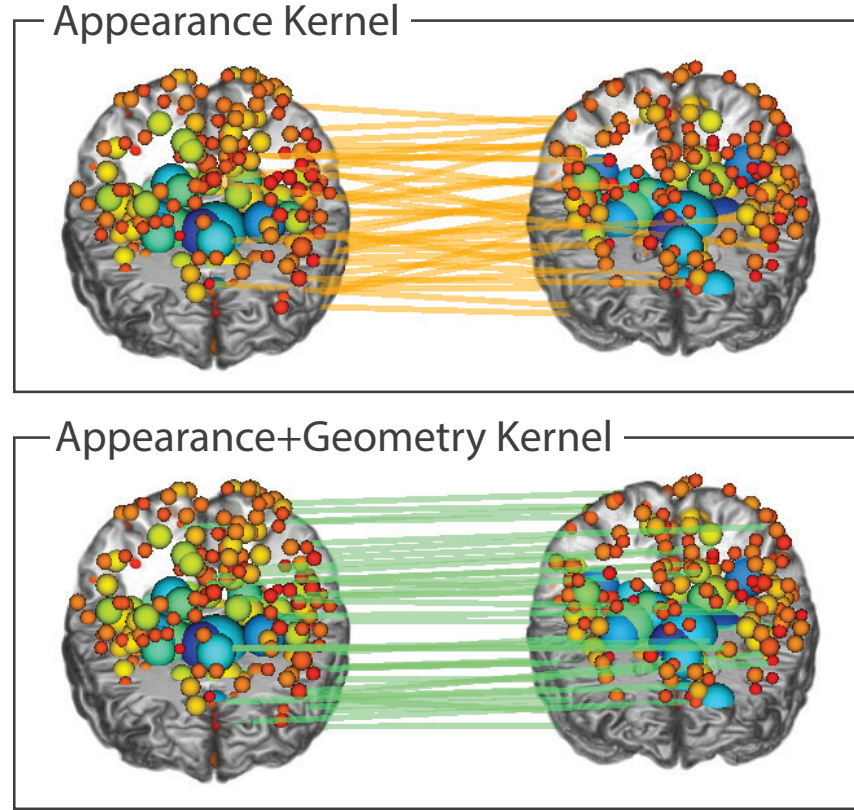


Figure 4.3 Visualizing the keypoint correspondences driving classification between a pair of identical twin brains, showing the 100 correspondences with the highest SSE kernel values. The appearance kernel alone (upper, Equation (4.6)) may include geometrically inconsistent matches (diagonal lines), however these are down weighted when combined with the geometry kernel (lower, Equations (4.9) and (4.10)), leading to geometrically consistent matches (horizontal lines) and improved prediction performance

siblings sharing 25-100% of their polymorphic genes and 2) broad genetic links between non-siblings sharing nominal genetic information due to sex and common racial ancestry. We hypothesize that closer genetic proximity will be reflected in higher pairwise similarity and thus lower Jaccard distance. We expect that pairwise distance based on our proposed soft set equivalence (SSE) $d_J(A, B) = -\log J_{SSE}(A, B)$ as defined in Equations (4.2) and (4.4) and our geometry kernel as defined by Equations (4.9) and (4.10) will lead to improved identification of pairwise relationship labels.

Prediction is based on pairwise Jaccard distance and a leave-one-out protocol, analogous to querying a subject MRI in a hospital PACS. Note that this is a deterministic procedure with no explicit training stage, where the primary hyperparameter is the number of nearest neighbors k per keypoint. Our computational workflow, as previously described, follows three steps 1) 3D SIFT-Rank keypoints are extracted from individual pre-processed images, 2) kNN correspondences are identified between keypoints from all N images, and 3) Jaccard distances are evaluated between all image pairs from kernels and kNN correspondences.

4.4.1 Data and Computational Details

Our data set consists of MRI scans of $N = 1010$ unique subjects from the Human Connectome Project (HCP) Q4 release ([Van Essen *et al.*, 2012](#)), aged 22-36 years (mean 29 years), acquired from a diverse population including 468 males and 542 females, and 434 unique families. The MRI data are provided as T1w volumes at isotropic 0.7mm voxel resolution and preprocessed via a standard neuroimaging pipeline, including rigid subject-to-atlas registration and skull-stripping. Keypoint extraction requires approximately 3 sec. / per image and results in an average of 1,400 keypoints per image for a total of 1,488,065 keypoints. We use a GPU implementation ([Pepin *et al.*, 2020](#)) of the 3D SIFT-Rank keypoint algorithm ([Toews *et al.*, 2013a](#)) that produces identical keypoints at a $7\times$ speedup. Approximate kNN correspondences between appearance descriptors are identified across the entire database using efficient KD-tree indexing ([Muja & Lowe, 2014](#)), where lookup requires 0.8 sec. / subject for $k=200$ nearest neighbors on an i7-5600@2.60Ghz machine with 16 GB RAM (1.64 GB used).

Given $N=1010$ images, there are a total of $N(N - 1)/2 = 509545$ pairwise relationships to be evaluated. Each subject pair is assigned by one of five possible relationship labels $L = \{MZ, DZ, FS, HS, UR\}$ for monozygotic twins (MZ), dizygotic twins (DZ), full non-twin siblings (FS), half-siblings (HS) and unrelated non-siblings (UR), for totals of $\{134, 71, 607, 44, 508689\}$ relationships per label. Family relationship labels are based on mother and father identity and zygosity (for twins), confirmed via genome-wide single nucleotide polymorphism (SNP) genotyping ([HCP, 2017](#)). Note the sparse structure of family

relationship labels, where each sibling label (MZ, DZ, FS, HS) is unique to one of 434 families and the vast majority of pairs are UR ($0.998 = 508689/509545$). The nominal probability of randomly guessing the correct family is approximately $0.0023 = 1/434$, and the challenge is thus to efficiently and accurately identify the small number of family relationships. Note that evaluating the similarity of $N(N - 1)/2$ pairs via brute force image matching or registration becomes computationally intractable for large N .

Image alignment is fundamental to the functioning of our proposed geometry kernel, which measures zero-mean deviations in keypoint location within a common spatial reference frame. We thus evaluated prediction following two different subject-to-atlas registration methods: the original 3D rigid transform provided by the HCP mapping images to the standard MNI atlas using the FLIRT algorithm (Glasser *et al.*, 2013; Jenkinson *et al.*, 2002), and feature-based alignment (FBA) (Toews *et al.*, 2013a) using keypoint correspondences to estimate a 3D similarity transform mapping images to a subject atlas, for 20 different randomly selected atlases. All alignment solutions were very close to the reference MNI alignment (rotation differed by $4.31^\circ \pm 2.84$, translation by $3\text{mm} \pm 3.5$), reflecting a degree of inter-subject variability given different atlases. Most notably, however, was that AUC values were virtually identical for all prediction trials, with a maximum standard deviation of $\epsilon = 0.0017$ (see Table 4.1), indicating that prediction was not sensitive to the alignment solution used. The results we report are thus consistent with the reference MNI alignment provided with the HCP data.

4.4.2 Close genetic proximity: Siblings

Figure 4.4 shows distributions of pairwise Jaccard distance conditioned on pairwise relationships. Our proposed kernel based on appearance and geometry (Figure 4.4, green) generally increases separation between different sibling relationships in comparison to appearance only (Figure 4.4, orange) as in (Chauvin *et al.*, 2020). A two-tailed Kolmogorov-Smirnoff (KS) test shows all distributions to be significantly different ($p\text{-value} < 1e - 10$) except those of DZ and FS siblings sharing approximately 50% of their genes ($p\text{-value} = 0.0199$). Furthermore, pairwise KS statistics were all lower for appearance+geometry kernels compared to appearance only,

indicating that the geometry kernels lead to increased separation of relationship categories. The Jaccard distance may be rapidly evaluated, and pairs which are outliers from their expected distributions may be easily flagged and inspected for irregularities. For example, the highest Jaccard distance outliers indicated noticeable spatial misalignment between a small number of subject pairs (e.g. Figure 4.4a. The lowest distance outliers for UR pairs may indicate potential sibling relationships. The blue circles near Figure 4.4b are pairs involving 3 subjects flagged as UR due to genotyping errors (including a MZ twin) (HCP, 2017). Here, nearest neighbor Jaccard distance was used to correctly predict the self-reported families for all three cases, a result confirmed by the HCP.

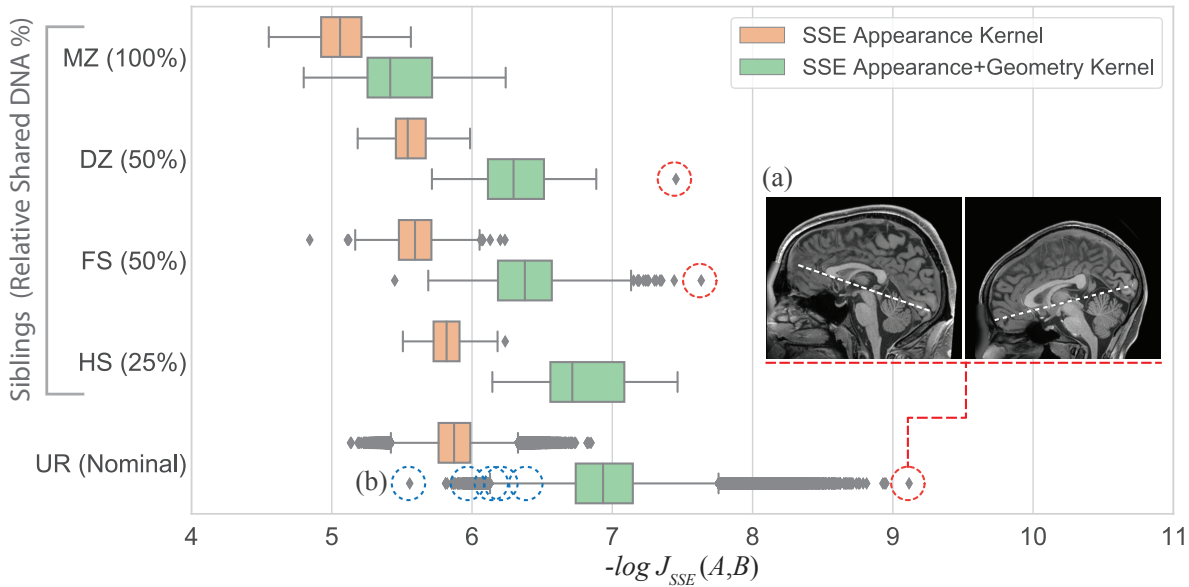


Figure 4.4 Distributions of Jaccard distances $-\log J_{SSE}(A, B)$ conditioned on pairwise labels $L = \{MZ, DZ, FS, HS, UR\}$, comparing kernels with and without geometry (green vs. orange). Note that the distance generally increases with genetic separation, and that geometry kernels increase separation between sibling relationship labels. Outliers, i.e. pairs outside of their expected relationship distance distributions, can be easily identified and inspected for irregularity, e.g. (a) unusually high distances may indicate spatial misalignment (red circles) and (b) unusually low UR distances may indicate cases of incorrect family labels (blue circles)

As sibling pairs exhibit significantly lower distance than UR pairs, we investigate the degree to which they can be distinguished from unrelated pairs based on a simple distance threshold.

Figure 4.5 shows the Receiver Operating Characteristic (ROC) curves for sibling relationships based on distance, comparing our SSE kernels for appearance and geometry, appearance only and the binary HSE kernel. Table 4.1 quantifies the improvement of SSE vs HSE, for various numbers of keypoint nearest neighbors (20, 100, 200), where the highest area-under-curve (AUC) values are obtained for SSE with $k=200$. The performance of SSE kernels generally increases with the number of NNs, and is always superior for combined appearance and geometry kernels. Conversely, HSE classification performance decreases, even falling below the diagonal in the most challenging case of HS pairs in Figure 4.5d, as additional hard correspondences appear to accumulate systematically on a subset of typical but unrelated subjects. For completeness, the highest UR prediction result was $AUC = 0.951$ for SSE, appearance and geometry kernels, $k=200$.

Table 4.1 Area Under the Curve (AUC) values for sibling classification, comparing kernels and numbers of nearest neighbors (k). $\epsilon = 0.0017$ is the maximum AUC standard deviation observed from 21 different subject-to-atlas registration trials. Note $SSE_{App+Geo}$ kernels significantly outperform SSE_{App} and HSE for the more difficult DZ and FS pairs

Kernel Label	$k = 20$			$k = 100$			$k = 200$		
	HSE	SSE_A	SSE_{A+G}	HSE	SSE_A	SSE_{A+G}	HSE	SSE_A	SSE_{A+G}
MZ	0.9983	0.9993	0.9996 $\pm \epsilon$	0.9544	0.9995	0.9998 $\pm \epsilon$	0.9272	0.9994	0.9999 $\pm \epsilon$
DZ	0.8922	0.8825	0.9044 $\pm \epsilon$	0.8018	0.9034	0.9250 $\pm \epsilon$	0.7541	0.9181	0.9391 $\pm \epsilon$
FS	0.8423	0.8433	0.8753 $\pm \epsilon$	0.7611	0.8569	0.8888 $\pm \epsilon$	0.7358	0.8680	0.8989 $\pm \epsilon$

4.4.3 Distant genetic proximity: unrelated subjects

Unrelated (UR) subject pairs share nominal amounts of genetic information, with subtle similarities due to demographic factors such as common sex and ancestral race. We thus expect whole-brain distance to be lowest for pairs of the same race and sex (R, S), highest for different race and sex ($\bar{R}, \bar{S}$), and intermediate for either same race ($R, \bar{S}$) or sex ($\bar{R}, S$). While the mean distances for conditional distributions in Figure 4.6 generally increase with differences in

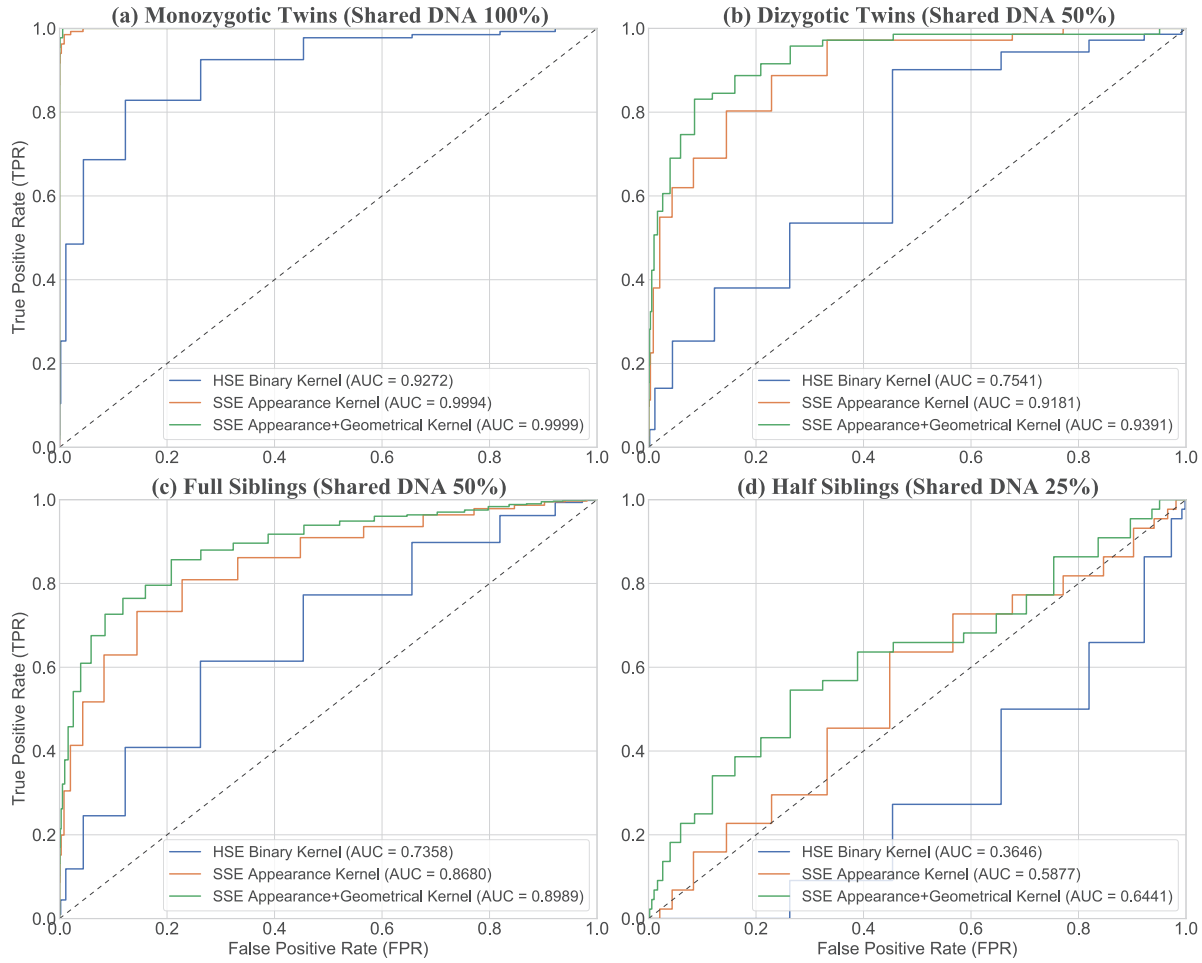


Figure 4.5 ROC curves for sibling identification based on Jaccard distance for (a) MZ, (b) DZ, (c) FS and (d) HS pairs, comparing three kernels for $k=200$ NNs. Note that the prediction AUC is always highest for our proposed kernel combining appearance and geometry (green), and decreases as expected with the degree genetic and developmental difference, e.g. in the order MZ, DZ, FS, HS

demographic labels, there are many exceptions where pairs with different labels exhibit lower distance than those with the same labels. We note this is consistent with pairwise genetic differences, where pairs of unrelated individuals from different populations may often exhibit higher genetic similarity than those from the same population (Witherspoon *et al.*, 2007).

Age difference between subjects is a potential confounding factor in whole brain Jaccard distance, and we plotted the variation of distance vs. age difference in Figure 4.7. Distance distributions

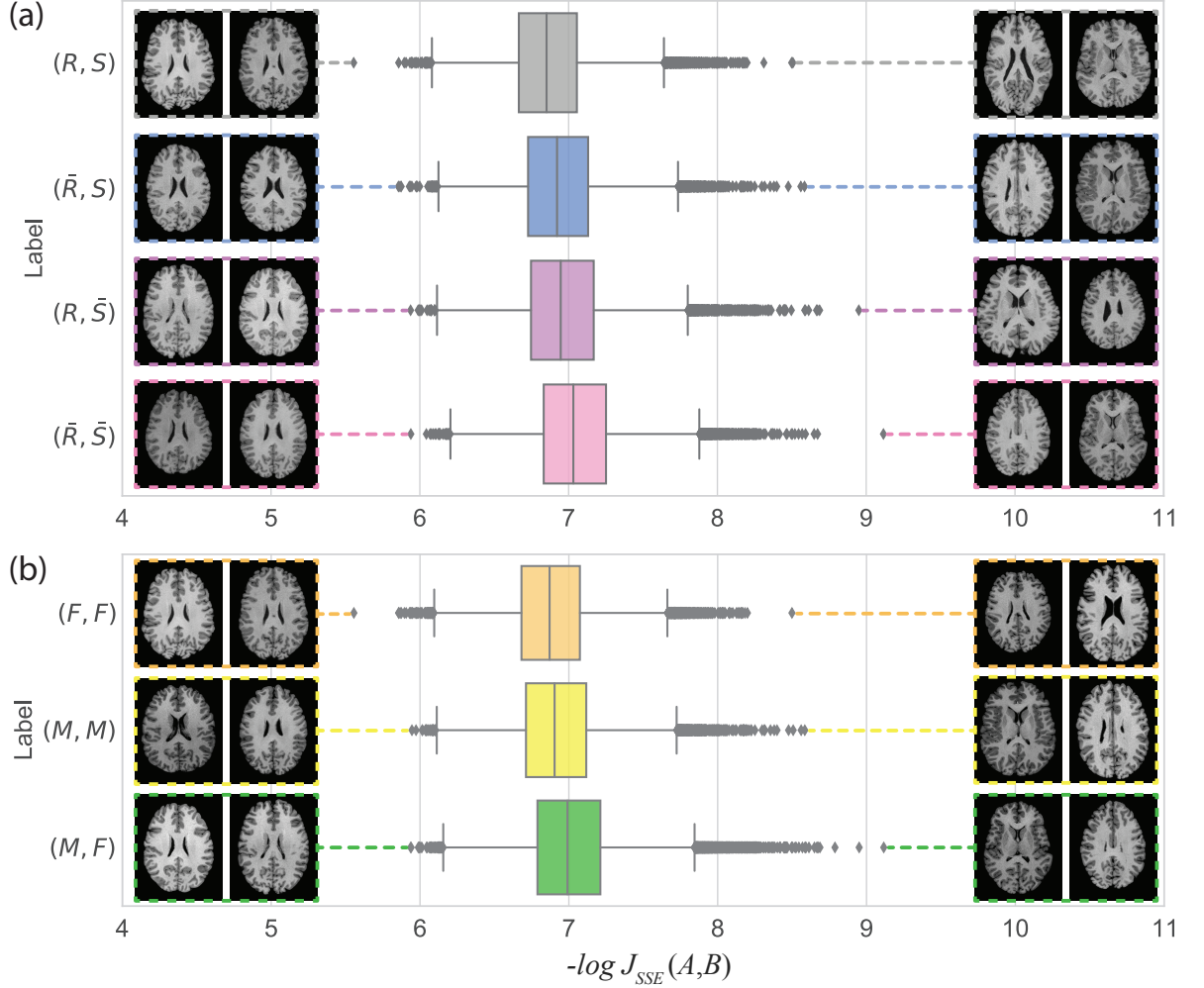


Figure 4.6 Jaccard distance $-\log J_{SSE}(A, B)$ distributions between unrelated pairs conditioned on shared demographic information. **(a)** same race, same sex (R, S) ; different race, same sex $(\bar{R}, S)$; same race, different sex $(R, \bar{S})$, and different race, different sex $(\bar{R}, \bar{S})$ and **(b)** Female-Female (F, F) , Male-Male (M, M) , Male-Female (M, F) . Image pairs corresponding to minimum and maximum distances for each distribution are shown for visualization

are virtually identical across the HCP subject age range spanning 22-36 years of age, indicating that age difference is not a major confound in this relatively young and healthy HCP cohort where brain morphology is relatively stable. Note that the Jaccard distance was found to increase with age difference in older subjects due to natural aging and neurodegenerative disease in (Chauvin *et al.*, 2020).

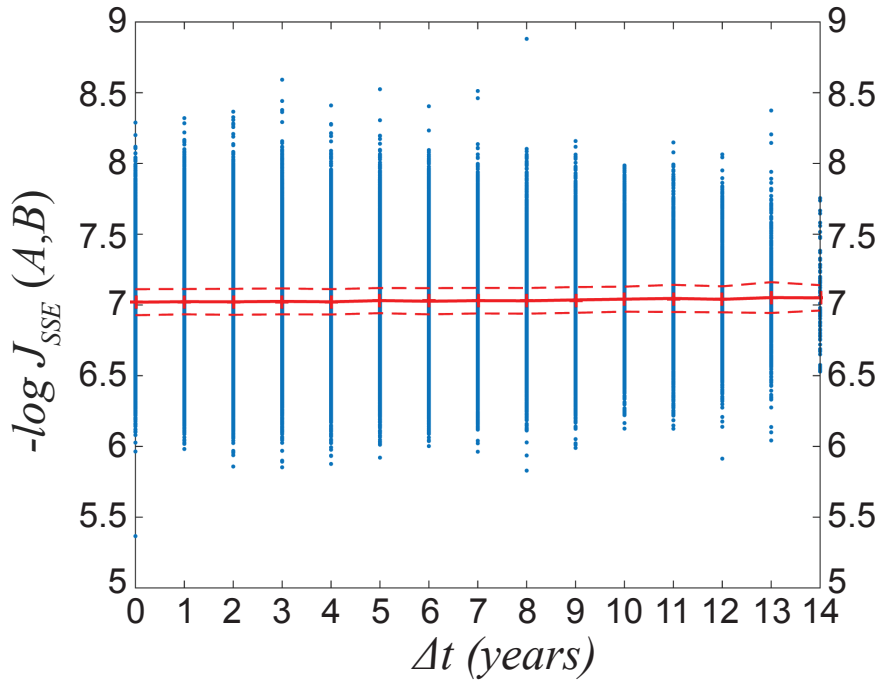


Figure 4.7 Distributions of Jaccard distance $-\log J_{SSE}(A, B)$ conditioned on age difference Δt between unrelated pairs (A, B) conditioned on age difference Δt . The mean (solid red line) and standard deviation (dashed red lines) are plotted for each Δt . Age difference has no significant impact on the distance for unrelated healthy young adult brains (age 22-36 years), eliminating a possible confounding factor

4.4.4 Group Prediction: Sex

While the pairwise Jaccard distance is highly informative regarding sibling relationships, it is insufficient for predicting group labels such as sex, age or disease. A simple modification can be used to predict group labels, by evaluating the distance between a single keypoint set A and supersets formed by the union of group members (excluding subject A), e.g. keypoints for *all Males* $d_J(A, \mathcal{M})$ and *all Females* $d_J(A, \mathcal{F})$. Supersets $\mathcal{M}$ and $\mathcal{F}$ are solely composed of unrelated subjects to avoid potential biases due to family relationships. These distances are combined in a basic linear classifier with a single threshold parameter τ to adjust for differences

in the numbers of keypoints per group based on the following equation:

$$Class(A) = \begin{cases} Female & \text{if } d_J(A, \mathcal{F}) - d_J(A, \mathcal{M}) + \tau > 0 \\ Male & \text{otherwise.} \end{cases} \quad (4.11)$$

Figure 4.8 shows ROCs curves for sex prediction obtained by varying τ over the range $(-\infty, \infty)$, comparing Jaccard distance computed with binary, appearance only and combined appearance and geometry kernels. A possible confound is brain size, which is on average slightly larger for males than females (Eliot *et al.*, 2021). As the appearance kernel is invariant to image scale, the AUC=0.93 reflects prediction accuracy independently of image size. Our proposed combined approach achieves the highest AUC=0.97, again outperforming other options.

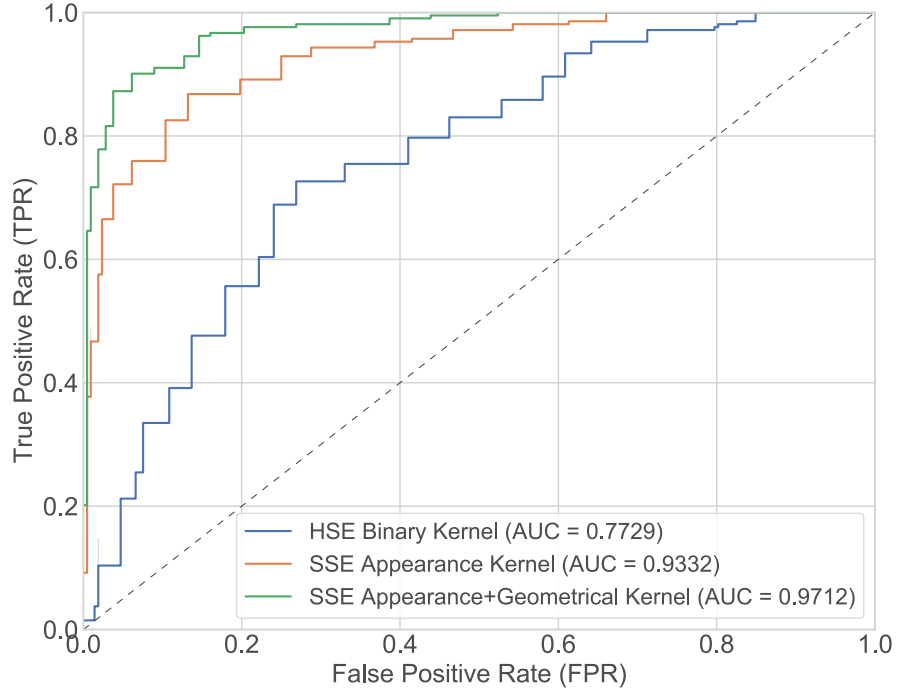


Figure 4.8 ROC curves and AUC for sex prediction based on individual-to-group Jaccard distance, comparing binary, appearance only, and combined appearance and geometry kernels.

4.5 Discussion

In this paper we propose a novel, highly specific distance measure between volumetric images represented as invariant keypoints. Our measure generalizes the Jaccard index to account for soft equivalence between keypoints, and a novel kernel estimator is proposed to model keypoint geometry in terms of location and scale within normalized image space. The soft Jaccard index is used as a distance measure to predict pairwise relationships between MRI brain scans of 1010 subjects from 434 families, including siblings and twins, and significantly improves upon previous work based on binary equivalence and appearance descriptors (Chauvin *et al.*, 2020; Kumar *et al.*, 2018b). We report the first results for predicting relationships from medical images, a new task, where monozygotic twins can be identified with virtually perfect accuracy. A minor modification allows the Jaccard distance to predict group labels such as sex with high accuracy. Our geometry kernel requires spatial normalization, however trials involving various linear subject-to-atlas alignment solutions including robust keypoint-based alignment (Toews *et al.*, 2013a) show that the prediction performance to be insensitive to the specific alignment solution used. The improved prediction afforded by our geometry kernel thus suggests that family members tend to align in a consistent manner regardless of the solution used.

Our method promises to be a useful tool for curating large medical image datasets for precision medicine and research purposes. A memory-based model using efficient and robust algorithms for 3D keypoint extraction and indexing (Toews *et al.*, 2013a,b) allows for fine-grained comparisons between $O(N^2)$ image pairs in $O(N \log N)$ computational complexity. The Jaccard distance may thus be used to rapidly validate relationship labels, e.g. unexpectedly low distance may indicate related individuals and unexpectedly high distance may indicate unrelated individuals or spatially misaligned pairs. Previous work identified mislabelled scans of individuals in large neuroimage datasets (Chauvin *et al.*, 2020), here our improved method allowed us predict the self-reported families of three subjects labelled as unrelated due to inconclusive genotyping (HCP, 2017), on the basis of nearest neighbor soft Jaccard distance. This exceptional result was confirmed by the Human Connectome Project, and serves as a concrete example as to how our method can be used to validate labels associated with medical imaging data.

The task of pairwise relationship prediction may be viewed as the finest grain of categorization, where the number of unique pairwise relationship labels is linear $O(N)$ in the number of data N , e.g. 434 families from $N=1010$ images, as opposed to typical classification where all data are associated with a small number of labels (e.g. male, female). Pairwise prediction thus represents a challenge for ubiquitous deep neural network methods, which generally require large numbers of training data per category, and training for specific modalities and body parts. In contrast, generic 3D SIFT keypoints may be used as-is with arbitrary imaging modalities and contexts with no training. In future work, generic keypoint correspondences here could potentially be used to train domain-specific keypoint models (Yi *et al.*, 2016; Detone *et al.*, 2018). Auto-encoders used for anomaly detection could potentially be adapted to generate subject-specific codes between pairs of images (Baur *et al.*, 2021). Analysis beyond pairs of subjects could be generalized via graph theory to model the clique structure of families (Blair & Peyton, 1993). All code required to reproduce our results may be obtained at <https://github.com/3dsift-rank>.

CONCLUSION AND RECOMMENDATIONS

In this thesis, we presented an atom-like model of SIFT features derived from fundamental operators such as the Laplacian and including discrete orientation and sign to address earlier models' drawbacks such as orientation ambiguity and challenges with contrast inversion in multi-modal images. Our approach subsequently proposes evaluating the likelihood of two particles emerging from the same distribution based on kernel density estimates of feature appearance and geometry, and ultimately extending it to volumetric images allowing for fast image registration and quantification of image similarity based on a Jaccard-like measure of set overlap. Our model has already proven useful for the neuroimaging community, by identifying mislabeled data in widely used large datasets or by efficiently registering 3D images in the case of stroke detection. It has been cited as "incredibly powerful in capturing a keypoint signature 'brainprint' [...] despite ageing and neurodegenerative disease progression" in (Fetit *et al.*, 2020) and as: "Impressively, by identifying keypoint signatures robust to scan-to-scan variation, the algorithm matched MR images from an individual collected as many as 11 years apart" in (Finn & Rosenberg, 2021). A wide range of studies could stem from this research, assessing for example the impact of age, gender or diseases in large scale analyses, as in (Rokooie, 2020).

Although the work presented in this thesis focus on evaluating similarity between pairs of features from distinct images, the geometrical kernel $\mathcal{K}(g_i, g_j)$ could also serve, for example, to establish a fingerprint of cortical regions with pairs of features from the same image, via a distance graph between features on the scale-space, accounting for feature scale (see Figure 5.1). Graph constraints could also be added like in (Gopinath *et al.*, 2022) to learn surface data of cortical structures through graph convolution networks (Gopinath, 2021).

Another conceivable future study route would be the combination of learn and memory-based models to accomplish some coarse to fine hierarchical classification, comparable to how the human brain recognises a face before identifying a specific individual. Deep learning algorithms

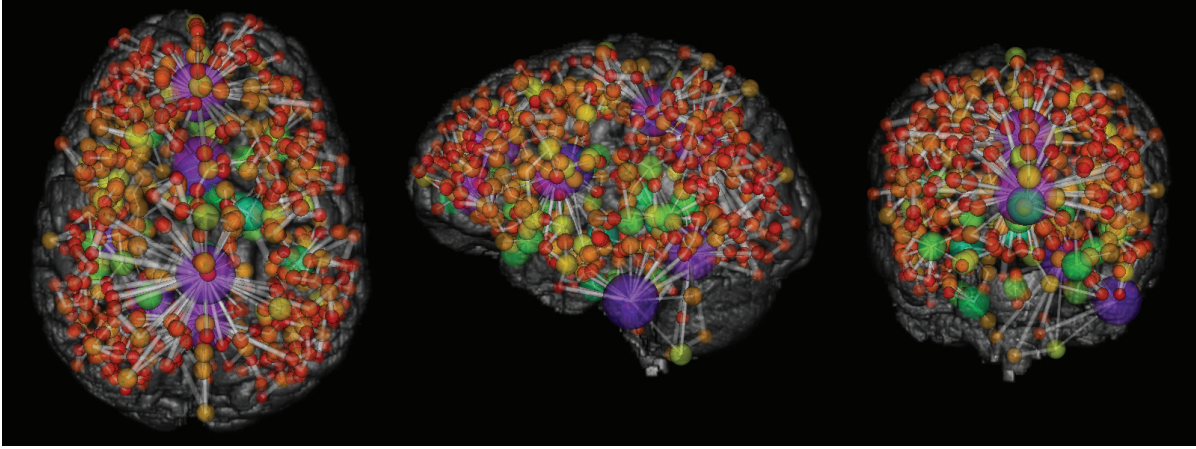


Figure 5.1 Illustration of first nearest neighbor graph between features in the scale-space based on the geometrical kernel $\mathcal{K}(g_i, g_j)$. Feature scales are represented by colors, from red for smaller ones to blue for larger ones.

might then be trained using local features to determine if they are more unique to their class (a general feature) or to their instance (a specific feature).

APPENDIX I

DIFFUSION ORIENTATIONS HISTOGRAMS (DOH) FOR DIFFUSION WEIGHTED IMAGE ANALYSIS

Laurent Chauvin^{a,b}, Kuldeep Kumar^a, Christian Desrosiers^a, Jacques De Guise^b, and
Matthew Toews^a

^a Laboratory for Imagery, Vision and Artificial Intelligence (LIVIA),
École de Technologie Supérieure, Montreal, Canada

^b Laboratoire de recherche en imagerie et orthopédie (LIO),
École de Technologie Supérieure, Montreal, Canada

Presented as poster in *Computational Diffusion MRI*, MICCAI 2017

Abstract

This paper proposes a novel keypoint descriptor for Diffusion Weighted Image (DWI) analysis, the Diffusion Orientation Histogram (DOH). The DOH descriptor quantizes local diffusion gradients into histograms over spatial location and orientation, in a manner analogous to the quantization of image gradients in the widely used Histogram of Oriented Gradients (HOG) technique. Diffusion gradient symmetry allows representing half of the orientation space at double the angular resolution, leading to a compact but highly informative descriptor. Quantitative preliminary experiments evaluate descriptors for the task of automatically identifying familial links (twins, non-twin siblings) from DWI keypoint correspondences. The DOH descriptor is found to be complementary to traditional HOG descriptors computed from scalar fractional anisotropy (FA) images, where concatenated DOH and HOG descriptors result in the highest rates of correct family member identification. Twin-twin descriptor correspondences are generally more concentrated about major white matter tracts, e.g. the internal capsule, in comparison to twin/non-twin sibling correspondences.

1 Introduction

Diffusion magnetic resonance imaging (dMRI) data offers the ability to observe patterns of brain connectivity in-vivo. dMRI data, particularly when sampled at high angular resolutions, is large and unwieldy, as diffusion information may be sampled over a large set of orientation directions. Large-scale Diffusion Weighted Images (DWI) analysis is thus typically performed based on reduced data representations for DWI, e.g. diffusion tensors (Le Bihan *et al.*, 2001; Basser *et al.*, 2000), tractography (Jahanshad *et al.*, 2011), etc.

In this paper we consider a potential alternative, the keypoint representation, where analysis focuses on a subset of salient, informative image local patches or features. Local feature information can be encoded into informative descriptors, and subsequently used in families of highly efficient analysis algorithms based on approximate nearest neighbors (NN), e.g. kernel density estimation (KDE). Approximate search routines can compute NN correspondences in $O(\log N)$ complexity in memory and computation time given N descriptors (Muja & Lowe, 2009), and thus scale gracefully to arbitrarily large data sets, as opposed to $O(N)$ for brute force/naive correspondence methods. Furthermore, NN methods come with important theoretical guarantees, e.g. they approach near optimal Bayes error rates as $N \rightarrow \infty$ (Cover & Hart, 1967).

Assuming salient DWI keypoints can be identified, how can DWI information be most effectively encoded? The computer vision literature shown descriptors based on histograms of orientated image gradients (HOG), e.g. SIFT (Lowe, 2004), HOG (Dalal & Triggs, 2005), BRIEF (Calonder *et al.*, 2010) to be among the most effective at achieving high correct correspondence rates. The HOG encoding is highly competitive with alternative encodings based on recent convolutional neural networks (CNN) (Dong & Soatto, 2015) (Zheng *et al.*, 2017).

In this work, we hypothesize that keypoints representing informative, local diffusion patterns can be identified and encoded for efficient DWI analysis, e.g. for tasks including registration, classification, etc. We propose a novel descriptor, the diffusion orientation histogram (DOH), that naturally extends the HOG descriptor to encode *diffusion gradient information*, as opposed to traditional image gradient information. We propose novel techniques for encoding diffusion

gradients, accounting for expected diffusion distance and diffusion orientation symmetry. Experiments demonstrate that DOH descriptors are effective and offer important, complementary information, in comparison to standard HOG descriptors encoded from scalar fractional anisotropy (FA) images.

2 Related Work

The DOH descriptor bridges two major fields of research, local keypoint methods and DWI analysis. Keypoint methods reduce large images into collections of local salient keypoints, that can be efficiently encoded for NN methods. The most effective descriptors have been based on local image gradient information (Lowe, 2004; Calonder *et al.*, 2010; Dalal & Triggs, 2005). They have been shown to be among the most effective for image correspondence tasks (Mikolajczyk & Schmid, 2005). They are the basis for the widely used SIFT descriptor (Lowe, 2004), which models image content in terms of multi-scale Gaussian derivatives (Lindeberg, 1994). HOG-like filters, i.e. localized oriented gradient filters, result from machine learning approaches such as independent component analysis (ICA) (Bell & Sejnowski, 1997) or CNNs (Krizhevsky *et al.*, 2012) when applied to natural images, and have been identified in the mammalian visual system, i.e. orientation-sensitive hypercolumns (Hubel & Wiesel, 1962).

In medical imaging, dMRI allows observation of water molecule diffusion in-vivo, locally and throughout the brain, and thus inference of neural connectivity patterns, as water tends to diffuse preferentially in the direction of axonal fibers (Le Bihan *et al.*, 1986). dMRI is thus a critical tool for in-vivo study of white matter geometry throughout the brain. A major challenge in large-scale DWI analysis, e.g. twin studies (Van Essen *et al.*, 2012; Jahanshad *et al.*, 2011), is coping with the sheer amount of directional data. Traditionally dMRI data are processed in terms of reduced data representations such scalar images (e.g. Fractional Anisotropy (FA), Mean Diffusivity) fiber tracts (Jahanshad *et al.*, 2011), diffusion tensors (DTI) (Le Bihan *et al.*, 2001; Basser *et al.*, 2000), etc.

3 Diffusion Orientation Histogram Descriptors

We propose encoding dMRI data in terms of Diffusion Orientation Histograms, a model-free natural extension of the ubiquitous HOG descriptor from computer vision. In this section we describe the computational framework for DWI and then the DOH encoding.

Diffusion Framework: Consider random variables of 3D spatial location $x \in R^3$, time $t \in R$, with corresponding displacement vectors Δx and Δt . Assuming temporally stationary diffusion and a constant diffusion time Δt , the diffusion of a water molecule can be modeled as a posterior density over a random displacement vector Δx conditioned on location x a time period Δt :

$$p(\Delta x, \Delta t | t, x) \propto p(\Delta x | \Delta t, x) \quad (\text{A I-1})$$

Equation (A I-1) is referred to as the Ensemble Average Propagator (EAP) (Descoteaux *et al.*, 2009) (or true diffusion propagator (Cory, 1990) under the *narrow pulse approximation*). Representing net 3D displacement vector $\Delta x = \{r, \angle\Theta\}$ in polar coordinates of magnitude $r = \|\Delta x\|$ and 2D orientation $\angle\Theta = \{\theta, \phi\}$, the EAP may be expressed using Bayes rule as:

$$p(\Delta x | \Delta t, x) = p(r, \angle\Theta | \Delta t, x) = p(r | \angle\Theta, \Delta t, x) p(\angle\Theta | \Delta t, x) \quad (\text{A I-2})$$

where in Equation (A I-2), the EAP is factored into conditional densities over orientation density $p(\angle\Theta | \Delta t, x)$ and magnitude $p(r | \angle\Theta, \Delta t, x)$ conditional on orientation $\angle\Theta$.

Without loss of generality, orientation density $p(\angle\Theta | \Delta t, x)$ is taken to be uniform, and the EAP is thus characterized by the conditional density of diffusion magnitude $p(r | \angle\Theta, \Delta t, x)$ along direction $\angle\Theta$. Adopting the Brownian motion model of diffusion (Einstein, 1905), this may be expressed as a unidimensional Gaussian density of the form

$$p(r | \angle\Theta, \Delta t, x) \propto \exp - \{r^2 / 4\Delta t ADC_{\angle\Theta}\} \quad (\text{A I-3})$$

centered on mean x with variance $2\Delta t \text{ ADC}_{\angle\Theta}$ (Alexander *et al.*, 2002), where $\text{ADC}_{\angle\Theta}$ is the apparent diffusion coefficient (ADC) measured along orientation $\angle\Theta$ in DWI. It is this density which we estimate and use to encode DWI gradient orientation information.

dmRI data comes in the form of images sampled at a set of diffusion-sensitizing gradient angles $\angle\Theta$. The DWI voxel intensity $S_{\angle\Theta}$ is proportional to the displacement of water molecules in the gradient direction $\angle\Theta$ applied during the acquisition, and the $\text{ADC}_{\angle\Theta}$ is estimated as

$$\text{ADC}_{\angle\Theta} = -\frac{1}{b} \ln \left(\frac{S_{\angle\Theta}}{S_0} \right) \quad (\text{A I-4})$$

where S_0 is the baseline signal without diffusion gradient and b is the b -value defined by the Stejskal-Tanner equation (Stejskal & Tanner, 1965) as $b = \gamma^2 G^2 \delta^2 (\Delta t - \frac{\delta}{3})$, with the gyromagnetic ratio γ , the sensitizing gradient pulse amplitude G , the gradient duration δ , and the diffusion time Δt between gradient pulses.

Diffusion Orientation Histogram (DOH) Descriptor: The DOH descriptor coarsely quantizes EAP information, i.e. $p(r|\angle\Theta, \Delta t, x)$, into histograms over spatial location x and orientation $\angle\Theta$ in a manner reminiscent of the widely used HOG descriptor. Intuitively, coarse quantization provides immunity to noise inherent to keypoint localization, e.g. small shifts or rotations.

Descriptors are computed at salient 3D keypoint regions defined by location and scale $\{x, \sigma\}$ within the image. We assume keypoints can be identified generic saliency operators in scale-space. Following the approach of (Toews *et al.*, 2013a), we experiment with a quantization structure of 8 spatial and 8 orientation bins for a 64-element descriptor. Given the symmetric nature of diffusion, we also consider a half-sphere structure with 8 approximately equally distributed orientations on the half-sphere. We refer to these sampling structures as full-sphere (FS) and half-sphere (HS).

The primary design question is how to encode gradient information in the form of diffusion coefficients $\text{ADC}_{\angle\Theta}$ given the EAP in Equation (A I-3), defined by the ratio $\frac{r^2}{\Delta t}$ of squared diffusion magnitude r and time Δt values, that may not be known. For this, we consider a soft

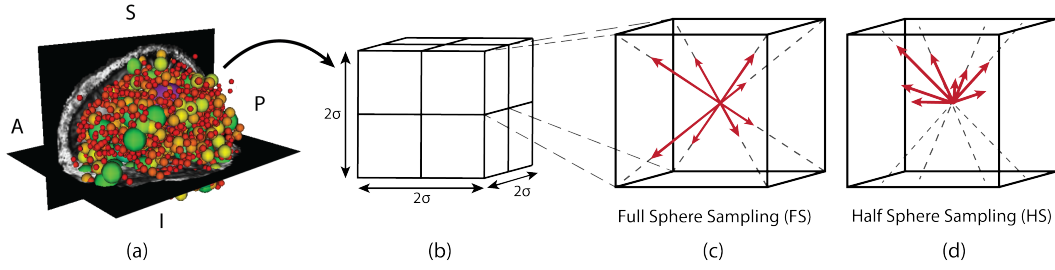


Figure-A I-1 The DOH descriptor is computed locally at keypoint locations, e.g. (a) keypoints detected in an FA volume. For each keypoint, local diffusion propagators are quantized into histograms of (b) 8 spatial bins and two 8-bin orientation sampling schemes: (c) full sphere (FS) and (d) half sphere (HS)

bin increment value based on the EAP in Equation (A I-3), which for an observed diffusion coefficient $ADC_{\angle\Theta}$ is

$$inc(ADC_{\angle\Theta}) \propto p(r|\angle\Theta, \Delta t, x). \quad (\text{A I-5})$$

4 Experiments

Experiments assess the utility of the DOH descriptor in feature matching experiments, where nearest neighbor (NN) inter-subject descriptor correspondences are used to automatically infer familial relationships based on a set of DWIs of siblings including twins and non-twin siblings. We use the Human Connectome Project (HCP) (Van Essen *et al.*, 2012) Q3 Release dMRI data, acquired on a Siemens Skyra 3T scanner, at isotropic $1.25mm$ resolution. For evaluation here, we consider a subset of 45 unique subject dMRI volumes from 15 families of 3 siblings each, including 2 monozygotic twins (T1, T2) and 1 non-twin sibling (NT), 23-35 years of age.

For each DWI, a fractional anisotropy (FA) is generated using the Dipy software package (Garyfalidis *et al.*, 2014). A set of 3D keypoints is then extracted as extrema of the difference-of-Gaussian scale-space (Toews *et al.*, 2013a) constructed from FA, and is used for all local descriptor evaluations. Keypoint extraction requires on the order of 20 seconds per image, each image results in approximately 2000 keypoints, see Figure I-1. Using a single set of keypoints ensures that descriptor comparisons are based on precisely the same underlying diffusion information.

To quantitatively assess descriptor performance, we consider the task of automatically recovering familial relationships from the structure of the inter-subject proximity graph, as proposed in (Toews & Wells III, 2016). Let $A = \{a_i\}$ and $B = \{b_j\}$ be sets of D -dimensional descriptors $a_i, b_j \in R^D$, where $|A| \neq |B|$. The distance between two sets of different cardinality $d(A, B)$ is then measured by their Jaccard distance, ranging from 0 to 1:

$$d(A, B) = J(A, B) = 1 - \frac{|A \cap B|}{|A \cup B|}, \quad (\text{A I-6})$$

where set intersection $|A \cap B|$ is defined by the number of NN descriptor correspondences between sets A and B . Note that the distance $d(a_i, b_j)$ between two descriptors a_i and b_i is defined by the Euclidean distance. In general, we expect the Jaccard distances from (A I-6) to reflect family relationships as siblings share more genetic material (especially twins) than unrelated subjects, such that:

$$d(T_1, T_2) < d(T_1, NTS) \approx d(T_2, NTS) < d(T_1, U) \approx d(T_2, U) \approx d(NTS, U) \quad (\text{A I-7})$$

T_1, T_2 and NTS being respectively twin and non-twin siblings of the same family, and U an unrelated subject from a different family. To assess the ability to recover family relationships based on sets of local features, the correct relationship recall rate is calculated for twin-twin γ_T and twin-non-twin γ_{NT} relationships separately, such as:

$$\gamma_T = \frac{|\{\Gamma_T\}|}{|\{\Omega_T\}|}, \gamma_{NT} = \frac{|\{\Gamma_{NT}\}|}{|\{\Omega_{NT}\}|}, \quad (\text{A I-8})$$

$\{\Gamma_T\}$ and $\{\Gamma_{NT}\}$ being respectively sets of correct Twin and Non-Twin recovered relationships based on local features, and $\{\Omega_T\}$ and $\{\Omega_{NT}\}$ respectively the sets of Twin and Non-Twin relationships in the whole dataset (here $\Omega_T = 15$ and $\Omega_{NT} = 30$).

We compare recall rates γ_T and γ_{NT} for five different keypoint descriptors (see Table I-1): 1) standard HOG descriptors computed from scalar FA, 2) DOH descriptors computed from DWI, 3) a concatenated descriptor combining HOG and DOH information 4) a DOH descriptor taking advantage of diffusion symmetry to sample half of the Orientation Distribution Function

Table-A I-1 Correct family relationship recovery rates for twins γ_T and non-twins γ_{NT} for five descriptor configurations

Config.	Type	Sampling	Orient. Bins	D	Relationship Recall	
					γ_T	γ_{NT}
1	FA	FS	8	64	0.90	0.40
2	DWI	FS	8	64	1.00	0.20
3	FA + DWI	FS + FS	8 + 8	128	1.00	0.33
4	DWI	HS	8	64	0.97	0.33
5	FA + DWI	FS + HS	8 + 8	128	1.00	0.40

(ODF) sphere (HS) at twice the sampling resolution and 5) a concatenated descriptor of 1) and 4). We expect more effective encodings to result in higher numbers of correct inter-subject correspondences, and thus inter-subject proximity graphs that more closely resemble the known family structure.

Results are shown in Table I-1, where best results are obtained using concatenated FA-HOG and DWI-DOH descriptors sampled over a half sphere. Figure I-3 illustrates the spatial distributions for NN feature correspondences. In row (a), DWI-DOH and FA-HOG descriptors lead to similar distributions of correspondences throughout the brain, and are thus generally complementary. In row (b), DWI correspondences are generally more concentrated in major white matter bundles for twins (b1) vs non-twins (b2,b3), this may indicate greater heritability. Finally, Figure I-2 show distributions of Jaccard distances for twins vs non-twins vs unrelated subject groups using descriptor configuration 5 (cf Table I-1), which generally reflect genotypical and phenotypical variations, as described in (A I-7).

5 Discussion

In this paper, we propose the diffusion orientation histogram (DOH) descriptor for DWI analysis, in a manner analogous to intensity gradient descriptors widely used in the computer vision literature. We provide the first preliminary keypoint-based analysis of sibling relationships from

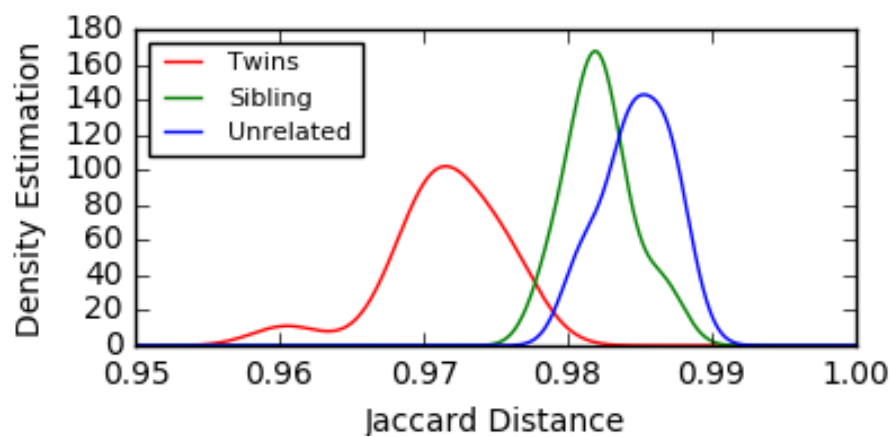


Figure-A I-2 Kernel density estimates of pairwise Jaccard distances for three groups of subject pairs: Twin, Sibling and Unrelated subject. Experimental results confirm theoretical expectations from (A I-7)

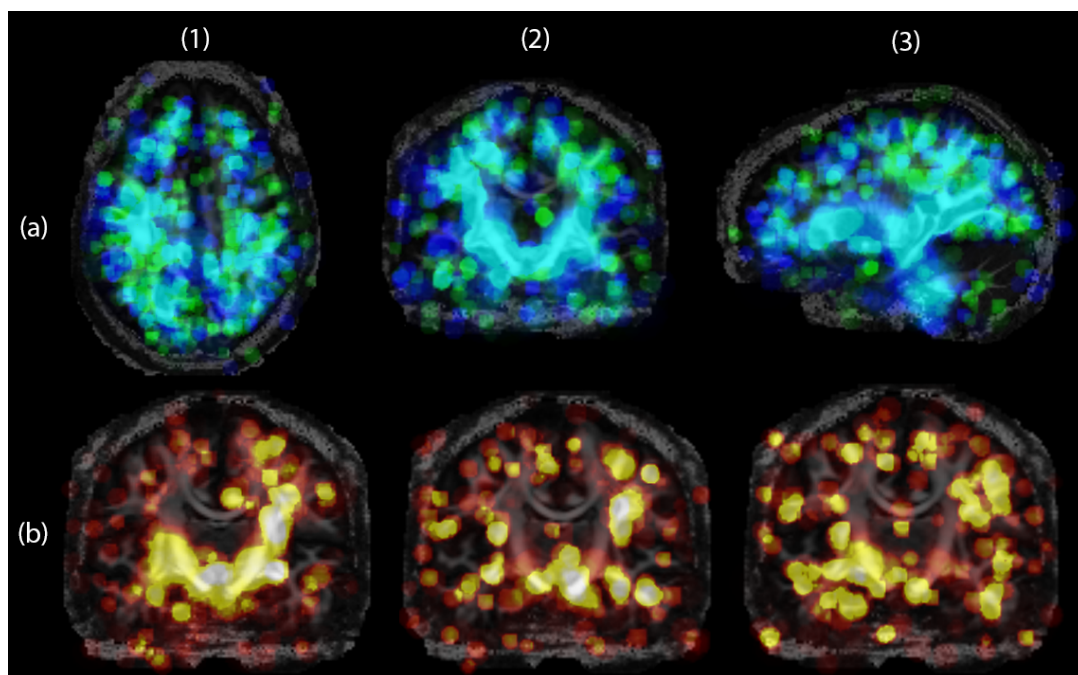


Figure-A I-3 Visualizing the spatial distributions correct NN feature correspondences. In row a), for DWI-DOH (green) vs FA-HOG descriptors (blue) in (a1-a3) axial, coronal and sagittal views. In row b), correspondences between twins T1-T2 in (b1) and non-twin siblings pairs in (b2) T1-NT (b3) T2-NT

DWI data, were results show the DOH encoding significantly augments scale FA feature analysis with complementary information, resulting in more effective identification of between family members based on local patterns of neural connectivity.

An important aspect is that of keypoint detection - this work adopts keypoints identified in scalar FA images in order to ensure a fair comparison of descriptors. Keypoints derived according to FA saliency are not necessarily optimal for full dMRI data, however, and future work will investigate salient feature detection methods specific to full dMRI data. The fact that DOH descriptors generally outperform FA-HOG descriptors here is thus a further indication of the effectiveness of the encoding. Experiments here are limited to a subset of HCP data with a specific family structure, future work will apply DOH to the entire dataset. We evaluate the DOH representation at sparse keypoints, DOH descriptor may prove a useful representation for dense image analysis.

It should also be noted that by design, the DOH descriptor also have the ability to compare diffusion data acquired under different protocols (e.g. cross-site, cross-scanner), as experimental preliminary results show robustness to variable q-space sampling scheme.

APPENDIX II

LAPLACIAN-OF-GAUSSIAN PARTICLE THEORY

Our work is based on Laplacian-of-Gaussian, here we emphasize the connection between fields and to scale-space. A single particle may be defined as a complex-valued wave function $I(\mathbf{x}, t)$ in space time $\mathbf{x}, t$ satisfying the Schrodinger Equation ([Schrödinger, 1926](#)):

$$\frac{i\partial I(\mathbf{x}, t)}{\partial t} = D\nabla^2 I(\mathbf{x}, t) \quad (\text{A II-1})$$

where $\frac{\partial I(\mathbf{x}, t)}{\partial t}$ is the partial derivative of $I(\mathbf{x}, t)$ with respect to time, $\nabla^2 I(\mathbf{x}, t) = \frac{d^2 I}{dx^2} + \frac{d^2 I}{dy^2} + \frac{d^2 I}{dz^2}$ is the Laplacian operator over spatial coordinates $\{x, y, z\}$, $i = \sqrt{-1}$ is the imaginary unit and D is a constant. In the case of a single real 3D image $I(\mathbf{x})$, the partial derivative $\partial I(\mathbf{x}, t)$ in Equation (A II-1) may not be evaluated. However the assumption of a passive diffusion process may be used to model the evolution of the image function over time t , and is expressed via the 3D heat or diffusion equation (Brownian motion ([Einstein, 1905](#))), equivalent to Equation (A II-1) without i and where D is a constant diffusion rate. The solution is the Gaussian function where the variance is proportional to t :

$$I(\mathbf{x}, t) = \frac{1}{\sqrt{(4\pi Dt)^3}} \exp\left(-\frac{\|\mathbf{x} - \mathbf{u}\|^2}{4Dt}\right) \quad (\text{A II-2})$$

The scale-space $I(\mathbf{x}, \sigma)$ may thus thought of as simulating the isotropic diffusion of image intensity information $I(\mathbf{x})$ for a time period $t = \sigma^2$, where the intensity $I(\mathbf{x}, \sigma)$ represents the image content according to a receptive field of spatial extent proportional to scale σ .

The isotropic Gaussian scale-space $I(\mathbf{x}, \sigma)$ is then defined as:

$$I(\mathbf{x}, \sigma) = I(\mathbf{x}) * N(\mathbf{x}, \sigma) \quad (\text{A II-3})$$

by convolution of the image $I(\mathbf{x})$ with Gaussian kernel $N(\mathbf{x}, \sigma) \propto e^{-(x^2+y^2+z^2)/2\sigma^2}$ with variance σ^2 .

The Laplacian-of-Gaussian operator is defined as follows:

$$\begin{aligned}\nabla^2 I(\mathbf{x}, \sigma) &= \frac{\partial^2}{\partial x^2} I(\mathbf{x}, \sigma) + \frac{\partial^2}{\partial y^2} I(\mathbf{x}, \sigma) + \frac{\partial^2}{\partial z^2} I(\mathbf{x}, \sigma) \\ &= \frac{x^2 + y^2 + z^2 - 3\sigma^2}{\sigma^4} e^{-(x^2+y^2+z^2)/2\sigma^2}\end{aligned}\quad (\text{A II-4})$$

The Laplacian-of-Gaussian operator may be approximated by the difference-of-Gaussian operator:

$$\nabla^2 I(\mathbf{x}, \sigma) \approx I(\mathbf{x}, \sigma) - I(\mathbf{x}, \kappa\sigma)$$

where κ is a multiplicative difference in scale sampling.

In digital image processing, early work by Marr and Hildreth used the Laplacian-of-Gaussian as an edge detector via zero crossings (Marr & Hildreth, 1980). Witkin in 1D audio signals (Witkin, 1984) and Koenderink in 2D image processing (Koenderink, 1984) demonstrated that a scale-space $I(\mathbf{x}, \sigma)$ may be constructed via convolution of the image $I(\mathbf{x}, \sigma) = I(\mathbf{x}) * G(\sigma)$ with an isotropic Gaussian kernel $G(\sigma) \propto \exp - \|\mathbf{x}\|^2 / 2\sigma^2$ of standard deviation σ . Burt described the Laplacian pyramid, where the Laplacian-of-Gaussian (LoG) function may be evaluated via a difference-of-Gaussian (DoG) (Burt & Adelson, 1987). Lindeberg showed the Gaussian kernel was uniquely consistent with scale-space axioms (Lindeberg, 1998) including non creation or enhancement of local image maxima, and thus served as a mechanism for identifying salient image features in a manner invariant to image rotation, translation and scaling. Lowe proposed identifying Difference-of-Gaussian maxima as distinctive image keypoints (Lowe, 2004), which identified distinctive keypoints which maximize the Laplacian-of-Gaussian operator $\nabla^2 I(\mathbf{x}, \sigma)$. Note also that center-surround filters reminiscent of the Laplacian are found in the receptive field of the mammalian visual cortex (Hubel & Wiesel, 1962).

APPENDIX III

CPD REGISTRATION DETAILS

In the case of rigid transform, the *solve* function is defined for rigid and affine transforms as (Myronenko & Song, 2010):

Algorithm-A III-1 Solve for rigid transform

```

1 solve_rigid( $\mathcal{F}, \mathcal{M}, P$ ):
2    $N_{\mathbf{P}} = \mathbf{1}^T \mathbf{P} \mathbf{1}$ ,  $\mu_{\mathcal{F}} = \frac{1}{N_{\mathbf{P}}} \mathcal{F}^T \mathbf{P}^T \mathbf{1}$ ,  $\mu_{\mathcal{M}} = \frac{1}{N_{\mathbf{P}}} \mathcal{M}^T \mathbf{P} \mathbf{1}$ ,
3    $\hat{\mathcal{F}} = \mathcal{F} - \mathbf{1} \mu_{\mathcal{F}}^T$ ,  $\hat{\mathcal{M}} = \mathcal{M} - \mathbf{1} \mu_{\mathcal{M}}^T$ ,
4    $\mathbf{A} = \hat{\mathcal{F}}^T \mathbf{P}^T \hat{\mathcal{M}}$ , compute SVD of  $\mathbf{A}$  such as  $\mathbf{A} = \mathbf{U} \mathbf{S} \mathbf{V}^T$ ,
5    $\mathbf{R} = \mathbf{U} \mathbf{C} \mathbf{V}^T$ , where  $\mathbf{C} = \text{d}(1, \dots, 1, \det(\mathbf{U} \mathbf{V}^T))$ ,
6    $s = \frac{\text{tr}(\mathbf{A}^T \mathbf{R})}{\text{tr}(\hat{\mathcal{M}}^T \text{d}(\mathbf{P} \mathbf{1}) \hat{\mathcal{M}})}$ ,
7    $\mathbf{t} = \mu_{\mathcal{F}} - \mathbf{s} \mathbf{R} \mu_{\mathcal{M}}$ ,
8    $\sigma^2 = \frac{1}{N_{\mathbf{P}} D} (\text{tr}(\hat{\mathcal{F}}^T \text{d}(\mathbf{P}^T \mathbf{1}) \hat{\mathcal{F}}) - s \text{tr}(\mathbf{A}^T \mathbf{R}))$ .
9   return  $\{s, \mathbf{R}, \mathbf{t}\}, \sigma^2$ 

```

Algorithm-A III-2 Solve for affine transform

```

1 solve_affine( $\mathcal{F}, \mathcal{M}, P$ ):
2    $N_{\mathbf{P}} = \mathbf{1}^T \mathbf{P} \mathbf{1}$ ,  $\mu_{\mathcal{F}} = \frac{1}{N_{\mathbf{P}}} \mathcal{F}^T \mathbf{P}^T \mathbf{1}$ ,  $\mu_{\mathcal{M}} = \frac{1}{N_{\mathbf{P}}} \mathcal{M}^T \mathbf{P} \mathbf{1}$ ,
3    $\hat{\mathcal{F}} = \mathcal{F} - \mathbf{1} \mu_{\mathcal{F}}^T$ ,  $\hat{\mathcal{M}} = \mathcal{M} - \mathbf{1} \mu_{\mathcal{M}}^T$ ,
4    $\mathbf{B} = (\hat{\mathcal{F}}^T \mathbf{P}^T \hat{\mathcal{M}})(\mathcal{M}^T \text{d}(\mathbf{P} \mathbf{1}) \hat{\mathcal{M}})^{-1}$ ,
5    $\mathbf{t} = \mu_{\mathcal{F}} - \mathbf{B} \mu_{\mathcal{M}}$ ,
6    $\sigma^2 = \frac{1}{N_{\mathbf{P}} D} (\text{tr}(\hat{\mathcal{F}}^T \text{d}(\mathbf{P}^T \mathbf{1}) \hat{\mathcal{F}}) - \text{tr}(\hat{\mathcal{F}}^T \mathbf{P}^T \hat{\mathcal{M}} \mathbf{B}^T))$ .
7   return  $\{\mathbf{B}, \mathbf{t}\}, \sigma^2$ 

```

BIBLIOGRAPHY

- Ahmad, B., Usama, M., Lu, J., Xiao, W., Wan, J. and Yang, J. (2019). Deep Convolutional Neural Network Using Triplet Loss to Distinguish the Identical Twins. *2019 IEEE Globecom Workshops*, pp. 1–6.
- Alcantarilla, P.F., Bartoli, A. and Davison, A.J. (2012). KAZE Features. *Computer Vision – ECCV 2012*, (Lecture Notes in Computer Science), 214–227.
- Alexander, A.L., Hasan, K.M., Lazar, M., Tsuruda, J.S. and Parker, D.L. (2001). Analysis of partial volume effects in diffusion-tensor MRI. *Magnetic Resonance in Medicine*, 45(5), 770–780.
- Alexander, D., Barker, G. and Arridge, S. (2002). Detection and modeling of non-Gaussian apparent diffusion coefficient profiles in human brain data. *Magnetic Resonance in Medicine*, 48(2), 331–340.
- Allaire, S., Kim, J.J., Breen, S.L.S., Jaffray, D.A. and Pekar, V. (2008). Full orientation invariance and improved feature selectivity of 3D SIFT with application to medical image analysis. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1–8.
- Amico, E. and Goñi, J.J. (2018). The quest for identifiability in human functional connectomes. *Scientific reports*, 8(1), 8254.
- Annas, G.J. and Elias, S. (2014). 23andMe and the FDA. *New England Journal of Medicine*, 370(11), 985–988.
- Arandjelović, R. and Zisserman, A. (2012). Three things everyone should know to improve object retrieval. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2911–2918.
- Armstrong, B.C., Ruiz-Blondet, M.V., Khalifian, N., Kurtz, K.J., Jin, Z. and Laszlo, S. (2015). Brainprint: Assessing the uniqueness, collectability, and permanence of a novel method for ERP biometrics. *Neurocomputing*, 166, 59–67.
- Arsigny, V., Commowick, O., Ayache, N. and Pennec, X. (2009). A Fast and Log-Euclidean Polyaffine Framework for Locally Linear Registration. *Journal of Mathematical Imaging and Vision*, 33(2), 222–238.
- Ashburner, J. and Friston, K.J. (2000). Voxel-based morphometry—the methods. *Neuroimage*, 11(6), 805–821.

- Auton, A., Donnelly, P., Eichler, E.E., Flicek, P., Gabriel, S.B., Gibbs, R.A., Green, E.D., Hurles, M.E., Knoppers, B.M., Korbel, J.O., Lander, E.S., Lee, C., Lehrach, H., Mardis, E.R., Marth, G.T. et al. (2015). A global reference for human genetic variation. *Nature*, 526(7571).
- Bajusz, D., Rácz, A. and Héberger, K. (2015). Why is Tanimoto index an appropriate choice for fingerprint-based similarity calculations? *Journal of Cheminformatics*, 7(1).
- Balntas, V., Lenc, K., Vedaldi, A. and Mikolajczyk, K. (2017). HPatches: A benchmark and evaluation of handcrafted and learned local descriptors. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5173–5182.
- Bardes, A., Ponce, J. and LeCun, Y. (2022). VICReg: Variance-Invariance-Covariance Regularization for Self-Supervised Learning. *arXiv:2105.04906 [cs]*.
- Basser, P.J., Mattiello, J. and LeBihan, D. (1994). MR diffusion tensor spectroscopy and imaging. *Biophysical journal*, 66(1), 259–267.
- Basser, P., Pajevic, S. and Pierpaoli, C. (2000). In vivo fiber tractography using DT-MRI data. *Magnetic Resonance in Medicine*.
- Baur, C., Denner, S., Wiestler, B., Navab, N. and Albarqouni, S. (2021). Autoencoders for unsupervised anomaly segmentation in brain MR images: A comparative study. *Medical Image Analysis (MIA)*, 69.
- Bay, H., Ess, A., Tuytelaars, T. and Van Gool, L. (2008). Speeded-up robust features (SURF). *Computer vision and image understanding*, 110(3), 346–359.
- Bell, A.J. and Sejnowski, T.J. (1997). The “independent components” of natural scenes are edge filters. *Vision Research*, 37(23), 3327–3338.
- Bellavia, F. and Colombo, C. (2020). Is there anything new to say about SIFT matching? *International Journal of Computer Vision*, 128(7).
- Belliveau, J., Kennedy, D., McKinstry, R., Buchbinder, B., Weisskoff, R., Cohen, M., Vevea, J., Brady, T. and Rosen, B. (1991). Functional mapping of the human visual cortex by magnetic resonance imaging. *Science*, 254(5032), 716–719.
- Bender, E.M., Gebru, T., McMillan-Major, A. and Shmitchell, S. (2021). On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, (FAccT ’21), 610–623.
- Bersvendsen, J., Toews, M., Danudibroto, A., Wells III, W.M., Urheim, S., Estépar, R.S.J. and Samset, E. (2016). Robust spatio-temporal registration of 4D cardiac ultrasound sequences.

- Medical imaging 2016: Ultrasonic imaging and tomography*, 9790, 122–128.
- Besl, P.J. and McKay, N.D. (1992). Method for registration of 3-D shapes. *Sensor Fusion IV: Control Paradigms and Data Structures*, 1611, 586–606.
- Bingjian, W., Quan, L., Yapeng, L., Fan, L., Liping, B., Gang, L. and Rui, L. (2011). Image registration method for multimodal images. *Applied Optics*, Vol. 50, Issue 13, pp. 1861–1867, 50(13), 1861–1867.
- Blair, J.R.S. and Peyton, B. (1993). An Introduction to Chordal Graphs and Clique Trees. *Graph Theory and Sparse Matrix Computation*, (The IMA Volumes in Mathematics and its Applications), 1–29.
- Bloem-Reddy, B. and Teh, Y.W. (2020). Probabilistic symmetries and invariant neural networks. *Journal of Machine Learning Research*, 21, 90–1.
- Boiman, O., Shechtman, E. and Irani, M. (2008). In defense of nearest-neighbor based image classification. *2008 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–8.
- Borchert, M.J., Devlin, J.A., Erlewein, S.R., Fleck, M., Harrington, J.A., Higuchi, T., Latacz, B.M., Voelksen, F., Wursten, E.J., Abbass, F., Bohman, M.A., Mooser, A.H., Popper, D., Wiesinger, M., Will, C. et al. (2022). A 16-parts-per-trillion measurement of the antiproton-to-proton charge–mass ratio. *Nature*, 601(7891), 53–57.
- Brin, S. (1995). Near neighbor search in large metric spaces.
- Brosch, T., Gmbh, P., Technologies, I. and Tam, R. (2013). Manifold Learning of Brain MRIs by Deep Learning Manifold Learning of Brain MRIs. *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 16(September).
- Buolamwini, J. and Gebru, T. (2018). Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. *Proceedings of the Conference on Fairness, Accountability and Transparency*, pp. 77–91.
- Burt, P.J. and Adelson, E.H. (1987). The Laplacian Pyramid as a Compact Image Code. In Fischler, M.A. and Firschein, O. (Eds.), *Readings in Computer Vision* (pp. 671–679).
- Calonder, M., Lepetit, V., Strecha, C. and Fua, P. (2010). Brief: Binary robust independent elementary features. *European conference on computer vision*, pp. 778–792.
- Calonder, M., Lepetit, V., Ozuysal, M., Trzcinski, T., Strecha, C. and Fua, P. (2011). BRIEF: Computing a local binary descriptor very fast. *IEEE transactions on pattern analysis and machine intelligence*, 34(7), 1281–1298.

- Carlier, J.B., Chauvin, L., Luo, J., Wells III, W.M., Machado, I., Harmouche, R. and Toews, M. (2021). GPU optimization of the 3D Scale-invariant Feature Transform Algorithm and a Novel BRIEF-inspired 3D Fast Descriptor. *arXiv:2112.10258 [cs]*.
- Castellani, U., Rossato, E., Murino, V., Bellani, M., Rambaldelli, G., Perlini, C., Tomelleri, L., Tansella, M. and Brambilla, P. (2012). Classification of schizophrenia using feature-based morphometry. *Journal of Neural Transmission*, 119(3), 395–404.
- Chauvin, L., Kumar, K., Desrosiers, C., Guise, J.D. and Toews, M. (2018). Diffusion Orientation Histograms (DOH) for Diffusion Weighted Image Analysis. *Computational Diffusion MRI*, pp. 91–99.
- Chauvin, L., Kumar, K., Desrosiers, C., De Guise, J., Wells, W. and Toews, M. (2019). Analyzing Brain Morphology on the Bag-of-Features Manifold. *International Conference on Information Processing in Medical Imaging (IPMI)*, 11492 LNCS, 45–56.
- Chauvin, L., Kumar, K., Wachinger, C., Vangel, M., de Guise, J., Desrosiers, C., Wells, W., Toews, M., Initiative, A.D.N., others and Alzheimer’s Disease Neuroimaging Initiative (ADNI). (2020). Neuroimage signature from salient keypoints is highly specific to individuals and shared by close relatives. *NeuroImage*, 204(20), 116–208.
- Chauvin, L., Kumar, K., Desrosiers, C., Wells, W. and Toews, M. (2021). Efficient Pairwise Neuroimage Analysis using the Soft Jaccard Index and 3D Keypoint Sets. *IEEE Transactions on Medical Imaging*.
- Chen, J. and Tian, J. (2009). Real-time multi-modal rigid registration based on a novel symmetric-SIFT descriptor. *Progress in Natural Science*, 19(5), 643–651.
- Chen, S. and Hu, X. (2018). Individual Identification Using the Functional Brain Fingerprint Detected by the Recurrent Neural Network. *Brain Connectivity*, 8(4), brain.2017.0561.
- Chen, Y., Storrs, J., Tan, L., Mazlack, L.J., Lee, J.H. and Lu, L.J. (2014). Detecting brain structural changes as biomarker from magnetic resonance images using a local feature based SVM approach. *Journal of neuroscience methods*, 221, 22–31.
- Cheung, W. and Hamarneh, G. (2009). n-SIFT: n-Dimensional Scale Invariant Feature Transform. *IEEE Transactions on Image Processing*, 18(9), 2012–2021.
- Christenson, J.H., Cronin, J.W., Fitch, V.L. and Turlay, R. (1964). Evidence for the 2π Decay of the K^0 Meson. *Physical Review Letters*, 13(4), 138–140.
- Chum, O., Philbin, J., Zisserman, A. and others. (2008). Near duplicate image detection: Min-hash and TF-IDF weighting. *The British Machine Vision Conference (BMVC)*, 810,

812–815.

- Cohen, T., Weiler, M., Kicanaoglu, B. and Welling, M. (2019). Gauge Equivariant Convolutional Networks and the Icosahedral CNN. *Proceedings of the 36th International Conference on Machine Learning*, pp. 1321–1330.
- Colclough, G.L., Smith, S.M., Nichols, T.E., Winkler, A.M., Sotiropoulos, S.N., Glasser, M.F., Van Essen, D.C. and Woolrich, M.W. (2017). The heritability of multi-modal connectivity in human brain activity. *eLife*, 6, e20178.
- Comaniciu, D. and Meer, P. (2002). Mean shift: a robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(5), 603–619.
- Cory, D. (1990). Measurement of translational displacement probabilities by NMR: an indicator of compartmentation. *Magnetic resonance in medicine*, 14(3), 435–444.
- Cover, T. and Hart, P. (1967). Nearest neighbor pattern classification. *IEEE Transactions on Information Theory (TIT)*, 13(1), 21–27.
- Cronin, J.W. (1981). CP symmetry violation—the search for its origin. *Reviews of Modern Physics*, 53(3), 373–383.
- Dai, Q., Carr, P., Sigal, L. and Hoiem, D. (2015). Family Member Identification from Photo Collections. *2015 IEEE Winter Conference on Applications of Computer Vision*, pp. 982–989.
- Dalal, N. and Triggs, B. (2005). Histograms of oriented gradients for human detection. *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, I, 886–893.
- Dempster, A.P., Laird, N.M. and Rubin, D.B. (1977). Maximum Likelihood from Incomplete Data Via the EM Algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(1), 1–22.
- Descoteaux, M., Deriche, R., Bihan, D.L., Mangin, J.F. and Poupon, C. (2009). Diffusion propagator imaging: using Laplace’s equation and multiple shell acquisitions to reconstruct the diffusion propagator. *International conference on information processing in medical imaging*, pp. 1–13.
- Desikan, R.S., Cabral, H.J., Hess, C.P., Dillon, W.P., Glastonbury, C.M., Weiner, M.W., Schmansky, N.J., Greve, D.N., Salat, D.H., Buckner, R.L., Fischl, B. and others. (2009). Automated MRI measures identify individuals with mild cognitive impairment and Alzheimer’s disease. *Brain*, 132(8), 2048–2057.

- Detone, D., Malisiewicz, T. and Rabinovich, A. (2018). SuperPoint: Self-Supervised Interest Point Detection and Description. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pp. 224–236.
- DiMaio, F.P., Soni, A.B., Phillips, G.N. and Shavlik, J.W. (2009). Spherical-harmonic decomposition for molecular recognition in electron-density maps. *International Journal of Data Mining and Bioinformatics*, 3(2), 205–227.
- Dong, J. and Soatto, S. (2015). Domain-size pooling in local descriptors: DSP-SIFT. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5097–5106.
- Einstein, A. (1905). Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen. *Annalen der physik*.
- Eliot, L., Ahmed, A., Khan, H. and Patel, J. (2021). Dump the “dimorphism”: Comprehensive synthesis of human brain studies reveals few male-female differences beyond size. *Neuroscience & Biobehavioral Reviews*.
- Esteves, C., Allen-Blanchette, C., Makadia, A. and Daniilidis, K. (2018). Learning SO(3) equivariant representations with spherical CNNs. *Proceedings of the european conference on computer vision (ECCV)*, pp. 52–68.
- Fedorov, A., Beichel, R., Kalpathy-Cramer, J., Finet, J., Fillion-Robin, J.C., Pujol, S., Bauer, C., Jennings, D., Fennessy, F., Sonka, M. and others. (2012). 3D Slicer as an image computing platform for the Quantitative Imaging Network. *Magnetic resonance imaging*, 30(9), 1323–1341.
- Fetit, A.E., Cupitt, J., Kart, T. and Rueckert, D. (2020). Training deep segmentation networks on texture-encoded input: application to neuroimaging of the developing neonatal brain. *Proceedings of the Third Conference on Medical Imaging with Deep Learning*, pp. 230–240.
- Finn, E.S. and Rosenberg, M.D. (2021). Beyond fingerprinting: Choosing predictive connectomes over reliable connectomes. *NeuroImage*, 239, 118254.
- Finn, E.S., Shen, X., Scheinost, D., Rosenberg, M.D., Huang, J., Chun, M.M., Papademetris, X. and Constable, R.T. (2015). Functional connectome fingerprinting: identifying individuals using patterns of brain connectivity. *Nature neuroscience*, 18(11), 1664–71.
- Fischl, B. (2012). FreeSurfer. *Neuroimage*, 62(2), 774–781.
- Flitton, G.T., Breckon, T.P. and Bouallagu, N.M. (2010). Object Recognition using 3D SIFT in Complex CT Volumes. *The British Machine Vision Conference (BMVC)*, pp. 11.1–11.12.

- Florack, L.M.J., ter Haar Romeny, B.M., Koenderink, J.J. and Viergever, M.A. (1992). Scale and the differential structure of images. *Image and Vision Computing*, 10(6), 376–388.
- Franz, A., Carlsen, I.C. and Renisch, S. (2006). An adaptive irregular grid approach using SIFT features for elastic medical image registration. In *Bildverarbeitung für die medizin 2006* (pp. 201–205). Springer.
- Friskén, S., Luo, M., Machado, I., Unadkat, P., Juvekar, P., Bunevicius, A., Toews, M., Wells, W.M., Miga, M.I. and Golby, A.J. (2019). Preliminary results comparing thin-plate splines with finite element methods for modeling brain deformation during neurosurgery using intraoperative ultrasound. *Medical Imaging 2019: Image-Guided Procedures, Robotic Interventions, and Modeling*, 10951, 1095120.
- Friskén, S., Luo, M., Juvekar, P., Bunevicius, A., Machado, I., Unadkat, P., Bertotti, M.M., Toews, M., Wells, W.M., Miga, M.I., Golby, A.J. and others. (2020). A comparison of thin-plate spline deformation and finite element modeling to compensate for brain shift during tumor resection. *International Journal of Computer Assisted Radiology and Surgery*, 15(1), 75–85.
- Gardner, A., Kanno, J., Duncan, C.A. and Selmic, R. (2014). Measuring Distance Between Unordered Sets of Different Sizes. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 137–143.
- Garyfallidis, E., Brett, M., Bagrat Amirbekian, B., Rokem, A., van der Walt, S., Descoteaux, M. and Nimmo-Smith, I. (2014). Dipy, a library for the analysis of diffusion MRI data. *Frontiers in Neuroinformatics*, 8(8).
- Geirhos, R., Rubisch, P., Michaelis, C., Bethge, M., Wichmann, F.A. and Brendel, W. (2019). ImageNet-trained CNNs are biased towards texture; increasing shape bias improves accuracy and robustness. *arXiv:1811.12231 [cs, q-bio, stat]*.
- Georgescu, I. (2020). CP violation and beyond. *Nature Reviews Physics*, 2(3), 120–122.
- Gerber, S., Tasdizen, T., Thomas Fletcher, P., Joshi, S., Whitaker, R. and Alzheimer's Disease Neuroimaging Initiative (ADNI). (2010). Manifold modeling for brain population analysis. *Medical Image Analysis (MIA)*, 14(5), 643–53.
- Gill, G., Toews, M. and Beichel, R.R. (2014). Robust initialization of active shape models for lung segmentation in CT scans: a feature-based atlas approach. *Journal of Biomedical Imaging*, 2014, 13.
- Glasser, M.F., Sotiropoulos, S.N., Wilson, J.A., Coalson, T.S., Fischl, B., Andersson, J.L., Xu, J., Jbabdi, S., Webster, M., Polimeni, J.R., Van Essen, D.C. and Jenkinson, M. (2013).

- The minimal preprocessing pipelines for the Human Connectome Project. *NeuroImage*, 80, 105–124.
- Golyanik, V., Ali, S.A. and Stricker, D. (2016). Gravitational approach for point set registration. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5802–5810.
- Gopinath, K. (2021). *Geometrical learning of brain surface data*. (PhD Thesis, École de Technologie Supérieure).
- Gopinath, K., Desrosiers, C. and Lombaert, H. (2022). Learnable Pooling in Graph Convolutional Networks for Brain Surface Analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(2), 864–876.
- Gordon, E.M., Laumann, T.O., Adeyemo, B., Gilmore, A.W., Nelson, S.M., Dosenbach, N.U.F. and Petersen, S.E. (2017). Individual-specific features of brain systems identified with resting state functional correlations. *NeuroImage*, 146, 918–939.
- Guo, C., Pleiss, G., Sun, Y. and Weinberger, K.Q. (2017). On calibration of modern neural networks. *International Conference on Machine Learning*, pp. 1321–1330.
- Hamburg, M.A. and Collins, F.S. (2010). The Path to Personalized Medicine. *New England Journal of Medicine*, 363(4), 301–304.
- Hardoon, D.R., Mourão-Miranda, J., Brammer, M. and Shawe-Taylor, J. (2007). Unsupervised analysis of fMRI data using kernel canonical correlation. *NeuroImage*, 37(4), 1250–1259.
- Harris, C., Stephens, M. and others. (1988). A combined corner and edge detector. *Alvey vision conference*, 15, 10–5244.
- HCP. (2017). 1200 Subjects Data Release - Connectome.
- Hermann, K., Chen, T. and Kornblith, S. (2020). The Origins and Prevalence of Texture Bias in Convolutional Neural Networks. *Advances in Neural Information Processing Systems*, 33, 19000–19015.
- Hossain, M.T. (2012). *An effective technique for multi-modal image registration*. (Ph.D. thesis, Monash University).
- Hossain, M.T., Teng, S.W., Lu, G. and Lackmann, M. (2010). An enhancement to SIFT-based techniques for image registration. *Proceedings - 2010 Digital Image Computing: Techniques and Applications, DICTA 2010*, 166–171.

- Hossain, M.T., Lv, G., Teng, S.W., Lu, G. and Lackmann, M. (2011). Improved symmetric-sift for multi-modal image registration. *2011 international conference on digital image computing: techniques and applications*, pp. 197–202.
- Hough, P.V. (1959). Machine Analysis of Bubble Chamber Pictures. *Proceedings of the international conference on High Energy Accelerators and Instrumentation*, pp. 554–556.
- Hu, J., Luo, Z., Wang, X., Sun, S., Yin, Y., Cao, K., Song, Q., Lyu, S. and Wu, X. (2021). End-to-end multimodal image registration via reinforcement learning. *Medical Image Analysis*, 68, 101878.
- Hubel, D.H. and Wiesel, T.N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of physiology*, 160(1), 106–154.
- Hubel, D.H. and Wiesel, T.N. (1968). Receptive fields and functional architecture of monkey striate cortex. *The Journal of physiology*, 195(1), 215–243.
- Jack Jr, C.R., Bernstein, M.A., Fox, N.C.N., Thompson, P., Alexander, G., Harvey, D., Borowski, B., Britson, P.J., L. Whitwell, J., Ward, C., others, Jack, C.R.C., Bernstein, M.A., Fox, N.C.N., Thompson, P. et al. (2008). The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods. *Journal of the International Society for Magnetic Resonance in Medicine (ISMRM)*, 27(4), 685–691.
- Jahanshad, N., Aganj, I., Lenglet, C., Joshi, A., Jin, Y., Barysheva, M., McMahon, K.L., De Zubizaray, G.I., Martin, N.G., Wright, M.J. and others. (2011). Sex differences in the human connectome: 4-Tesla high angular resolution diffusion imaging (HARDI) tractography in 234 young adult twins. *2011 IEEE international symposium on biomedical imaging: From nano to macro*, pp. 939–943.
- Jauer, P., Kuhleemann, I., Bruder, R., Schweikard, A. and Ernst, F. (2019). Efficient registration of high-resolution feature enhanced point clouds. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(5), 1102–1115.
- Jenkinson, M., Bannister, P., Brady, M. and Smith, S. (2002). Improved Optimization for the Robust and Accurate Linear Registration and Motion Correction of Brain Images. *NeuroImage*, 825–841.
- Jiang, W., Trulls, E., Hosang, J., Tagliasacchi, A. and Yi, K.M. (2021). COTR: Correspondence Transformer for Matching Across Images. *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 6207–6217.
- Jiao, Z., Liu, R., Yi, P. and Zhou, D. (2019). A Point Cloud Registration Algorithm Based on 3D-SIFT. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial*

- Intelligence and Lecture Notes in Bioinformatics*), 11345 LNCS, 24–31.
- Juefei-Xu, F. and Savvides, M. (2013). An augmented linear discriminant analysis approach for identifying identical twins with the aid of facial asymmetry features. *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pp. 56–63.
- Kadir, T. and Brady, M. (2001). Saliency, Scale and Image Description. *International Journal of Computer Vision*, 45(2), 83–105.
- Kelman, A., Sofka, M. and Stewart, C.V. (2007). Keypoint descriptors for matching across multiple image modalities and non-linear intensity variations. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1–7.
- Kiley, M.R. and Hossain, M.S. (2020). Who are My Family Members? A Solution Based on Image Processing and Machine Learning. *International Journal of Image and Graphics*, 20(04), 2050033.
- Koenderink, J.J. (1984). The structure of images. *Biological Cybernetics*, 50(5), 363–370.
- Krizhevsky, A., Sutskever, I. and Hinton, G.E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. *Advances In Neural Information Processing Systems (NeurIPS)*, pp. 1097–1105.
- Kumar, K., Desrosiers, C., Siddiqi, K., Colliot, O. and Toews, M. (2017). Fiberprint: A subject fingerprint based on sparse code pooling for white matter fiber analysis. *NeuroImage*, 158, 242–259.
- Kumar, K., Chauvin, L., Toews, M., Colliot, O., Desrosiers, C. and others. (2018a). Multi-Modal Analysis of Genetically-Related Subjects Using SIFT Descriptors in Brain MRI. *Computational Diffusion MRI*, pp. 219–228.
- Kumar, K., Toews, M., Chauvin, L., Colliot, O., Desrosiers, C., Toews, M., Colliot, O., Desrosiers, C., Chauvin, L., Colliot, O. and Desrosiers, C. (2018b). Multi-modal brain fingerprinting: a manifold approximation based framework. *NeuroImage*, 183, 209726.
- Lai, M. (2015). Deep learning for medical image segmentation. *arXiv preprint arXiv:1505.02000*.
- Laptev, I. (2005). On Space-Time Interest Points. *International Journal of Computer Vision*, 64(2), 107–123.
- Lauterbur, P.C. (1973). Image formation by induced local interactions: examples employing nuclear magnetic resonance. *Nature*, 242(5394), 190–191.

- Le Bihan, D., Breton, E., Lallemand, D., Grenier, P., Cabanis, E. and Laval-Jeantet, M. (1986). MR imaging of intravoxel incoherent motions: application to diffusion and perfusion in neurologic disorders. *Radiology*, 161(2), 401–407.
- Le Bihan, D., Mangin, J.F., Poupon, C., Clark, C.A., Pappata, S., Molko, N. and Chabriat, H. (2001). Diffusion tensor imaging: Concepts and applications. *Journal of Magnetic Resonance Imaging*, 13(4), 534–546.
- LeCun, Y., Boser, B., Denker, J.S., Henderson, D., Howard, R.E., Hubbard, W. and Jackel, L.D. (1989). Backpropagation applied to handwritten zip code recognition. *Neural computation*, 1(4), 541–551.
- LeCun, Y., Bengio, Y. and Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436.
- Lehnert, R. (2022). Mirror symmetry validated for proton and its antimatter twin. *Nature*, 601(7891), 32–33.
- Lerch, J.P., van der Kouwe, A.J.W., Raznahan, A., Paus, T., Johansen-Berg, H., Miller, K.L., Smith, S.M., Fischl, B. and Sotiropoulos, S.N. (2017). Studying neuroanatomy using MRI. *Nature Neuroscience*, 20(3), 314–326.
- Levandowsky, M. and Winter, D. (1971). Distance between sets. *Nature*, 234(5323), 34.
- Lindeberg, T. (1994). Scale-space theory: A basic tool for analyzing structures at different scales. *Journal of applied statistics*, 21(1-2), 225–270.
- Lindeberg, T. (1998). Feature Detection with Automatic Scale Selection. *International Journal of Computer Vision (IJCV)*, 30(2), 79 – 116.
- Liu, Y., Hel-Or, H., Kaplan, C.S. and Gool, L.V. (2010). Computational Symmetry in Computer Vision and Computer Graphics. *Foundations and Trends in Computer Graphics and Vision*, 5(1–2), 1–195.
- Lowe, D.G. (1999). Object recognition from local scale-invariant features. *Proceedings of the 7th IEEE International Conference on Computer Vision*, 2, 1150–1157 vol.2.
- Lowe, D.G. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision (IJCV)*, 60(2), 91–110.
- Luo, C., Li, X., Wang, L., He, J., Li, D. and Zhou, J. (2018a). How Does the Data set Affect CNN-based Image Classification Performance? *2018 5th International Conference on Systems and Informatics (ICSAI)*, pp. 361–366.

- Luo, J., Frisken, S., Machado, I., Zhang, M., Pieper, S., Golland, P., Toews, M., Unadkat, P., Sedghi, A., Zhou, H. and others. (2018b). Using the variogram for vector outlier screening: application to feature-based image registration. *International journal of computer assisted radiology and surgery*, 13(12), 1871–1880.
- Luo, J., Toews, M., Machado, I., Frisken, S., Zhang, M., Preiswerk, F., Sedghi, A., Ding, H., Pieper, S., Golland, P. and others. (2018c). A feature-driven active framework for ultrasound-based brain shift compensation. *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 30–38.
- Luo, W., Li, Y., Urtasun, R. and Zemel, R. (2016). Understanding the Effective Receptive Field in Deep Convolutional Neural Networks. *Advances in Neural Information Processing Systems*, 29.
- Lv G. (2019). Self-similarity and symmetry with SIFT for multi-modal image registration. *IEEE Access*, 7, 52202–52213.
- Ma, W., Wen, Z., Wu, Y., Jiao, L., Gong, M., Zheng, Y. and Liu, L. (2017). Remote sensing image registration with modified sift and enhanced feature matching. *IEEE Geoscience and Remote Sensing Letters*, 14(1), 3–7.
- Machado, I., Toews, M., Luo, J., Unadkat, P., Essayed, W., George, E., Teodoro, P., Carvalho, H., Martins, J., Golland, P. and others. (2018a). Deformable mri-ultrasound registration via attribute matching and mutual-saliency weighting for image-guided neurosurgery. In *Simulation, Image Processing, and Ultrasound Systems for Assisted Diagnosis and Navigation* (pp. 165–171). Springer, Cham.
- Machado, I., Toews, M., Luo, J., Unadkat, P., Essayed, W., George, E., Teodoro, P., Carvalho, H., Martins, J., Golland, P., Pieper, S., Frisken, S., Golby, A., Wells, W., others et al. (2018b). Non-rigid registration of 3D ultrasound for neurosurgery using automatic feature detection and matching. *International Journal of Computer Assisted Radiology and Surgery (IJCARS)*, 13(10), 1525–1538.
- Marcus, D.S., Wang, T.H., Parker, J., Csernansky, J.G., Morris, J.C. and Buckner, R.L. (2007). Open Access Series of Imaging Studies (OASIS): cross-sectional MRI data in young, middle aged, nondemented, and demented older adults. *Journal of cognitive neuroscience*, 19(9), 1498–1507.
- Marczewski, E. and Steinhaus, H. (1958). On a certain distance of sets and the corresponding distance of functions. *Colloquium Mathematicum*, 6, 319–327.
- Marr, D. and Hildreth, E. (1980). Theory of edge detection. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, 207(1167), 187–217.

- Menze, B.H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., Lanczi, L., Gerstner, E., Weber, M.A., Arbel, T., Avants, B.B. et al. (2015). The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS). *IEEE transactions on medical imaging*, 34(10), 1993–2024.
- Mikolajczyk, K. and Schmid, C. (2005). A performance evaluation of local descriptors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(10), 1615–1630.
- Mikolajczyk, K. and Schmid, C. (2004). Scale & affine invariant interest point detectors. *International Journal of Computer Vision*, 60(1), 63–86.
- Miranda-Dominguez, O., Mills, B.D., Carpenter, S.D., Grant, K.A., Kroenke, C.D., Nigg, J.T. and Fair, D.A. (2014). Connectotyping: model based fingerprinting of the functional connectome. *PloS one*, 9(11), e111048.
- Mishchuk, A., Mishkin, D., Radenovic, F. and Matas, J. (2017). Working hard to know your neighbor’s margins: Local descriptor learning loss. *arXiv preprint arXiv:1705.10872*.
- Molodtsov, D. (1999). Soft set theory—First results. *Computers & Mathematics with Applications*, 37(4), 19–31.
- Moreels, P. and Perona, P. (2007). Evaluation of Features Detectors and Descriptors based on 3D Objects. *International Journal of Computer Vision*, 73(3), 263–284.
- Moyer, D., Abaci Turk, E., Grant, P.E., Wells, W.M. and Golland, P. (2021). Equivariant Filters for Efficient Tracking in 3D Imaging. *Medical Image Computing and Computer Assisted Intervention – MICCAI 2021*, (Lecture Notes in Computer Science), 193–202.
- Muja, M. and Lowe, D. (2014). Scalable nearest neighbor algorithms for high dimensional data. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 2227–2240.
- Muja, M. and Lowe, D.G. (2009). Fast approximate nearest neighbors with automatic algorithm configuration. *VISAPP*, 2(331-340), 2.
- Mukherjee, D., Jonathan Wu, Q.M. and Wang, G. (2015). A comparative experimental study of image feature detectors and descriptors. *Machine Vision and Applications*, 26(4), 443–466.
- Myronenko, A. and Song, X. (2010). Point Set Registration: Coherent Point Drift. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 32(12), 2262–2275.
- Newton, I. (1704). *Opticks, or, a treatise of the reflections, refractions, inflections & colours of light*. Courier Corporation.

- Nguyen, A., Yosinski, J. and Clune, J. (2015). Deep neural networks are easily fooled: High confidence predictions for unrecognizable images. *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 427–436.
- Ni, D., Qu, Y., Yang, X., Chui, Y.P., Wong, T.T., Ho, S.S.M. and Heng, P.A. (2008). Volumetric ultrasound panorama based on 3D SIFT. *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 52–60.
- Noether, E. (1971). Invariant variation problems. *Transport Theory and Statistical Physics*, 1(3), 186–207.
- Noh, H., Araujo, A., Sim, J., Weyand, T. and Han, B. (2017). Large-Scale Image Retrieval With Attentive Deep Local Features. *Proceedings of the IEEE international conference on computer vision*, pp. 3456–3465.
- Olver, P., Sapiro, G. and Tannenbaum, A. (1994). Differential Invariant Signatures and Flows in Computer Vision: A Symmetry Group Approach. In ter Haar Romeny, B.M. (Ed.), *Geometry-Driven Diffusion in Computer Vision* (pp. 255–306). Dordrecht: Springer Netherlands.
- Ono, M., Kubik, S. and Abernathey, C.D. (1990). *Atlas of the cerebral sulci*. Thieme.
- Ono, Y., Trulls, E., Fua, P., Moo Yi, K. and Yi, K.M. (2018). LF-Net: Learning Local Features from Images. *Advances in Neural Information Processing Systems (NeurIPS)*, 31.
- Paone, J.R., Flynn, P.J., Philips, P.J., Bowyer, K.W., Bruegge, R.W., Grother, P.J., Quinn, G.W., Pruitt, M.T. and Grant, J.M. (2014). Double trouble: Differentiating identical twins by face recognition. *IEEE Transactions on Information Forensics and Security (TIFS)*, 9(2).
- Park, J.H., Kim, O.H. and Kwun, Y.C. (2012). Some properties of equivalence soft set relations. *Computers & Mathematics with Applications*, 63(6), 1079–1088.
- Pascoli, S. and Turner, J. (2020). Matter–antimatter symmetry violated. *Nature*, 580(7803), 323–324.
- Penfield, W. and Boldrey, E. (1937). Somatic motor and sensory representation in the cerebral cortex of man as studied by electrical stimulation. *Brain*, 60(4), 389–443.
- Pepin, E., Carlier, J.B., Chauvin, L., Toews, M. and Harmouche, R. (2020). Large-Scale Unbiased Neuroimage Indexing via 3D GPU-SIFT Filtering and Keypoint Masking. In *Machine Learning in Clinical Neuroimaging and Radiogenomics in Neuro-oncology* (pp. 108–118). Springer.

- Pomerleau, F., Colas, F. and Siegwart, R. (2015). A Review of Point Cloud Registration Algorithms for Mobile Robotics. *Foundations and Trends in Robotics*, 4(1), 1–104.
- Regan, E.A., Hokanson, J.E., Murphy, J.R., Make, B., Lynch, D.A., Beaty, T.H., Curran-Everett, D., Silverman, E.K. and Crapo, J.D. (2011). Genetic Epidemiology of COPD (COPDGene) Study Design. *COPD: Journal of Chronic Obstructive Pulmonary Disease*, 7(1), 32–43.
- Riggi, F., Toews, M. and Arbel, T. (2006). Fundamental Matrix Estimation via TIP - Transfer of Invariant Parameters. *18th International Conference on Pattern Recognition (ICPR'06)*, 2, 21–24.
- Rister, B., Horowitz, M.A. and Rubin, D.L. (2017). Volumetric image registration from invariant keypoints. *IEEE Transactions on Image Processing*, 26(10), 4900–4910.
- Rogers, D.J. and Tanimoto, T.T. (1960). A Computer Program for Classifying Plants. *Science (New York, N.Y.)*, 132(3434), 1115–1118.
- Rokooie, M. (2020). *Visualizing large medical image datasets using the 3D scale-invariant feature transform (SIFT)*. (Master's thesis, École de technologie supérieure).
- Rosenblatt, F. (1958). The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review*, 65(6), 386.
- Rosenblatt, F. (1961). *Principles of neurodynamics. perceptrons and the theory of brain mechanisms*.
- Rublee, E., Rabaud, V., Konolige, K. and Bradski, G. (2011). ORB: An efficient alternative to SIFT or SURF. *IEEE International Conference on Computer Vision (ICCV)*, pp. 2564–2571.
- Sabuncu, M.R., Ge, T., Holmes, A.J., Smoller, J.W., Buckner, R.L., Fischl, B. and Alzheimer's Disease Neuroimaging Initiative (ADNI). (2016). Morphometricity as a measure of the neuroanatomical signature of a trait. *Proceedings of the National Academy of Sciences (PNAS)*, 113(39), E5749–E5756.
- Saiti, E. and Theoharis, T. (2020). An application independent review of multimodal 3D registration methods. *Computers & Graphics*, 91, 153–178.
- Samper-González, J., Burgos, N., Bottani, S., Fontanella, S., Lu, P., Marcoux, A., Routier, A., Guillon, J., Bacci, M., Wen, J., Bertrand, A., Bertin, H., Habert, M.O., Durrleman, S., Evgeniou, T. et al. (2018). Reproducible evaluation of classification methods in Alzheimer's disease: Framework and application to MRI and PET data. *NeuroImage*, 183, 504–521.

- Schrödinger, E. (1926). An Undulatory Theory of the Mechanics of Atoms and Molecules. *Physical Review*, 28(6), 1049–1070.
- Schönberger, J.L., Hardmeier, H., Sattler, T., Pollefeys, M., Schonberger, J.L., Hardmeier, H., Sattler, T. and Pollefeys, M. (2017). Comparative Evaluation of Hand-Crafted and Learned Local Features. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1482–1491.
- Scovanner, P., Ali, S., International, M.S.P.o.t.t.A., 2007, U. and Shah, M. (2007). A 3-dimensional sift descriptor and its application to action recognition. *Proceedings of the ACM international conference on Multimedia*, pp. 357–360.
- Serna, I., Peña, A., Morales, A. and Fierrez, J. (2021). InsideBias: Measuring Bias in Deep Networks and Application to Face Gender Biometrics. *2020 25th International Conference on Pattern Recognition (ICPR)*, pp. 3720–3727.
- Shahamat, H. and Pouyan, A.A. (2015). Feature selection using genetic algorithm for classification of schizophrenia using fMRI data. *Journal of AI and Data Mining*, 3(1), 30–37.
- Sharp, G.C., Lee, S.W. and Wehe, D.K. (2002). ICP registration using invariant features. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(1), 90–102.
- Shi, J. and Tomasi. (1994). Good features to track. *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 593–600.
- Smith, S.M., Jenkinson, M., Woolrich, M.W., Beckmann, C.F., Behrens, T.E., Johansen-Berg, H., Bannister, P.R., De Luca, M., Drobnjak, I., Flitney, D.E. and others. (2004). Advances in functional and structural MR image analysis and implementation as FSL. *Neuroimage*, 23, S208–S219.
- Song, Y., Cai, W., Huang, H., Zhou, Y., Feng, D.D., Wang, Y., Fulham, M.J. and Chen, M. (2015). Large Margin Local Estimate With Applications to Medical Image Classification. *IEEE Transactions on Medical Imaging*, 34(6), 1362–1377.
- Sozzi, M. (2008). *Discrete Symmetries and CP Violation: From Experiment to Theory*. Oxford University Press.
- Spezialetti, R., Salti, S. and Stefano, L.D. (2019). Learning an Effective Equivariant 3D Descriptor Without Supervision. *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 6401–6410.
- Stejskal, E.O. and Tanner, J.E. (1965). Spin diffusion measurements: spin echoes in the presence of a time-dependent field gradient. *The journal of chemical physics*, 42(1), 288–292.

- Su, J., Vargas, D.V. and Sakurai, K. (2019). One pixel attack for fooling deep neural networks. *IEEE Transactions on Evolutionary Computation*, 23(5), 828–841.
- Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., Downey, P., Elliott, P., Green, J., Landray, M., Liu, B., Matthews, P., Ong, G., Pell, J., Silman, A. et al. (2015). UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age. *PLOS Medicine*, 12(3), e1001779.
- Takao, H., Hayashi, N. and Ohtomo, K. (2015). Brain morphology is individual-specific information. *Magnetic resonance imaging*, 33(6), 816–21.
- Tanaka, H. and Kunin, D. (2021). Noether’s Learning Dynamics: Role of Symmetry Breaking in Neural Networks. *Advances in Neural Information Processing Systems*, 34, 25646–25660.
- Teng, S.W., Hossain, M.T. and Lu, G. (2015). Multimodal image registration technique based on improved local feature descriptors. *Journal of Electronic Imaging*, 24(1), 013013.
- Terrell, G.R. and Scott, D.W. (1992). Variable Kernel Density Estimation. *The Annals of Statistics*, 1236–1265.
- Thompson, P.M., Cannon, T.D., Narr, K.L., Van Erp, T., Poutanen, V.P., Huttunen, M., Lönqvist, J., Standertskjöld-Nordenstam, C.G., Kaprio, J., Khaledy, M., Dail, R., Zoumalan, C.I., Toga, A.W. and others. (2001). Genetic influences on brain structure. *Nature Neuroscience*, 4(12), 1253.
- Toews, M. and Arbel, T. (2007). A statistical parts-based model of anatomical variability. *IEEE Transactions on Medical Imaging*, 26(4), 497–508.
- Toews, M. and Wells, W. (2009). SIFT-Rank: Ordinal description for invariant feature correspondence. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 172–177.
- Toews, M. and Wells, W.M. (2010). A mutual-information scale-space for image feature detection and feature-based classification of volumetric brain images. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 111–116.
- Toews, M. and Wells III, W.M. (2016). How are siblings similar? How similar are siblings? Large-scale imaging genetics using local image features. *International Symposium on Biomedical Imaging (ISBI)*, pp. 847–850.
- Toews, M., Wells III, W., Collins, D.L. and Arbel, T. (2010). Feature-based morphometry: Discovering group-related anatomical patterns. *NeuroImage*, 49(3), 2318–2327.

- Toews, M., Wells, W.M. and Zöllei, L. (2012). Feature-based developmental model of the infant brain in structural MRI. *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 7511 LNCS, 204–211.
- Toews, M., Wells III, W.M. and Wells, W. (2013a). Efficient and Robust Model-to-Image Alignment using 3D Scale-Invariant Features. *Medical Image Analysis (MIA)*, 17(3), 271–282.
- Toews, M., Zöllei, L. and Wells, W.M. (2013b). Feature-based Alignment of Volumetric Multimodal Images. *International Conference on Information Processing in Medical Imaging (IPMI)*, 23, 25–36.
- Toews, M., Wachinger, C., Estepar, R.S.J. and Wells, W. (2015). A Feature-Based Approach to Big Data Analysis of Medical Images. *International Conference on Information Processing in Medical Imaging (IPMI)*, 24, 339–350.
- Torrallba, A. and Efros, A.A. (2011). Unbiased look at dataset bias. *CVPR 2011*, pp. 1521–1528.
- Toussaint, G.T. (2002). Proximity graphs for nearest neighbor decision rules: recent progress. *Interface*, 34.
- Tyszkiewicz, M., Fua, P. and Trulls, E. (2020). DISK: Learning local features with policy gradient. *Advances in Neural Information Processing Systems*, 33, 14254–14265.
- Urschler, M., Bauer, J., Ditt, H. and Bischof, H. (2006). SIFT and shape context for feature-based nonlinear registration of thoracic CT images. *International Workshop on Computer Vision Approaches to Medical Image Analysis*, pp. 73–84.
- Valizadeh, S.A., Liem, F., Méritat, S., Hänggi, J. and Jäncke, L. (2018). Identification of individual subjects on the basis of their brain anatomical features. *Scientific reports*, 8(1), 5611.
- van der Lee, S.J., Roshchupkin, G.V., Adams, H.H., Schmidt, H., Hofer, E., Saba, Y., Schmidt, R., Hofman, A., Amin, N., van Duijn, C.M., Vernooij, M.W., Ikram, M.A. and Niessen, W.J. (2017). Gray matter heritability in family-based and population-based studies using voxel-based morphometry. *Human Brain Mapping (HBM)*, 38(5).
- Van Essen, D.C., Smith, S.M., Barch, D.M., Behrens, T.E.J., Yacoub, E., Ugurbil, K., Consortium, W.M.H.C.P. and others. (2013). The WU-Minn human connectome project: an overview. *NeuroImage*, 80, 62–79.
- Van Essen, D.C., Donahue, C., Dierker, D.L. and Glasser, M.F. (2016). *Parcellations and Connectivity Patterns in Human and Macaque Cerebral Cortex*. Springer.

- Van Essen, D., Ugurbil, K., Auerbach, E., Barch, D., Behrens, T., Bucholz, R., Chang, A., Chen, L., Corbetta, M., Curtiss, S., Della Penna, S., Feinberg, D., Glasser, M., Harel, N., Heath, A. et al. (2012). The Human Connectome Project: A data acquisition perspective. *NeuroImage*, 62(4), 2222–2231.
- Viola, P. and Wells, W.M. (1997). Alignment by Maximization of Mutual Information. *International Journal of Computer Vision*, 24(2), 137–154.
- Wachinger, C. and Navab, N. (2012). Entropy and Laplacian images: Structural representations for multi-modal registration. *Medical Image Analysis*, 16(1), 1–17.
- Wachinger, C., Golland, P., Kremen, W., Fischl, B., Reuter, M., Initiative, A.D.N. and others. (2015a). BrainPrint: A discriminative characterization of brain morphology. *NeuroImage*, 109, 232–248.
- Wachinger, C., Toews, M., Langs, G., Wells, W. and Golland, P. (2015b). Keypoint transfer segmentation. *International Conference on Information Processing in Medical Imaging*, pp. 233–245.
- Wachinger, C., Toews, M., Langs, G., Wells, W. and Golland, P. (2018). Keypoint transfer for fast whole-body segmentation. *IEEE Transactions on Medical Imaging (TMI)*, 39(2), 273–282.
- Wachinger, C., Becker, B.G., Rieckmann, A. and Pölsterl, S. (2019). Quantifying Confounding Bias in Neuroimaging Datasets with Causal Inference. *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019*, (Lecture Notes in Computer Science), 484–492.
- Witherspoon, D.J., Wooding, S., Rogers, A.R., Marchani, E.E., Watkins, W.S., Batzer, M.A. and Jorde, L.B. (2007). Genetic Similarities Within and Between Human Populations. *Genetics*, 176(1), 351–359.
- Witkin, A. (1984). Scale-space filtering: A new approach to multi-scale description. *IEEE International Conference on Acoustics, Speech, and Signal Processing*, 9, 150–153.
- Xia, R., Zhao, J. and Liu, Y. (2013). A robust feature-based registration method of multimodal image using phase congruency and coherent point drift. *MIPPR 2013: Pattern recognition and computer vision*, 8919, 891903.
- Yeh, F.C., Vettel, J.M., Singh, A., Poczos, B., Grafton, S.T., Erickson, K.I., Tseng, W.Y.I. and Verstynen, T.D. (2016). Quantifying differences and similarities in whole-brain white matter architecture using local connectome fingerprints. *PLoS computational biology*, 12(11), e1005203.

- Yi, K.M., Trulls, E., Lepetit, V. and Fua, P. (2016). LIFT: Learned invariant feature transform. *Proceedings of the European Conference on Computer Vision (ECCV)*, 9910 LNCS, 467–483.
- Yuan, Y., Chao, M. and Lo, Y.C. (2017). Automatic Skin Lesion Segmentation Using Deep Fully Convolutional Networks with Jaccard Distance. *IEEE Transactions on Medical Imaging (TMI)*, 36(9).
- Zheng, L., Yang, Y. and Tian, Q. (2017). SIFT Meets CNN: A Decade Survey of Instance Retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(5), 1–1.