



Πανεπιστήμιο Δυτικής Αττικής
Σχολή Μηχανικών
Τμήμα Ηλεκτρολόγων & Ηλεκτρονικών μηχανικών

Διπλωματική Εργασία

**Τεχνικές εύρωστης εκμάθησης μετρικών σε χώρους πολλαπλότητας
πινάκων συνδιακύμανσης για χρήση σε εφαρμογές ταξινόμησης προτύπων
και μηχανικής μάθησης**

Φοιτητής: Αντώνιος Καλαϊτζάκης
ΑΜ: 262017039

Επιβλέποντες Καθηγητές

Δρ. Ηλίας Ζώης,
Αναπληρωτής Καθηγητής

Δρ. Αλέξανδρος Αλεξανδρίδης,
Καθηγητής

Αθήνα-Αιγάλεω, Μάρτιος 2024



University of West Attica
Faculty of Engineering
Department of Electrical & Electronics Engineering

Diploma Thesis

**Robust Distance Measure for Similarity-Based Classification on the SPD
Manifold**

Student: Antonios Kalaitzakis
Registration Number: 262017039

Supervisors

Dr. Elias Zois,
Associate Professor

Dr. Alexandros Alexandridis,
Professor

Athens-Egaleo, March 2024

Η Διπλωματική Εργασία έγινε αποδεκτή και βαθμολογήθηκε από την εξής τριμελή επιτροπή:

Ζώης Ηλίας, Αναπληρωτής Καθηγητής	Αλεξανδρίδης Αλέξανδρος, Καθηγητής	Χωριανόπουλος Χρήστος, Επίκουρος Καθηγητής

Copyright © Με επιφύλαξη παντός δικαιώματος. All rights reserved.

**ΠΑΝΕΠΙΣΤΗΜΙΟ ΔΥΤΙΚΗΣ ΑΤΤΙΚΗΣ και Καλαϊτζάκης Αντώνιος,
Μάρτιος 2024**

Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα. Ερωτήματα που αφορούν τη χρήση της εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς τους συγγραφείς.

Οι απόψεις και τα συμπεράσματα που περιέχονται σε αυτό το έγγραφο εκφράζουν τον/τη συγγραφέα του και δεν πρέπει να ερμηνευθεί ότι αντιπροσωπεύουν τις θέσεις του επιβλέποντος, της επιτροπής εξέτασης ή τις επίσημες θέσεις του Τμήματος και του Ιδρύματος.

ΔΗΛΩΣΗ ΣΥΓΓΡΑΦΕΑ ΔΙΠΛΩΜΑΤΙΚΗΣ ΕΡΓΑΣΙΑΣ

Ο/η κάτωθι υπογεγραμμένος ΚΑΛΑΪΤΖΑΚΗΣ ΑΝΤΩΝΙΟΣ του ΘΕΟΦΙΛΟΥ, με αριθμό μητρώου 262017039 φοιτητής του Πανεπιστημίου Δυτικής Αττικής της Σχολής ΜΗΧΑΝΙΚΩΝ του Τμήματος ΗΛΕΚΤΡΟΛΟΓΩΝ ΚΑΙ ΗΛΕΚΤΡΟΝΙΚΩΝ ΜΗΧΑΝΙΚΩΝ,

δηλώνω υπεύθυνα ότι:

«Είμαι συγγραφέας αυτής της διπλωματικής εργασίας και ότι κάθε βοήθεια την οποία είχα για την προετοιμασία της είναι πλήρως αναγνωρισμένη και αναφέρεται στην εργασία. Επίσης, οι όποιες πηγές από τις οποίες έκανα χρήση δεδομένων, ιδεών ή λέξεων, είτε ακριβώς είτε παραφρασμένες, αναφέρονται στο σύνολό τους, με πλήρη αναφορά στους συγγραφείς, τον εκδοτικό οίκο ή το περιοδικό, συμπεριλαμβανομένων και των πηγών που ενδεχομένως χρησιμοποιήθηκαν από το διαδίκτυο. Επίσης, βεβαιώνω ότι αυτή η εργασία έχει συγγραφεί από μένα αποκλειστικά και αποτελεί προϊόν πνευματικής ιδιοκτησίας τόσο δικής μου, όσο και του Ιδρύματος.

Παράβαση της ανωτέρω ακαδημαϊκής μου ευθύνης αποτελεί ουσιώδη λόγο για την ανάκληση του διπλώματός μου.

Επιθυμώ την απαγόρευση πρόσβασης στο πλήρες κείμενο της εργασίας μου μέχρι _____ και έπειτα από αίτησή μου στη Βιβλιοθήκη και έγκριση του επιβλέποντος/ουσας καθηγητή/ήτριας.»

Ο/Η Δηλών/ούσα
(Ονοματεπώνυμο φοιτητή/ήτριας)
Καλαϊτζάκης Αντώνιος

(Υπογραφή φοιτητή/ήτριας)

Με την ολοκλήρωση της διπλωματικής μου εργασίας φτάνω στο τέλος των προπτυχιακών σπουδών, και γι' αυτό, νιώθω την ανάγκη να εκφράσω τις ευχαριστίες μου στους ανθρώπους μου με βοήθησαν να φτάσω ως εδώ. Θα ήθελα αρχικά να ευχαριστήσω θερμά τους επιβλέποντες καθηγητές μου, τον κ. Ζώη Ηλία και τον κ. Αλεξανδρίδη Αλέξανδρο, για την καθοδήγηση και την πολύτιμη βοήθειά τους στην ολοκλήρωση της διπλωματικής μου εργασίας.

Επιπλέον, θα ήθελα να ευχαριστήσω τους εκπαιδευτικούς από όλα τα χρόνια της ζωής μου, για τη μόρφωση που μου παρείχαν. Ευχαριστώ ιδιαίτερα τους καθηγητές μου κ. Ιωάννη Κουβαράκη, κ. Χαράλαμπο Δερμάτη, κ. Ιωάννη Μαλτέζο, του τμήματος Ηλεκτρονικής του 1ου ΕΠΑΛ Περάματος, που με ενέπνευσαν και μου έδειξαν πρώτοι τις δυνατότητες των ηλεκτρονικών.

Θα ήθελα επίσης να ευχαριστήσω την οικογένειά μου για τη στήριξη και την ενθάρρυνση που μου παρείχαν σε κάθε βήμα μου. Τέλος, . . .



Περίληψη

Παρά την εξέλιξη της τεχνολογίας, οι χειρόγραφες υπογραφές αποτελούν εδώ και αιώνες ένα από τα πιο διαδεδομένα μέσα βιομετρικής ταυτοποίησης. Τις τελευταίες δεκαετίες, η ανθρώπινη υπογραφή έχει ενταχθεί στο ψηφιακό κόσμο και ερευνώνται τρόποι για την ανάπτυξη συστημάτων, που εκτελούν αυτόματα τη διαδικασία της αναγνώρισης ατόμων μέσω αυτής. Στην παρούσα διπλωματική εργασία, παρουσιάζεται η υλοποίηση ενός συστήματος ταξινόμησης χειρόγραφων υπογραφών, με τη χρήση ενός πρωτοπόρου αλγόριθμου Αναγνώρισης Προτύπων. Κατά την εφαρμογή αυτή, κατασκευάζεται ένα σύστημα βιομετρικής αναγνώρισης χρηστών, που δέχεται τις χειρόγραφες υπογραφές τους στη μορφή ψηφιακών εικόνων. Η προσέγγιση του προβλήματος της Αναγνώρισης Προτύπων που ακολουθείται από τον παρόντα αλγόριθμο, αναγνωρίζει διαφορετικούς υπογράφοντες χωρίς να κατασκευάζει διαφορετικά συστήματα αναγνώρισης για τον κάθε ένα. Μεταξύ άλλων προτερημάτων, το προτεινόμενο σύστημα προσφέρει ένα συνδυασμό ασφάλειας και διευκόλυνσης των εξουσιοδοτημένων, λόγω των αυξημένων δυνατοτήτων του υπό ανάλυση αλγόριθμου.

Η απεικόνιση της ψηφιακής χειρόγραφης υπογραφής, από τον χώρο της ψηφιακής εικόνας σε μια μαθηματική οντότητα, γίνεται μέσω της δημιουργίας ενός πίνακα συνδιακύμανσης. Αυτό έχει ως αποτέλεσμα, κάθε εικόνα να αντιστοιχεί σε ένα σημείο, το οποίο βρίσκεται πάνω σε μια μαθηματική επιφάνεια με ιδιαίτερες ιδιότητες. Η κυριότερη από αυτές, είναι ότι τα σημεία που βρίσκονται σε αυτή την επιφάνεια, δεν υπόκεινται στους γνωστούς “Ευκλείδειους” κανόνες. Αντιθέτως, τα σημεία αυτά, αναπαρίστανται από ένα Συμμετρικό και Θετικά Ορισμένο Πίνακα (SPD Matrix) και ανήκουν σε ένα Riemannian χώρο Πολλαπλοτήτων (Manifolds), με μέλη το σύνολο των SPD πινάκων. Λόγω της δομής του χώρου αυτού, οι μαθηματικές πράξεις επαναδιατυπώνονται και υπόκεινται σε συγκεκριμένους κανόνες, που παραπέμπουν στις αρχές της διαφορικής γεωμετρίας. Ο κύριος λόγος για τη χρήση τέτοιων Πολλαπλοτήτων, έχει να κάνει με την εκμετάλλευση της μη γραμμικής συσχέτισης, που προέρχεται από τον υπολογισμό του πίνακα συνδιακύμανσης μιας εικόνας. Καθώς ο υπό ανάλυση αλγόριθμος βασίζεται στη μέτρηση των αποστάσεων μεταξύ των αναπαραστάσεων των δεδομένων, η χρήση της σχέσης εκείνης που περιγράφει σωστά την έννοια της απόστασης, αποτελεί ένα βασικό στοιχείο για τη διαδικασία της ταξινόμησης. Παρόλο που έχουν αναπτυχθεί μετρικές που ορίζουν την έννοια της απόστασης σε αυτούς τους χώρους, ως μέρος της λειτουργίας του αλγόριθμου, επιλέγεται να γίνει Εκμάθηση Μετρικής, κατά την οποία, ανακαλύπτεται η σχέση που διατυπώνει με ακρίβεια, την απόσταση μεταξύ των στοιχείων του χώρου.

Περαιτέρω, μια από τις καινοτομίες του προτεινόμενου αλγόριθμου, αποτελεί το γεγονός ότι, από την κάθε αναπαράσταση εξάγονται πολλαπλές άλλες, μικρότερων διαστάσεων, στις οποίες αναδεικνύονται διαφορετικά χαρακτηριστικά των δεδομένων. Η αντιμετώπιση αυτή όπως παρουσιάζεται, επιφέρει μεγαλύτερη ακρίβεια κατά την αναγνώριση υπογραφών των χρηστών. Επιπλέον, λόγω του τρόπου λειτουργίας του, το σύστημα αυτό παρέχει τη δυνατότητα της Μεταφοράς Μάθησης, δηλαδή της χρήσης του και σε άλλες εφαρμογές ταξινόμησης. Τέλος, για την τεκμηρίωση των παραπάνω υποστηριζόμενων δυνατοτήτων του αλγόριθμου, το σύστημα που κατασκευάζεται αξιολογείται με τη χρήση δύο διαφορετικών Βάσεων Δεδομένων που περιέχουν υπογραφές.

Λέξεις - κλειδιά

Βιομετρική Αναγνώριση, Χειρόγραφη Υπογραφή, Αναγνώριση Προτύπων, Υπολογιστική Όραση, Συμμετρική και Θετικά Ορισμένη Πολλαπλότητα, Εκμάθηση Μετρικής, Μεταφορά Μάθησης

Abstract

Despite the advancement of technology, handwritten signatures have been one of the most widespread means of biometric authentication, throughout the ages. Over the last decade, handwritten signatures have been successfully integrated into today's digital world and ways are being explored, in which the authentication process of one's identity is automatically verified, via their signature. In the present thesis, the development of such system is being showcased, using a novel Pattern Recognition algorithm. Specifically, a biometric authentication system is being built, which classifies handwritten signatures, presented in the form of digital images. This system is Writer Independent, which means that it has the ability to recognize different signers without the need of developing different classifiers for each one of them. Among its qualities, this system offers a combination of security and user convenience, because of the enhanced capabilities of the underlying algorithm.

The mapping of the digital handwritten signature, from the digital image space, to a mathematical entity, is done through the use of a covariance matrix. As a result, each digital image corresponds to a point on the surface of a mathematical entity, with particular properties. The main one is the fact that, the points on this entity, are not subject to the well known "Euclidean" rules. On the contrary, these points are represented by Symmetric Positive Definite (SPD) Matrices, and are on the Riemannian space of the SPD Manifold. Because of its structure, mathematical operations are reformulated and subject to specific rules which refer to the principles of differential geometry. The main reason behind utilizing such Manifolds, has to do with the exploitation of the non-linear correlation, which stems from the calculation of the covariance matrix of an image. Since the proposed algorithm utilizes the measurement of distances between representations of the data, the use of that mathematical expression which correctly describes the concept of distance, is a key element for the classification process. Although metrics that define the notion of distance on the SPD Manifold have already been developed, Metric Learning is being utilized, in order for the algorithm to discover the expression that accurately formulates the distance between points of this space.

Furthermore, one of the innovative characteristics of the proposed algorithm, is the fact that from each representation multiple other of smaller dimensions are extracted, in which different characteristics of the data are emphasized. In this way, as it will be later showcased, user signatures are categorized with higher accuracy. Moreover, due to its way of operation, this system provides the ability of Transfer Learning, i.e., its use in other classification applications. Lastly, to document the above supported capabilities of the algorithm, the system being built is evaluated using two different databases, containing handwritten signatures.

Keywords

Biometric Recognition, Handwritten Signature, Pattern Recognition, Computer Vision, Symmetric Positive Definite Manifold, Metric Learning, Transfer Learning

Περιεχόμενα

1	Αναδρομή στην Υπολογιστική Όραση	1
1.1	Ιστορική Αναδρομή	4
1.1.1	Δεκαετία 1970: Η γέννηση της Υπολογιστικής Όρασης	4
1.1.2	Δεκαετία 1980: Νέες τεχνικές και μαθηματική μοντελοποίηση	5
1.1.3	Δεκαετία 1990: 3D Γραφικά	5
1.1.4	Δεκαετία 2000: Υπολογιστική Φωτογραφία	6
1.1.5	Δεκαετία 2010: Σύγχρονη Υπολογιστική Όραση	7
1.2	Εφαρμογές	8
1.3	Αλγόριθμοι Μηχανικής Μάθησης και Χώρος Χαρακτηριστικών	11
1.3.1	Βασικά στοιχεία της Οπισθοδιάδοσης	12
2	Riemannian Γεωμετρία	14
2.1	Τοπολογικοί Χώροι	14
2.1.1	Χώροι Hausdorff	15
2.1.2	Βάση Τοπολογικού Χώρου	15
2.2	Τοπολογικές Πολλαπλότητες	15
2.2.1	Χάρτες Συντεταγμένων	16
2.2.2	Συναρτήσεις Μετάβασης	16
2.2.3	Άτλαντες	18
2.2.4	Σύνολο Πηλίκου	18
2.2.5	Λεία Δομή	18
2.3	Λείες Πολλαπλότητες	19
2.3.1	Εφαπτόμενος Χώρος	19
2.3.2	Γεωδειακή	20
2.4	Εφαπτόμενη Δέσμη	20
2.5	Εκθετική Απεικόνιση	21
2.6	Riemannian Πολλαπλότητες	21
2.6.1	Πολλαπλότητες των Συμμετρικών και Θετικά Ορισμένων Πινάκων	22
2.6.2	Stiefel Πολλαπλότητες	22
2.6.3	Grassmannian Πολλαπλότητες	22
2.7	Χρήσιμες Μαθηματικές Πράξεις Πινάκων	23
2.7.1	Λογαριθμική Συνάρτηση	23
2.7.2	Εκθετική Συνάρτηση	24
2.7.3	Συνδιακύμανση	24
2.8	Μετρικές Πολλαπλοτήτων των SPD Πινάκων	25
2.8.1	Affine Invariant Riemannian Metric - AIRM	25
2.8.2	Log-Euclidean Metric - LEM	25
2.8.3	Alpha-Beta Divergence	26
3	Αλγόριθμος Point-to-Set and Set-to-Set Distance	28
3.1	Σύγκριση του PSSSD με άλλες μεθόδους	28
3.1.1	Μέθοδοι Ταξινόμησης	28
3.1.2	Μέθοδοι Εκμάθησης Μετρικής	29
3.2	Παρουσίαση της μεθόδου που ακολουθείται	30
3.2.1	Μετασχηματισμός Σημείου σε Σύνολο	30
3.2.2	Set-to-Set Μετρική	32

3.3	Μέθοδος Βελτιστοποίησης	36
3.3.1	Παράγωγος της Συνάρτησης Κόστους ως προς τη παράμετρο M	37
3.3.2	Παράγωγος της Συνάρτησης Κόστους ως προς τη παράμετρο A	37
3.3.3	Παράγωγος της Συνάρτησης Κόστους ως προς τη παράμετρο W	38
3.3.4	Ανανέωση των παραμέτρων	39
3.4	Υπολογιστική Πολυπλοκότητα	40
3.4.1	Υπολογιστική Πολυπλοκότητα της Συνάρτησης Κόστους	40
3.4.2	Υπολογιστική Πολυπλοκότητα των μερικών παραγώγων	40
3.4.3	Υπολογιστική Πολυπλοκότητα της ανανέωσης των Riemannian παραμέτρων	40
4	Γενικά περί Βιομετρικής Ταυτοποίησης	41
4.1	Κλασικές μέθοδοι Ταυτοποίησης της Ανθρώπινης Παρουσίας	41
4.2	Βιομετρικές μέθοδοι Ταυτοποίησης	42
4.2.1	Στατικά Βιομετρικά Χαρακτηριστικά	42
4.2.2	Δυναμικά Βιομετρικά Χαρακτηριστικά	43
4.3	Προκλήσεις των Συστημάτων Βιομετρικής Αναγνώρισης	45
4.3.1	Ασφάλεια	45
4.3.2	Πλαστογράφιση	46
4.4	Αυτόματη Επιβεβαίωση της Ανθρώπινης Ταυτότητας μέσω Υπογραφής	46
4.5	Μέθοδοι Ανάλυσης Υπογραφών	48
4.5.1	Δυναμική Ανάλυση Υπογραφών	48
4.5.2	Στατική Ανάλυση Υπογραφών	48
4.6	Μέθοδοι Ταξινόμησης Υπογραφών	49
4.6.1	Writer Dependent Μέθοδος	49
4.6.2	Writer Independent Μέθοδος	50
5	Πειράματα	52
5.1	Βάσεις Δεδομένων	52
5.2	Προεπεξεργασία Εικόνων	52
5.3	Εξαγωγή Χαρακτηριστικών	54
5.4	Εξαγωγή Πινάκων Συνδιακύμανσης	55
5.5	Πειραματική Διαδικασία	56
5.5.1	Στάδιο Εκπαίδευσης και Αξιολόγησης	57
5.5.2	Στάδιο Ελέγχου	61
5.6	Αποτελέσματα	64
5.7	Σχολιασμός Αποτελεσμάτων	65
6	Συμπεράσματα	68

1 Αναδρομή στην Υπολογιστική Όραση

Προτού παρουσιαστεί το πρόβλημα που στοχεύει να επιλύσει ο προτεινόμενος αλγόριθμος, ο τρόπος λειτουργίας του και η χρήση του στην παρούσα διπλωματική εργασία, θεωρείται πρέπει να γίνει αναφορά σε θεμελιώδεις έννοιες, στις οποίες βασίζεται η φιλοσοφία πίσω από την προσέγγιση της επίλυσης του προβλήματος. Τα θεμελιώδη αυτά συστατικά αποτελούν: η Αναγνώριση Προτύπων (Pattern Recognition), η Μηχανική Μάθηση (Machine Learning) και η Υπολογιστική Όραση (Computer Vision).

Η **Υπολογιστική Νοημοσύνη** (Artificial Intelligence) [1], είναι ο κλάδος της Επιστήμης των Υπολογιστών (Computer Science), ο οποίος ασχολείται με τη σχεδίαση και την υλοποίηση ευφυών δυνατοτήτων όμοιες με αυτές των ανθρώπων, σε υπολογιστικά συστήματα. Στόχος του κλάδου είναι η δημιουργία ευφυών συστημάτων, τα οποία έχουν την ικανότητα να μαθαίνουν, να προσαρμόζονται και να κατανοούν μέσα από δεδομένα, με σκοπό να αναπτύσσουν λογικά συμπεράσματα και να εκτελούν ενέργειες που θα οδηγήσουν στην επίλυση ενός προβλήματος. Τα συστήματα αυτά, συνήθως υλοποιούνται ως αλγόριθμοι.

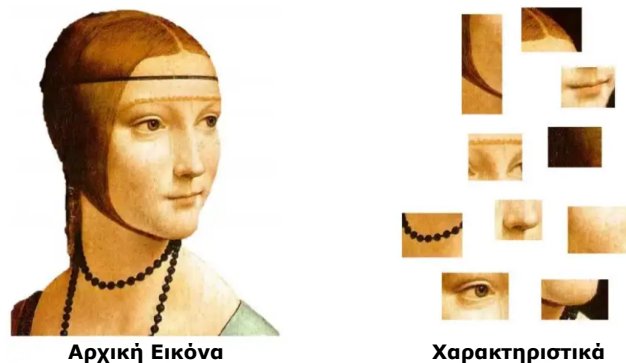
Η **Αναγνώριση Προτύπων** (Pattern Recognition) [2, 3], ασχολείται με την εύρεση προτύπων (patterns) σε δεδομένα, τα οποία τα καθιστούν όσο πιο ξεχωριστά γίνεται, ώστε να είναι ευκολότερη η διάκριση τους μεταξύ άλλων, παρόμοιων δεδομένων.

Τα **Δεδομένα** (Data) [4, 5], είναι ένα σύνολο διακριτών στοιχείων, τα οποία παρέχουν πληροφορία για ένα αντικείμενο. Τα αρχικά, ακατέργαστα δεδομένα (raw data), αποτελούν μετρήσεις ή παρατηρήσεις που προέρχονται από μια πηγή και αφορούν την εφαρμογή που χρήζει επίλυσης. Η πηγή αυτή μπορεί να είναι ένα ελεγχόμενο περιβάλλον, όπως ένα εργαστήριο, όπου πολλές συνθήκες είναι ελεγχόμενες και ορισμένες ή ένα μη-ελεγχόμενο περιβάλλον, όπως ο έξω κόσμος, όπου τα δεδομένα που θα συλλεχθούν, επηρεάζονται από πολλούς εξωτερικούς παράγοντες. Ένα σύνολο δεδομένων, αποτελεί το αντιπροσωπευτικό δείγμα του υπό εξέταση προβλήματος και ονομάζεται ως **Βάση Δεδομένων** (Dataset).

Τα στοιχεία αυτά μπορεί να είναι αριθμοί, λέξεις, δυαδικές καταστάσεις αλλά και πιο σύνθετες ποσότητες πληροφορίας όπως σήματα, εικόνες και βίντεο. Σε κάποιες περιπτώσεις, τα δεδομένα αυτά συνοδεύονται από επιπλέον πληροφορία, η οποία τα χαρακτηρίζει ως προς μια ή παραπάνω [6, 7] ιδιότητές τους. Η πληροφορία αυτή ονομάζεται **ετικέτα των δεδομένων** (label of the data). Τα δεδομένα τα οποία συνοδεύονται από ετικέτα, χαρακτηρίζονται ως **“δεδομένα με ετικέτα”** (labeled data), διαφορετικά, χαρακτηρίζονται ως **“δεδομένα χωρίς ετικέτα”** (unlabeled data). Οι ετικέτες των δεδομένων, ανάλογα με το είδος τους αλλά και με τον τρόπο με τον οποίο επιλέγεται να χρησιμοποιηθούν, χαρακτηρίζονται είτε ως αριθμητικές (numerical labels), είτε ως κατηγορηματικές (categorical labels) [8]. Οι αριθμητικές ετικέτες είναι αυτές οι οποίες χρησιμοποιούνται ως αριθμοί, με σκοπό για παράδειγμα να εφαρμοστούν σε αυτές μαθηματικές πράξεις. Οι κατηγορηματικές, μπορεί να είναι λέξεις αλλά και αριθμοί, οι οποίοι επιλέγεται να χρησιμοποιηθούν ως κατηγορίες. Για παράδειγμα, μπορεί μια εικόνα που απεικονίζει ένα σκύλο, να συνοδεύεται από τις κατηγορηματικές ετικέτες: “σκύλος” και “Labrador Retriever”, ενώ μπορεί να παραλείπονται άλλα χαρακτηριστικά του, όπως το χρώμα του. Άλλες φορές, μπορεί οι ετικέτες των δεδομένων να είναι αριθμητικές και να χαρακτηρίζουν τα δεδομένα ως προς τις κλάσεις τους, για παράδειγμα “-1” και “1” όπως συνηθίζεται για σύνολα δεδομένων δύο κλάσεων ή “1”, “2”, “3”, κ.α όταν αυτά περιέχουν δεδομένα παραπάνω από δύο κλάσεων.

Για να χρησιμοποιηθούν τα αρχικά δεδομένα στην επίλυση ενός προβλήματος, θα πρέπει να υποστούν επεξεργασία. Η επεξεργασία τους γίνεται, τόσο για λόγους συμβατότητας με τον αλγόριθμο που θα τροφοδοτηθούν, όσο και για να αναδειχθεί η χρήσιμη πληροφορία που μπορούν να προσφέρουν. Ανάλογα με τον αλγόριθμο, τα δεδομένα αυτά μπορεί να χρειαστεί

να αναπαρασταθούν όχι μέσω της αρχικής τους μορφής, ως ακατέργαστα (raw) δεδομένα δηλαδή, αλλά μέσω κάποιων ποσοτήτων που θα τα χαρακτηρίζουν. Οι μετρήσιμες ποσότητες που προκύπτουν από πράξεις επί των αρχικών δεδομένων και τα διαφοροποιούν από τα υπόλοιπα, ονομάζονται γνωρίσματα ή **χαρακτηριστικά** (features) [2, 3, 9]. Τις περισσότερες φορές, τα αρχικά δεδομένα ενός συνόλου, είναι πολύ σύνθετα για να υπάρξει ένα και μόνο χαρακτηριστικό το οποίο διαφοροποιεί το κάθε αντικείμενο από τα υπόλοιπα. Για αυτό το λόγο, επιδιώκεται η χρήση πολλών και ανεξάρτητων μεταξύ τους χαρακτηριστικών, τα οποία θα βοηθήσουν στην παραπάνω διαφοροποίηση των δεδομένων. Η αναπαράσταση του συνόλου των χαρακτηριστικών που χρησιμοποιούνται για να περιγράψουν τα δεδομένα, γίνεται με τη μορφή του **διανύσματος χαρακτηριστικών πολλών διαστάσεων** (multidimensional feature vector) [2, 10]. Η δημιουργία του διανύσματος χαρακτηριστικών θα διέπεται από τους ίδιους κανόνες, ανεξαρτήτως της εικόνας εισόδου. Τα δεδομένα, έπειτα από την αναπαράστασή τους από τις ποσότητες των χαρακτηριστικών τους, μπορούν να αναπαρασταθούν ως σημεία, στο **χώρο των χαρακτηριστικών** (feature space), ένα μαθηματικό διανυσματικό χώρο πολλών διαστάσεων, που διαισθητικά μοιάζει με έναν “επίπεδο” Ευκλείδειο χώρο.



Εικόνα 1: Μια αρχική εικόνα και τα κυριότερα χαρακτηριστικά που την περιγράφουν, με βάση έναν αλγόριθμο επιλογής χαρακτηριστικών¹.

Η **Υπολογιστική Όραση** (Computer Vision) [11], είναι το υποπεδίο της Υπολογιστικής Νοημοσύνης, που ασχολείται με την υλοποίηση της όρασης σε υπολογιστές. Συγκεκριμένα, ο κλάδος αυτός ασχολείται με τη λήψη, την επεξεργασία και την ανάλυση σκηνών (scene analysis) με σκοπό την εξαγωγή πληροφοριών μέσα από αυτές. Οι σκηνές αυτές μπορεί να προέρχονται από ήδη υπάρχουσες ψηφιακές εικόνες ή ακόμα και να δημιουργούνται εκείνη τη στιγμή, μέσα από μια ζωντανή ροή βίντεο. Στην παρούσα εργασία, τα δεδομένα που τροφοδοτούνται στον αλγόριθμο που θα αναλυθεί σε επόμενη ενότητα, είναι σε μορφή ψηφιακών εικόνων οι οποίες προέρχονται από διάφορα Σύνολα Δεδομένων (Datasets). Οι ψηφιακές αυτές εικόνες όπως θα παρουσιαστεί τη συνέχεια, επεξεργάζονται με σκοπό να αναδειχθεί η ωφέλιμη πληροφορία που περιέχουν και αναλύονται με σκοπό να γίνει διαφοροποίηση μεταξύ των στοιχείων που απεικονίζουν. Οι παραπάνω διαδικασίες αφορούν το κλάδο της Υπολογιστικής Όρασης. Η διαδικασία αυτή, αποτελεί ένα από τα σημαντικότερα στάδια των εφαρμογών που χρησιμοποιούν την Υπολογιστική Όραση, καθώς η ποιότητα της πληροφορίας που συλλέγεται, παίζει καθοριστικό ρόλο στην ολική επίδοση της εφαρμογής.

Η **Μηχανική Μάθηση** (Machine Learning) [12], είναι το σύνολο των διαδικασιών κατά τις οποίες ένα σύστημα, μαθαίνει από δεδομένα με στόχο να ανταποκριθεί σε προβλήματα για τα οποία δεν έχει ρητά προγραμματιστεί. Ο όρος Μηχανική Μάθηση, επινοήθηκε από τον

¹Bag of Features: Simplifying Image Recognition for Non-Experts

Arthur Samuel και αποτελεί υποπεδίο της Υπολογιστικής Νοημοσύνης και της Αναγνώρισης Προτύπων. Αυτό που ξεχωρίζει τη Μηχανική Μάθηση από την Αναγνώριση Προτύπων, είναι το γεγονός ότι οι αλγόριθμοι Μηχανικής Μάθησης έχουν τη δυνατότητα της τροποποίησης των παραμέτρων τους με βάση τα δεδομένα με τα οποία τροφοδοτούνται, πετυχαίνοντας έτσι την προσαρμογή τους στις ιδιαιτερότητες της εκάστοτε εφαρμογής. Συνεπώς, οι αλγόριθμοι Μηχανικής Μάθησης δε σχεδιάζονται μόνο για μια εφαρμογή, αλλά με σκοπό την επίλυση ενός τύπου προβλήματος, όπως θα παρουσιαστεί στη συνέχεια. Η εκπαίδευσή τους ή μάθηση όπως χαρακτηρίζεται, είναι η διαδικασία κατά την οποία γίνεται τροποποίηση των παραμέτρων του αλγορίθμου, για την προσαρμογή του στα δεδομένα με τα οποία τροφοδοτείται. Από αυτή τη διαδικασία μαθαίνονται οι σχέσεις που διέπουν την είσοδο με την έξοδο και αποθηκεύονται με τη μορφή παραμέτρων. Οι αλγόριθμοι Μηχανικής Μάθησης, χωρίζονται σε τέσσερις κύριες κατηγορίες, ως προς το τρόπο που γίνεται η εκπαίδευσή τους [13]: στην Επιβλεπόμενη, στη Μη-Επιβλεπόμενη, στην Ημιεπιβλεπόμενη και στην Ενισχυτική μάθηση.

- **Επιβλεπόμενη Μάθηση (Supervised Learning):** Η εκπαίδευση των αλγορίθμων αυτής της κατηγορίας γίνεται μέσω παραδειγμάτων. Τα δεδομένα με τα οποία τροφοδοτείται ένας αλγόριθμος, συνοδεύονται από επιπλέον πληροφορία, η οποία ορίζει το πλαίσιο στο οποίο θα γίνουν οι προβλέψεις του. Παραδείγματα προβλημάτων που επιλύουν οι αλγόριθμοι αυτής της κατηγορίας, αποτελούν τα προβλήματα Ταξινόμησης και Παλινδρόμησης.
- **Μη-Επιβλεπόμενη Μάθηση (Unsupervised Learning):** Σε αυτή την κατηγορία, οι αλγόριθμοι αναζητούν αυτόνομα τη δομή και τη συσχέτιση των δεδομένων με τα οποία τροφοδοτούνται. Τα δεδομένα που παρέχονται, δεν περιέχουν κάποια προγενέστερη πληροφορία όσο αφορά τις ιδιότητές τους και τη συσχέτισή τους. Παραδείγματα προβλημάτων που επιλύουν οι αλγόριθμοι αυτής της κατηγορίας αποτελούν τα προβλήματα Συσταδοποίησης και Μείωσης Διάστασης.
- **Ημιεπιβλεπόμενη Μάθηση (Semi-Supervised Learning):** Η τροφοδότηση των αλγορίθμων αυτής της κατηγορίας γίνεται και με δεδομένα με ετικέτα, αλλά και με δεδομένα χωρίς. Σκοπός των αλγορίθμων αυτής της κατηγορίας, είναι να συσχετίσουν τα δεδομένα που συνοδεύονται από ετικέτα, με τα δεδομένα χωρίς.
- **Ενισχυτική Μάθηση (Reinforcement Learning):** Σε αυτούς τους αλγόριθμους δίνονται ένα σετ από επιλογές, παραμέτρους και εξόδους καθώς και κάποιοι κανόνες. Με βάση τις επιλογές που γίνονται από τους αλγόριθμους, προκύπτει κάποιο αποτέλεσμα. Ο αλγόριθμος προσαρμόζει τις παραμέτρους του, με βάση την αξιολόγηση του αποτελέσματος που του επιφέρουν οι πράξεις του.

Όταν για το πρόβλημα μιας εφαρμογής, γίνεται η επιλογή της επίλυσής του με αλγόριθμους Μηχανικής Μάθησης, αυτό που προέχει να αποφασιστεί δεν είναι το πώς θα μάθει ένας αλγόριθμος από τα δεδομένα, αλλά το αποτέλεσμα στο οποίο επιθυμούμε αυτός να καταλήξει. Προηγουμένως αναφέρθηκαν τα κυριότερα είδη Μηχανικής Μάθησης που υπάρχουν, με σκοπό να γίνει μια εισαγωγή στις μεθόδους με τις οποίες οι αλγόριθμοι μαθαίνουν. Παρακάτω, θα αναλυθούν τα κυριότερα είδη προβλημάτων που επιλύονται από αλγόριθμους Μηχανικής Μάθησης:

- **Ταξινόμηση (Classification) [2, 14, 15]:** Στόχος της, είναι η κατηγοριοποίηση αντικειμένων σε κλάσεις (classes). Ο αλγόριθμος μαθαίνει να διαχωρίζει δεδομένα γνωστών κλάσεων, και στη συνέχεια καλείται να κατατάξει νέα δεδομένα, άγνωστης κλάσης, στην

κατηγορία την οποία εκτιμά ότι αυτά ανήκουν. Η Ταξινόμηση αποτελεί εφαρμογή της επιβλεπόμενης Μηχανικής Μάθησης, αφού για τη διαδικασία της μάθησης, ο αλγόριθμος χρειάζεται να έχει γνώση των της κατηγορίας στην οποία εντάσσεται ένα αντικείμενο. Η αντίστοιχη διαδικασία μη-επιβλεπόμενης Μηχανικής Μάθησης είναι η Συσταδοποίηση, η οποία θα αναλυθεί στη συνέχεια.

- **Παλινδρόμηση (Regression)** [12]: Αυτή η διαδικασία, έχει στόχο την προσέγγιση μιας συνάρτησης, δεδομένου ότι κάποια σημεία της είναι γνωστά. Συγκεκριμένα, ο αλγόριθμος μαθαίνει τη σχέση μεταξύ ενός σετ ανεξάρτητων μεταβλητών (εισόδου) και της εξαρτημένης μεταβλητής (εξόδου). Τέλος, καλείται να προσδιορίσει την τιμή της εξόδου του για ένα σετ μεταβλητών εισόδου πάνω στις οποίες δεν έχει εκπαιδευτεί. Όπως γίνεται αντιληπτό, οι αλγόριθμοι που εκτελούν Παλινδρόμηση εκπαιδεύονται με επίβλεψη, αφού παρέχονται δεδομένα τα οποία θεωρούνται σωστά και κατευθύνουν το πλαίσιο στο οποίο θα πρέπει να κινηθεί ο αλγόριθμος.
- **Συσταδοποίηση ή Ομαδοποίηση (Clustering)** [2, 12, 14, 15]: Είναι η διαδικασία κατά την οποία, ένας αλγόριθμος ανακαλύπτει την κατανομή και τη δομή των δεδομένων με τα οποία τροφοδοτείται, με σκοπό να τα ομαδοποιήσει. Σε αντίθεση με την Ταξινόμηση, τα δεδομένα αυτά δε συνοδεύονται από ετικέτες, επομένως οι κλάσεις τους είναι άγνωστες. Όπως και στην Ταξινόμηση, τα δεδομένα τα οποία ανήκουν στην ίδια ομάδα θεωρούνται όμοια, ενώ δύο αντικείμενα δύο διαφορετικών ομάδων θεωρούνται ανόμοια.
- **Μείωση Διάστασης (Dimensionality Reduction)** [2, 14, 15]: Είναι η διαδικασία κατά την οποία γίνεται μείωση του αριθμού των χαρακτηριστικών που χρειάζονται για την αναπαράσταση των δεδομένων, διατηρώντας τη χρήσιμη πληροφορία που τα χαρακτηρίζει. Οι μέθοδοι Μείωσης Διαστάσεων είναι πολλές, γενικά όμως, οι αλγόριθμοι που εκτελούν αυτή τη διαδικασία ανακαλύπτουν ποια χαρακτηριστικά προσφέρουν παραπάνω πληροφορία και απορρίπτουν εκείνα που δεν είναι χρήσιμα. Η διαδικασία αυτή είναι γνωστή και ως μείωση της διάστασης των δεδομένων και γίνεται με σκοπό να καταπολεμηθεί το φαινόμενο της Κατάρας της Διαστατικότητας (Curse of Dimensionality), όπου όσο η διαστατικότητα των δεδομένων αυξάνεται, τόσο πιο αραιά τοποθετούνται αυτά στο χώρο. Επίσης, όσο περισσότερα χαρακτηριστικά χρησιμοποιούνται για την αναπαράσταση των δεδομένων, τόσο μεγαλύτερο είναι και το υπολογιστικό κόστος του αλγορίθμου σε μνήμη, αποθηκευτικό χώρο και επεξεργαστική ισχύ. Αυτό το είδος προβλήματος μπορεί να χρησιμοποιηθεί και ως ενδιάμεσο στάδιο, με σκοπό να γίνει επίλυση άλλων προβλημάτων που αναφέρθηκαν προηγουμένως, με δεδομένα απαλλαγμένα από τα παραπάνω μειονεκτήματα.

1.1 Ιστορική Αναδρομή

Η Υπολογιστική Όραση αποτελεί ένα σχετικά πρόσφατο κεφάλαιο της Υπολογιστικής Νοημοσύνης, αφού όπως θα δούμε παρακάτω, ξεκίνησε σαν επέκτασή της, αλλά κατέληξε να είναι ένα ευρύ και σπουδαίο πεδίο της, πάνω στο οποίο βασίζονται πολλές εφαρμογές, οι οποίες δε θα υπήρχαν χωρίς την εξέλιξή του.

1.1.1 Δεκαετία 1970: Η γέννηση της Υπολογιστικής Όρασης

Η Υπολογιστική Όραση ξεκίνησε να αναπτύσσεται τη δεκαετία του 1970, όπου οι τότε ερευνητές επιθυμούσαν να τη χρησιμοποιήσουν ως συμπλήρωμα της τότε Τεχνητής Νοημοσύνης, στην κατασκευή συστημάτων που θα μιμούνταν την ανθρώπινη νοημοσύνη [11]. Για εκεί-

νους, η Υπολογιστική Όραση αποτελούσε το εύκολο κομμάτι του παζλ στην κατασκευή ενός ολοκληρωμένου ευφυούς συστήματος, ενώ διαδικασίες όπως η λογική υψηλού επιπέδου και η λήψη αποφάσεων θεωρούνταν δυσκολότερες. Πλέον βέβαια, είναι αντιληπτή η δυσκολία και η σημασία ανάπτυξης εύρωστων (robust) συστημάτων Υπολογιστικής Όρασης. Αυτό που διαφοροποίησε την Υπολογιστική Όραση από την Ψηφιακή Επεξεργασία Εικόνας εκείνη την εποχή, ήταν η πρόθεση των ερευνητών να ανακτήσουν την τρισδιάστατη πληροφορία μιας σκηνής. Είχαν σκοπό δηλαδή, να προσδιορίσουν το βάθος ενός αντικειμένου στο χώρο που απεικονίζεται σε μια σκηνή. Μια από τις μεθόδους που χρησιμοποιήθηκαν αυτή την περίοδο, η οποία χρησιμοποιείται μέχρι και σήμερα σε πολλές εφαρμογές, είναι η αναγνώριση ακμών (edge detection) σε αντικείμενα για την εξαγωγή του σχήματός τους, μια ιδέα η οποία έδωσε το έναυσμα για την ανάλυση εικόνων με χαρακτηριστικά χαμηλού επιπέδου (low-level features).

1.1.2 Δεκαετία 1980: Νέες τεχνικές και μαθηματική μοντελοποίηση

Κατά την επόμενη δεκαετία, δόθηκε έμφαση στην ανάπτυξη πιο εξελιγμένων τεχνικών εξαγωγής πληροφορίας από σκηνές, οι οποίες είχαν σαφώς οριοθετημένο μαθηματικό υπόβαθρο. Τέτοιες είναι η χωρική κλίμακα (scale space) [16–18], η αναπαράσταση πυραμίδας (pyramid representation) [19], με τη δεύτερη να είναι η βελτίωση της πρώτης. Οι τεχνικές αυτές δέχονται μια εικόνα και παράγουν ένα σύνολο από θολωμένες και μόνο για την περίπτωση της αναπαράστασης πυραμίδας, υποβιβασμένες σε ανάλυση εικόνες (Εικόνα 2a). Συγκεκριμένα, οι εικόνες αυτές παράγονται ανά επίπεδα, όπου κάθε επίπεδο έχει και ένα διαφορετικό βαθμό θόλωσης και υποβιβασμού ανάλυσης, με σκοπό την εκτίμηση της τρισδιάστατης πληροφορίας που κρύβεται σε μια δισδιάστατη εικόνα, μέσω της προσέγγισης της τοποθεσίας (του βάθους) των αντικειμένων που απεικονίζονται σε μια σκηνή. Εκείνη την περίοδο αναπτύχθηκε ο αλγόριθμος του μετασχηματισμού εικόνας σε wavelets (Discrete Wavelet Transform) [20, 21], μια τεχνική παρόμοια με τον μετασχηματισμό Fourier, όπου μια εικόνα αναλύεται στις συχνότητες που την απαρτίζουν.

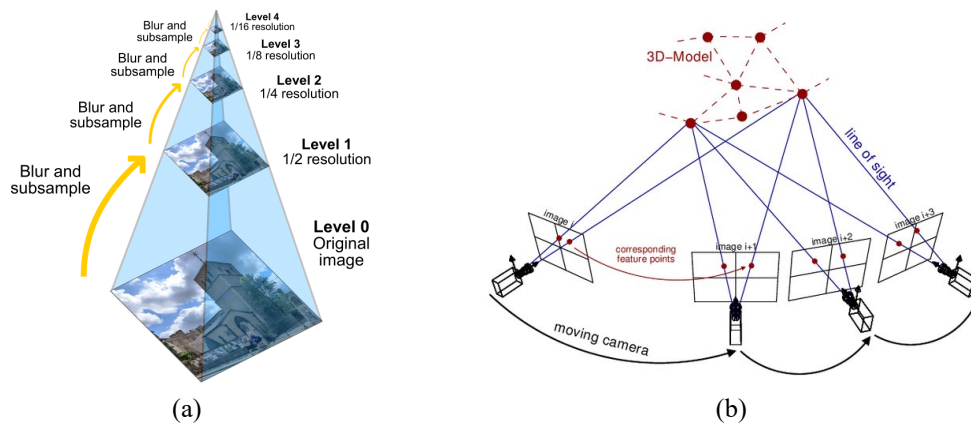
Άλλες τεχνικές που αναπτύχθηκαν εκείνη την περίοδο, αφορούν την ανάκτηση του σχήματος ενός αντικειμένου, από τη μελέτη της πληροφορίας που παρέχεται από τη φωτεινότητα και τις σκιές του, την υφή (texture) του ή την εστίαση της κάμερας στο αντικείμενο αυτό, κ.α. Παράλληλα, δόθηκε βάση στη βελτίωση των ήδη υπαρχών τεχνικών αναγνώρισης ακμών και αναπτύχθηκαν αξιοσημείωτοι αλγόριθμοι. Ένας από αυτούς, ο οποίος είναι πολύ γνωστός και χρησιμοποιείται μέχρι σήμερα, είναι ο αλγόριθμος αναγνώρισης ακμών του Canny, που κατασκευάστηκε το 1986 [22]. Επιπλέον, μια σημαντική διαπίστωση που έγινε από τους ερευνητές εκείνης της περιόδου η οποία αφορά το κομμάτι των μαθηματικών, είναι ότι, όπως παρατηρήθηκε, πολλές τεχνικές που είχαν μέχρι τώρα αναπτυχθεί (ανίχνευση ακμών, ανάκτηση σχήματος μέσω διάφορων μεθόδων, κ.α), μπορούσαν να γενικευτούν και να εκφραστούν ως μοντέλα της θεωρίας των τυχαίων Μαρκοβιανών πεδίων (Markov Random Field). Αυτή η διαπίστωση είναι σπουδαία, διότι οι τεχνικές αυτές στη γενικευμένη τους μορφή μπορούν να χρησιμοποιηθούν σε συνδυασμό με αλγόριθμους βελτιστοποίησης, για την εύρεση βέλτιστων λύσεων σε εφαρμογές.

1.1.3 Δεκαετία 1990: 3D Γραφικά

Κατά αυτή την περίοδο, δόθηκε έμφαση στην επίλυση προβλημάτων τα οποία είχαν ήδη ξεκινήσει να διερευνώνται την προηγούμενη δεκαετία. Ένα από τα προβλήματα αυτά, είναι εκείνο της ανάκτησης της δομής ενός χώρου μέσω μιας σειράς εικόνων, το οποίο ονομάζεται δομή από κίνηση (structure from motion) (Εικόνα 2b) [23, 24]. Λόγω της ενασχόλησης των ερευνητών με την επίλυση αυτού του προβλήματος, η οποία έγινε με διάφορες προσεγγίσεις, εξοπλίστηκαν με παραπάνω κατανόηση για τη βαθμονόμηση καμερών. Αργότερα, έφτασαν στο

συμπέρασμα πως οι τεχνικές που υλοποιούσαν ήταν οι ίδιες με αυτές που χρησιμοποιούνταν σε έναν άλλο κλάδο της επιστήμης που ονομάζεται φωτογραμμετρία (photogrammetry), ο οποίος ασχολείται με τον προσδιορισμό διαστάσεων αντικειμένων από φωτογραφίες [25]. Λόγω αυτού, κατασκευάστηκαν πλήρως αυτοματοποιημένα μοντέλα τα οποία υλοποιούσαν τεχνικές βελτιστοποίησης και εκτελούσαν τη διαδικασία καταγραφής της δομής από κίνηση. Κάποιες ακόμα τεχνικές που είδαν ανάπτυξη ήταν η ιχνηλασία αντικειμένων από το περίγραμμά τους (contour tracking) και η κατάτμηση αντικειμένων από το υπόβαθρο της σκηνής στην οποία απεικονίζονται (image segmentation).

Λόγω της ανάπτυξης μεθόδων στατιστικής μάθησης (statistical learning) εκείνης της περιόδου, αναπτύχθηκε η πρώτη εφαρμογή αναγνώρισης προσώπων, ονόματι eigenfaces, από τους Kirby και Sirovich το 1990. Μια επιτυχία εκείνης της περιόδου ήταν η δυνατότητα αλληλεπίδρασης εφαρμογών της Υπολογιστικής Όρασης με γραφικά υπολογιστών, το οποίο επέτρεψε τη σύνθεση σταθερών (image morphing) και κινούμενων εικόνων (image-based model rendering). Άλλες παρόμοιες εφαρμογές ήταν η συρραφή εικόνων (image stitching) από τις οποίες προέκυπταν πανοραμικές εικόνες, καθώς και η κατασκευή τρισδιάστατων μοντέλων από μια συλλογή φωτογραφιών.



Εικόνα 2: (a) Αναπαράσταση πυραμίδας μιας εικόνας² (b) Παράδειγμα ανάκτησης της δομής μιας σκηνής με μια εφαρμογή structure from motion³.

1.1.4 Δεκαετία 2000: Υπολογιστική Φωτογραφία

Κατά τη δεκαετία αυτή, συνεχίστηκε η αλληλεπίδραση μεταξύ γραφικών υπολογιστών και της Υπολογιστικής Όρασης, ενώ αυτή παράλληλα βελτιωνόταν και οι εφαρμογές της διευρύνονταν από την ολοένα και περισσότερο εισαγωγή της Μηχανικής Μάθησης στο κλάδο. Πολλές εφαρμογές που αναφέρθηκαν προηγουμένως, όπως η σύνθεση εικόνων με στοιχεία φωτογραφιών, χρησιμοποιώντας γραφικά υπολογιστών (image-based rendering), η φωτογράφιση με υψηλό δυναμικό εύρος (high dynamic range, HDR) (Εικόνα 3a) και η συρραφή εικόνων (image stitching), κατατάχτηκαν σε ένα κλάδο που ονομάζεται υπολογιστική φωτογραφία (computational photography), λόγω του σε μεγάλο βαθμό συνδυασμού αλγορίθμων επεξεργασίας και ανάλυσης εικόνων με την κλασική φωτογραφία. Οι δύο τελευταίες εφαρμογές υπάρχουν ενσωματωμένες σε κάθε σύγχρονο κινητό τηλέφωνο. Άλλες εξίσου σημαντικές εφαρμογές αποτελεί

²Pyramid (image processing)

³Exploring the use of 3D GIS as an analytical tool in archaeological excavation practice

η σύνθεση υφής (texture synthesis) και ο συνδυασμός της έκθεσης (exposure) φωτογραφιών που απεικονίζουν την ίδια ή κομμάτια της ίδιας σκηνής, πετυχαίνοντας παρόμοιο αποτέλεσμα με τη φωτογράφιση υψηλού δυναμικού εύρους (HDR).

Μια σημαντική τάση της περιόδου αυτής, η οποία είναι σχετική με την παρούσα διπλωματική εργασία, ήταν η προσπάθεια αναγνώρισης αντικειμένων μέσω χαρακτηριστικών (feature-based recognition), η οποία επιτυγχάνεται με τον συνδυασμό τεχνικών Μηχανικής Μάθησης και βελτιστοποίησης. Πέρα από την παραπάνω εφαρμογή της Υπολογιστικής Όρασης, πολλές άλλες άνθισαν και ακόμα περισσότερες γεννήθηκαν λόγω του συνδυασμού της με τη Μηχανική Μάθηση. Σε συνδυασμό με την τις πλέον αυξημένες δυνατότητες των υπολογιστικών συστημάτων της εποχής και τη διαθεσιμότητα πληθώρας εικόνων από το ίντερνετ, η Υπολογιστική Όραση εξελίχθηκε στο κλάδο που αναγνωρίζουμε σήμερα.

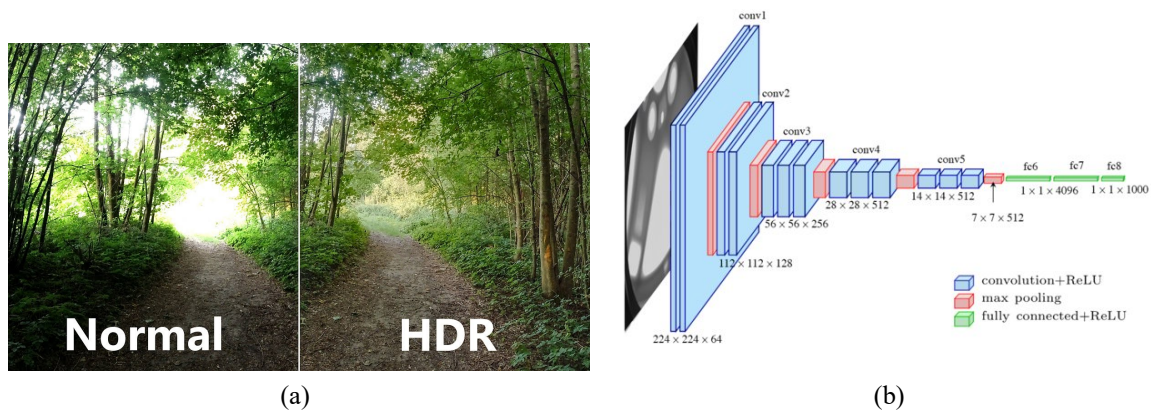
1.1.5 Δεκαετία 2010: Σύγχρονη Υπολογιστική Όραση

Η ύπαρξη μεγάλου πλήθους δεδομένων (εικόνων) στο ίντερνετ, πολλές φορές με ετικέτα και σε κάποιες περιπτώσεις υψηλής ποιότητας, όπως τα ImageNet, Microsoft COCO και LVIS, ώθησε την κατασκευή αλγόριθμων Μηχανικής Μάθησης οι οποίοι είχαν μεγάλη επιτυχία στην αναγνώριση αντικειμένων. Επιπλέον, λόγω της αυξημένης δυνατότητας σε υπολογιστική ισχύ και την ικανότητα της παράλληλης επεξεργασίας που παρεχόταν από τις κάρτες γραφικών (graphical processing units, GPUs), κατασκευάστηκαν αλγόριθμοι Μηχανικής Μάθησης οι οποίοι παρόλο που ήταν απαιτητικοί σε υπολογιστικούς πόρους είχαν μεγάλη επιτυχία στην αναγνώριση αντικειμένων. Οι αλγόριθμοι στους οποίους γίνεται αναφορά δεν είναι όμοιοι με τους κλασικούς αλγόριθμους Μηχανικής Μάθησης και κατατάσσονται στην κατηγορία των αλγορίθμων βαθιάς μάθησης (deep learning algorithms). Οι αλγόριθμοι αυτοί αποτελούν μια κατηγορία νευρωνικών δικτύων και ονομάζονται νευρωνικά δίκτυα βαθιάς μάθησης (deep learning neural networks ή deep neural networks) [26] (Εικόνα 3b).

Το είδος των νευρωνικών αυτών δικτύων, περιέχουν πολλούς περισσότερους νευρώνες από ένα κλασικό νευρωνικό δίκτυο. Λόγω του αυξημένου πλήθους των νευρώνων τους, τα δίκτυα αυτά έχουν τη δυνατότητα να μαθαίνουν πιο περίπλοκα πρότυπα που χαρακτηρίζουν τα δεδομένα και ως αποτέλεσμα, να έχουν συνήθως καλύτερη επίδοση στις εφαρμογές που καλούνται να επιλύσουν. Μειονεκτούν στο γεγονός ότι για να είναι αποτελεσματική η εκπαίδευσή τους, απαιτείται η τροφοδοσία τους με πολλά παραπάνω δεδομένα, σε σύγκριση με αυτά που θα απαιτούσε ένας άλλος αλγόριθμος Μηχανικής Μάθησης και ως επακόλουθο αυτού, οι απαιτήσεις τους σε υπολογιστικούς πόρους είναι μεγαλύτερες και η εκπαίδευσή τους πιο χρονοβόρα, καθώς το πλήθος των νευρώνων τους είναι μεγαλύτερο. Ένα δημοφιλές δίκτυο βαθιάς μάθησης (deep neural network) που είχε αναδειχθεί τότε, είναι το AlexNet [27], το οποίο είχε μεγάλη επιτυχία και η δημοσίευσή του στο κοινό ώθησε σε μεγάλο βαθμό την ανάπτυξη και άλλων δικτύων βαθιάς μάθησης. Ακόμα και σήμερα, τα βαθιά νευρωνικά δίκτυα προτιμούνται έναντι κλασικών νευρωνικών δικτύων και άλλων αλγορίθμων Μηχανικής Μάθησης, σε διάφορες εφαρμογές πολλών κλάδων, συμπεριλαμβανομένου και της Υπολογιστικής Όρασης. Κάποιες από αυτές είναι αναγνώριση ανθρώπινων κινήσεων (action recognition) και πόζας (body pose estimation) σε εικόνες και βίντεο.

Περαιτέρω ανάπτυξη είδαν και αισθητήρες ειδικά κατασκευασμένοι για εφαρμογές Υπολογιστικής Όρασης, όπως οι αισθητήρες βάθους (depth sensors), οι οποίοι χρησιμοποιούνται για την ιχνηλασία ανθρώπινων κινήσεων (human movement tracking) και τη σύνθεση τρισδιάστατων μοντέλων (3D model rendering). Λόγω διάφορων παραγόντων όμως, όπως του κόστους τους, χρησιμοποιούνται από εφαρμογές που μπορούν να τους υποστηρίξουν. Άλλες εφαρμογές αρκούνται στη χρήση λογισμικού, για τη λήψη του βάθους μιας σκηνής. Κάποιες εφαρμογές

που θέλουν την πραγματική μέτρηση του βάθους και χρησιμοποιούν αυτούς τους αισθητήρες είναι η αυτόνομη πλοήγηση, σε επίγεια και εναέρια οχήματα, μέσω της χαρτογράφησης περιβάλλοντος με ταυτόχρονη ιχνηλασία θέσης (simultaneous localization and mapping, SLAM) [28], καθώς και της οπτικής αδρανειακής οδομετρίας (visual inertial odometry) [29]. Εν κατακλείδι, η δεκαετία αυτή χαρακτηρίζεται από τον συνδυασμό της Υπολογιστικής Όρασης με τη Μηχανική Μάθηση, η οποία αποτελεί πλέον το εργαλείο με το οποίο ακόμα πιο περίπλοκες και εντυπωσιακές εφαρμογές μπορούν να υλοποιηθούν.



Εικόνα 3: (a) Φωτογραφία του ίδιου τοπίου με και χωρίς τη χρήση HDR⁴ (b) Αρχιτεκτονική του νευρωνικού δικτύου βαθιάς μάθησης VGG-16⁵.

1.2 Εφαρμογές

Η ενσωμάτωση της Υπολογιστικής Όρασης στην Υπολογιστική Νοημοσύνη, ανοίγει τις πόρτες στην αυτοματοποίηση μυριάδων εφαρμογών που επωφελούνται από την υλοποίηση της ψηφιακής όρασης. Μερικές από αυτές είναι οι παρακάτω [30]:

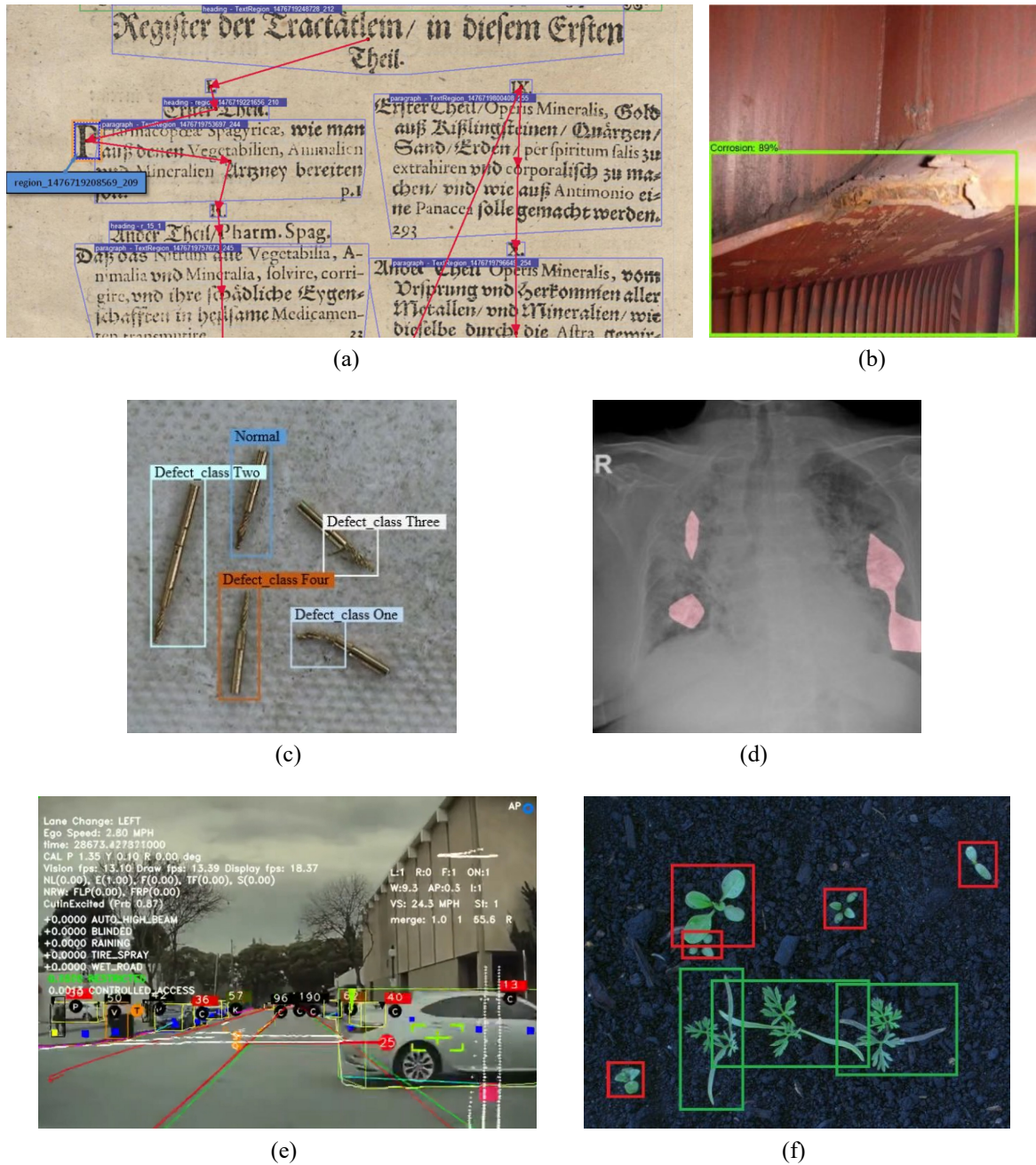
- **Οπτική Αναγνώριση Χαρακτήρων:** Αποτελεί μια από τις πρώτες εφαρμογές της Υπολογιστικής Όρασης, η οποία χρησιμοποιείται μέχρι και σήμερα, για την αναγνώριση χαρακτήρων σε χειρόγραφα κείμενα με σκοπό την ψηφιοποίησή τους (Εικόνα 4a).
- **Επιθεώρηση Υποδομών και Εξοπλισμού:** Η επιθεώρηση υποδομών χρησιμοποιείται κατά κόρον για την αξιολόγηση της κατάστασης σταθερών εγκαταστάσεων όπως κτηρίων, γεφυρών, κεραιών, γραμμών μεταφοράς αλλά και κινητών εγκαταστάσεων, όπως αεροπλάνων, πλοίων και άλλων ογκωδών μηχανημάτων (Εικόνα 4b). Σαν διαδικασία μπορεί να γίνει γρήγορα και απομακρυσμένα, από μη-επανδρωμένα εναέρια οχήματα με πλεονεκτήματα όπως αυτό της ασφάλειας του επιθεωρητή και της δυνατότητας επιθεώρησης σε δύσβατα ή αδύνατα στην πρόσβασή τους σημεία μιας εγκατάστασης.
- **Ποιοτικός Έλεγχος:** Μια ακόμα συχνή χρήση της Υπολογιστικής Όρασης βρίσκεται στις γραμμές παραγωγής, όπου ο ποιοτικός έλεγχος μπορεί να γίνεται άμεσα από έξυπνα συστήματα, με ελάχιστη ή και καθόλου ανθρώπινη παρέμβαση. Τα συστήματα αυτά χρησιμοποιούν κάμερες ή και άλλους αισθητήρες, καταγράφοντας τη μορφή των προϊόντων

⁴HDR Photography: A Practical Guide

⁵Automatic localization of casting defects with convolutional neural networks

που εξάγει η γραμμή παραγωγής, με σκοπό την κατηγοριοποίησή τους σε ελαττωματικά ή λειτουργικά (Εικόνα 4c).

- **Ιατρικές Απεικονίσεις:** Η επιστήμη της Ιατρικής επωφελείται από τη χρήση της Υπολογιστικής Όρασης για την υποβοήθηση ιατρών, στη διάγνωση ασθενειών όπως του καρκίνου, των τραυμάτων των εγκεφαλικών επεισοδίων, της πνευμονίας και πρόσφατα του COVID-19, μέσω της ανάλυσης ιατρικών απεικονίσεων όπως των ακτινογραφιών (Εικόνα 4d), των υπερήχων, των μαγνητικών και αξονικών τομογραφιών, των ενδοσκοπήσεων κ.α. Η χρήση της Υπολογιστικής Όρασης σε αυτό το κλάδο έχει βοηθήσει με άμεσο τρόπο την ανθρώπινη ζωή και αναμφισβήτητα αποτελεί μια από της σημαντικότερες εφαρμογές της [31].
- **Παρακολούθηση:** Η χρήσεις της Υπολογιστικής Όρασης σε αυτό το τομέα είναι πολλές και προσφέρουν αυτοματισμό και υποβοήθηση στους χειριστές συστημάτων παρακολούθησης, σε διάφορες διαβαθμίσεις. Παραδείγματα δυνατοτήτων που προσφέρονται είναι η αναγνώριση αντικείμενων, προσώπων, ανθρώπινων κινήσεων και συμπεριφορών, η αναγνώριση παραβιάσεων σε χώρους, η ανάγνωση πινακίδων αυτοκινήτων, η εξαγωγή στατιστικών αναλύσεων και πολλών άλλων [32].
- **Αυτόνομη Οδήγηση:** Εταιρίες αυτοκινήτων όπως η Tesla [33, 34], κατάφεραν να αναπτύξουν, τα τελευταία χρόνια, αυτοκίνητα τα οποία σε ένα βαθμό έχουν τη δυνατότητα της αυτόνομης πλοήγησης. Παρόλο που η αυτόνομη οδήγηση δεν είναι στο στάδιο που μπορεί να αντικαταστήσει έναν ανθρώπινο οδηγό, υπάρχουν λειτουργίες όπως υποβοήθηση στη στάθμευση και την αναγνώριση επικείμενων ατυχημάτων, οι οποίες βασίζονται σε συστήματα όρασης, μεταξύ άλλων, οι οποίες έχουν εξελίξει πολύ την έννοια της σύγχρονης οδήγησης (Εικόνα 4e). Επιπλέον, η αυτόνομη οδήγηση δεν περιορίζεται μόνο στα επίγεια οχήματα, αλλά συναντάται και σε εναέρια οχήματα, στη μορφή αυτόματων πιλότων, σε μη-επανδρωμένα εναέρια οχήματα, όπου αυτά μπορούν να εκτελέσουν περίπλοκες προκαθορισμένες διαδρομές ή να ακολουθούν αντικείμενα, αποφεύγοντας εμπόδια όπως δέντρα και πουλιά [35].
- **Γεωργία και Κτηνοτροφία:** Χρησιμοποιείται για την αυτοματοποίηση και την παρακολούθηση διάφορων διαδικασιών και πραγμάτων, διευκολύνοντας κάποιες διεργασίες, αυξάνοντας έτσι την απόδοση και την ποιότητα των καλλιεργειών και των ζώων που εκτρέφονται. Κάποια παραδείγματα εφαρμογών αποτελούν η παρακολούθηση της υγείας και της συμπεριφοράς ζωντανών, την παρακολούθηση της υγείας και της ανάπτυξης των καλλιεργειών, την εύρεση και τη στοχευμένη εξόντωση παρασίτων (Εικόνα 4f), πετυχαίνοντας έτσι τη λιγότερη χρήση φυτοφαρμάκων και άλλων μη-στοχευμένων μεθόδων προστασίας των καλλιεργειών.
- **Βιομετρική Αναγνώριση:** Χρησιμοποιείται για τη φύλαξη πληροφοριών και την παροχή πρόσβασης σε πληροφορίες ή υπηρεσίες, σε άτομα με τη δικαιοδοσία να τις αποκτήσουν. Λόγω της ασφάλειας που παρέχει, αφού τα βιομετρικά χαρακτηριστικά ενός ατόμου είναι μοναδικά, χρησιμοποιείται ως μέσο ταυτοποίησης για εφαρμογές που χρήζουν υψηλή ασφάλεια. Η Βιομετρική Αναγνώριση μέσω κάμερας αποτελεί μια από τις πιο συχνά χρησιμοποιούμενες εφαρμογές της Υπολογιστικής Όρασης από το μέσο άνθρωπο στην καθημερινότητά του, στη μορφή της αναγνώρισης προσώπου για το ξεκλείδωμα των προσωπικών ηλεκτρονικών του συσκευών, όπως το κινητό τηλέφωνο.



Εικόνα 4: Εφαρμογές της Υπολογιστικής Όρασης: (a) οπτική αναγνώριση χαρακτήρων σε ιστορικό χειρόγραφο⁶, (b) ανίχνευση σκουριάς σε βιομηχανική εγκατάσταση⁷, (c) ποιοτικός έλεγχος σε γραμμή παραγωγής⁸, (d) ανίχνευση προσβεβλημένου ιστού πνεύμονα από παθόντα του COVID-19⁹, (e) ο κόσμος μέσα από τα μάτια ενός αυτοκινήτου Tesla¹⁰, (f) ανίχνευση παρασίτων σε καλλιέργεια¹¹.

⁶Issue 13: OCR

⁷16 Applications of Computer Vision in Construction (2024 Guide)

⁸The 100 Most Popular Computer Vision Applications in 2024

⁹Top 19 Applications Of Deep Learning and Computer Vision In Healthcare

¹⁰This is how Tesla's Autopilot sees the world

¹¹Detecting Weed Through Edge Computer Vision

1.3 Αλγόριθμοι Μηχανικής Μάθησης και Χώρος Χαρακτηριστικών

Σε αυτή τη διπλωματική εργασία, θα αναλυθεί ένας ήδη δημοσιευμένος αλγόριθμος, ο οποίος έχει σαν κύριο στόχο τη Ταξινόμηση δεδομένων, ενώ παράλληλα επιτυγχάνει και τη Μείωση Διαστάσεων τους. Προτού όμως αναφερθούμε στη λειτουργία του αλγορίθμου, είναι αναγκαίο να γίνει μια προεπισκόπηση συνήθων άλλων διαδικασιών που έχουν παρόμοιο στόχο, εκ των οποίων μία από αυτές χρησιμοποιείται σε ένα στάδιο του αλγορίθμου. Η διαδικασία της Ταξινόμησης, όπως έχει ήδη αναφερθεί, είναι η διαδικασία διαχωρισμού αντικειμένων στις σωστές κλάσεις, τις οποίες ανήκουν. Αυτή επιτυγχάνεται με διάφορες μεθόδους, πάνω στις οποίες βασίζονται πολλοί αλγόριθμοι Ταξινόμησης, όπως ο Support Vector Machines (SVM) [36], ο k-Nearest Neighbors (k-NN) [37, 38], ο Naïve Bayes [39], τα δέντρα επιλογών (decision trees) [40], κ.α. Για τον αλγόριθμο που θα αναλυθεί, τα αντικείμενα, κατατάσσονται με βάση το πόσο απέχουν από άλλα αντικείμενα με τα οποία γειτνιάζουν. Η μέθοδος αυτή είναι γνωστή ως Similarity-based Classification και είναι αυτή που υλοποιεί ο αλγόριθμος K-Nearest Neighbors. Ο λόγος που χρησιμοποιείται αυτή η μέθοδος έναντι κάποιας άλλης είναι γιατί το μοναδικό κριτήριο που λαμβάνεται υπόψη για τον προσδιορισμό της κατηγορίας στην οποία ανήκει ένα σημείο, είναι η απόστασή του από άλλα γειτονικά σημεία. Σε αντίθεση με αλγόριθμους όπως του SVM, που για τη λειτουργία του χρησιμοποιεί υπερεπίπεδα (hyperplanes) και βασίζονται στη γεωμετρία του χώρου, στην περίπτωση του Similarity Based Classification, γίνεται η θεώρηση ότι τα σημεία αυτά βρίσκονται σε ένα διανυσματικό χώρο, που ισοδυναμεί με έναν Ευκλείδειο χώρο.

Τα αντικείμενα αυτά αποτελούν τα αρχικά (raw) δεδομένα και από αυτά εξάγονται χαρακτηριστικά, με σκοπό να παρασταθούν στο χώρο χαρακτηριστικών ως σημεία. Με αυτό το τρόπο, ανεξαρτήτως της μορφής που είχαν τα δεδομένα αρχικά (αριθμοί, κείμενο, εικόνα, βίντεο), μετατρέπονται τελικά σε όμοιας δομής μαθηματικές ποσότητες που τα χαρακτηρίζουν, στην οποία μπορούν να εφαρμοστούν διάφορα μαθηματικά εργαλεία και σχέσεις για να εξαχθούν τα επιθυμητά συμπεράσματα. Ο χώρος χαρακτηριστικών που αναπαρίστανται αυτά τα σημεία είναι ένας διανυσματικός χώρος και η δομή του έχει κάποια χαρακτηριστικά ανάλογα με τις γεωμετρικές ιδιότητες των σημείων.

Όταν τα δεδομένα διέπονται από τους κανόνες της Ευκλείδειας γεωμετρίας, τότε ο χώρος στον οποίο βρίσκονται είναι ο Ευκλείδειος διανυσματικός χώρος n διαστάσεων \mathbb{R}^n , στον οποίο αυτά είναι ισότροπα κατανεμημένα. Για την εύρεση των αποστάσεων μεταξύ των σημείων αυτού του χώρου χρησιμοποιείται η Ευκλείδεια απόσταση¹² που προκύπτει από την Ευκλείδεια μετρική, η οποία χρησιμοποιείται για τον προσδιορισμό της απόστασης μεταξύ σημείων του χώρου. Για χρόνια, οι εφαρμογές της Μηχανικής Μάθησης διεξάγονταν με τη θεώρηση ότι τα δεδομένα τα οποία χρησιμοποιούνται κατατάσσονται στον Ευκλείδειο διανυσματικό χώρο. Στην περίπτωση όμως που τα δεδομένα, λόγω των γεωμετρικών τους ιδιοτήτων, δεν κατατάσσονται στον Ευκλείδειο χώρο η διαχείρισή τους βασιζόμενη σε αυτή τη θεώρηση είναι λάθος. Τα δεδομένα τότε ανήκουν σε ένα χώρο του οποίου η δομή υπόκειται σε γεωμετρικούς περιορισμούς λόγω των ιδιοτήτων τους. Η μέτρηση της απόστασης μεταξύ δύο σημείων σε εκείνους τους χώρους δε γίνεται μέσω της Ευκλείδειας απόστασης αλλά μέσω των γεωδειακών, οι οποίες είναι καμπύλες που οδηγούν από το ένα σημείο του στο άλλο διασχίζοντας μόνο σημεία αυτού του χώρου, μη βγαίνοντας εκτός αυτού. Η μέτρηση του μήκους των γεωδειακών δίνει την απόσταση των σημείων αυτού του χώρου.

Σε αρκετές των περιπτώσεων, ο χώρος των χαρακτηριστικών στον οποίο κατατάσσονται

¹²Η Ευκλείδεια απόσταση [41] μεταξύ δύο σημείων, εκφρασμένα ως διανύσματα p και q ορίζεται ως: $d(p, q) = \|p - q\|_2 = \sqrt{\sum_{i=1}^n (p_i - q_i)^2}$, όπου $\|\cdot\|_2$ η Ευκλείδεια νόρμα [42] και n η διάσταση των p και q , με $\dim(p) = \dim(q) = n$

τα δεδομένα είναι μια Riemannian Πολλαπλότητα, ο οποίος όπως αναφέρεται αναλυτικότερα στην επόμενη ενότητα, είναι ένας χώρος με λεία δομή που συνοδεύεται από μία ή περισσότερες Riemannian μετρικές, οι οποίες σέβονται τους γεωμετρικούς περιορισμούς των δεδομένων και είναι κατάλληλες για τη μέτρηση αποστάσεων μεταξύ σημείων του χώρου. Η χρήση της σωστής μετρικής εγγυάται πως η απόσταση μεταξύ σημείων του, θα αντικατοπτρίζει σωστά το πόσο αυτά απέχουν το ένα από το άλλο. Αυτό σημαίνει πως τα όμοια δεδομένα θα απέχουν λίγο το ένα από το άλλο, ενώ τα ανόμοια θα απέχουν πολύ μεταξύ τους.

Τα δεδομένα που χρησιμοποιούνται στην παρούσα διπλωματική εργασία είναι της μορφής πινάκων συνδιακύμανσης και ανήκουν στο Riemannian χώρο των Πολλαπλοτήτων, των Συμμετρικών και Θετικά Ορισμένων Πινάκων (Symmetric Positive Definite, SPD Manifold). Για αυτές τις πολλαπλότητες έχουν δημιουργηθεί μετρικές όπως η Affine Invariant Riemannian Metric (AIRM) και η Log-Euclidean Metric (LEM), που αναλύονται στις υποενότητες 2.8.1 και 2.8.2 αντίστοιχα. Οι μετρικές αυτές έχουν αναπτυχθεί για τις συγκεκριμένες πολλαπλότητες και επιτυγχάνουν τη μέτρηση αποστάσεων μεταξύ σημείων, σεβόμενες τη γεωμετρία του χώρου. Παρόλο που οι μετρικές αυτές έχουν ισχυρό μαθηματικό υπόβαθρο, λόγω των διαφορετικών κατανομών των δεδομένων σε κάθε εφαρμογή, οι παραπάνω σχέσεις μπορεί να μην είναι τόσο ακριβείς στον προσδιορισμό της απόστασης μεταξύ σημείων. Ένας τρόπος προσέγγισης αυτού του προβλήματος είναι η εκμάθηση της σχέσης που εκφράζει καλύτερα την απόσταση μεταξύ σημείων. Η διαδικασία αυτή ονομάζεται **Εκμάθηση Μετρικής** (Metric Learning) και αποτελεί μια εφαρμογή επιβλεπόμενης μάθησης. Η Εκμάθηση Μετρικής επιτυγχάνεται μέσω της τροφοδοσίας ενός αλγορίθμου με ζεύγη όμοιων και ανόμοιων σημείων της πολλαπλότητας στην οποία ανήκουν. Αυτή η διαδικασία χρησιμοποιείται από τον παρών αλγόριθμο, για τον ακριβέστερο και γρηγορότερο προσδιορισμό της απόστασης των δεδομένων. Έπειτα, αλγόριθμοι όπως ο k -Nearest Neighbors χρησιμοποιούνται για να προσδιοριστεί η κατηγορία στην οποία ανήκει ένα σημείο, με βάση την απόστασή του από τα πλησιέστερα σε εκείνο σημεία.

1.3.1 Βασικά στοιχεία της Οπισθοδιάδοσης

Η **Οπισθοδιάδοση** (Backpropagation), είναι μια από τις διαδικασίες που ακολουθούνται για την εκπαίδευση, τη διόρθωση δηλαδή της προσαρμογής ενός μοντέλου Μηχανικής Μάθησης [43]. Κατά τη λειτουργία ενός αλγορίθμου Μηχανικής Μάθησης, η Οπισθοδιάδοση χρησιμοποιείται ως ακολούθως:

- **Εμπρόσθια Τροφοδότηση:** Μετάδοση των δεδομένων εισόδου μέσα από το μοντέλο και υπολογισμός της εξόδου του.
- **Οπισθοδιάδοση:** Διόρθωση των παραμέτρων του μοντέλου για τη μείωση του σφάλματος της εξόδου του, μέσω των παρακάτω βημάτων:
 - Υπολογισμός του σφάλματος, δηλαδή της διαφοράς μεταξύ της εξόδου που υπολογίστηκε (πρόβλεψη, \hat{y}) και της επιθυμητής τιμής (πραγματική, y). Η συνάρτηση που υπολογίζει το σφάλμα μεταξύ προβλεπόμενης και πραγματικής τιμής, αποκαλείται **αντικειμενική συνάρτηση** (objective function) ή **συνάρτηση κόστους** ή **σφάλματος** (loss ή error function).
 - Για κάθε παράμετρο υπολογίζεται η μεταβολή της εξόδου όταν αυτή μεταβάλλεται. Με άλλα λόγια, υπολογίζεται η μερική παράγωγος της συνάρτησης κόστους ως προς την κάθε παράμετρο. Η κάθε παράμετρος προσαρμόζεται ώστε να μειώνεται η τιμή της συνάρτησης κόστους.

- Επανάληψη της διαδικασίας, με σκοπό τη μείωση του σφάλματος μεταξύ προβλεπόμενης και επιθυμητής εξόδου, επιτυγχάνοντας κατά συνέπεια την καλύτερη προσαρμογή του στη συνάρτηση που καλείται να προσεγγίσει.

Ο αλγόριθμος Οπισθοδιάδοσης που περιγράφεται παραπάνω είναι ο λεγόμενος αλγόριθμος **Gradient Descent** [44, 45], ο οποίος διορθώνει κάθε μια από τις παραμέτρους του αλγορίθμου λαμβάνοντας υπόψιν του όλα τα παρεχόμενα δεδομένα εκπαίδευσης. Η διαδικασία αυτή είναι χρονοβόρα και υπολογιστικά απαιτητική, ειδικά στις περιπτώσεις που γίνεται χρήση μεγάλου όγκου δεδομένων εκπαίδευσης, όπως κατά την εκπαίδευση νευρωνικών δικτύων βαθιάς μάθησης. Για αυτό το λόγο, σε εφαρμογές που διαχειρίζονται μεγάλο όγκο δεδομένων χρησιμοποιείται μια παραλλαγή του, η οποία χρησιμοποιεί ένα υποσύνολο των δεδομένων εκπαίδευσης και ονομάζεται **Minibatch Stochastic Gradient Descent** [45, 46]. Η διόρθωση των παραμέτρων με ένα μέρος των δεδομένων εκπαίδευσης έχει ως αποτέλεσμα την ταχύτερη εκπαίδευση του αλγορίθμου, συγκριτικά όμως με τον Gradient Descent, δεν επιτυγχάνεται η ίδια μείωση του σφάλματος της εξόδου του.

2 Riemannian Γεωμετρία

Το αντικείμενο αυτής της διπλωματικής εργασίας, βασίζεται από μαθηματικής φύσεως, σε ένα μεγάλο βαθμό, στις έννοιες των Πολλαπλοτήτων (Manifolds) και κυρίως στις Πολλαπλότητες Συμμετρικών και Θετικά Ορισμένων Πινάκων (Manifolds of Symmetric Positive Definite Matrices or SPD Manifolds). Για να αναφερθούμε όμως στις Πολλαπλότητες θα χρειαστεί να εισάγουμε κάποιες βασικές έννοιες από την Τοπολογία, καθώς οι Πολλαπλότητες αποτελούν τοπολογικούς χώρους με κάποιες ιδιαιτερότητες που θα αναφερθούν στη συνέχεια.

2.1 Τοπολογικοί Χώροι

Ένα σύνολο (set) X , είναι μια συλλογή διακριτών αντικειμένων ή στοιχείων. Τα αντικείμενα αυτά μπορεί να είναι οποιαδήποτε μαθηματική οντότητα όπως: αριθμοί (τα σύνολα των πραγματικών \mathbb{R} , των μιγαδικών αριθμών \mathbb{C} , κ.ά.), σημεία, γραμμές στο επίπεδο (σύνολο σημείων [47]), πίνακες, γεωμετρικά σχήματα ή και άλλα σύνολα.

Ένα σύνολο από μόνο του δεν έχει δομή, και ένας τρόπος να λάβει είναι μέσω του συνδυασμού του με ένα άλλο σύνολο, το οποίο ονομάζεται **τοπολογία**. Η τοπολογία σε ένα σύνολο παρέχει τρεις σημαντικές δυνατότητες [48]: αυτή της σύγκλισης, των συνεχών απεικονίσεων και η έννοια των γειτονιών. Η έννοια της σύγκλισης είναι η προσέγγιση μιας ακολουθίας όλο και περισσότερο σε μια τελική τιμή, όταν ο δείκτης της τείνει στο άπειρο. Παρ' όλα αυτά, για τους ορισμούς που θα ακολουθήσουν οι τελευταίες δύο ιδιότητες είναι οι σημαντικότερες.

Οι λεγόμενες **γειτονιές** είναι υποσύνολα του X και φανερώνουν το πως συσχετίζονται μεταξύ τους τα στοιχεία του συνόλου. Τα υποσύνολα αυτά ονομάζονται **ανοικτά υποσύνολα**.

Ένα σύνολο X σε συνδυασμό με μία τοπολογία \mathcal{T} συνθέτουν ένα **Τοπολογικό Χώρο**, ο οποίος συμβολίζεται ως (X, \mathcal{T}) . Το σχήμα του χώρου αυτού διατηρείται και δεν αλλοιώνεται από ομοπαράλληλες (affine) παραμορφώσεις (τεντώματα, περιστροφές, κ.ά.). Εάν αυτός ο χώρος υποστεί τομές ή ενώσεις, τότε η υπάρχουσα τοπολογία του δεν ισχύει, αφού έχουν αλλάξει τα ανοικτά υποσύνολα τα οποία περιέχονται στο σύνολο της τοπολογίας.

Μια τοπολογία \mathcal{T} ενός συνόλου X είναι μια συλλογή ανοικτών υποσυνόλων (γειτονιών) του X και για αυτή ισχύει ότι:

- i. Το κενό σύνολο \emptyset καθώς και το ίδιο το σύνολο X εμπεριέχονται στην τοπολογία \mathcal{T} , δηλαδή $\emptyset, X \in \mathcal{T}$.
- ii. Η τομή πεπερασμένου πλήθους υποσυνόλων της τοπολογίας \mathcal{T} είναι επίσης υποσύνολο της τοπολογίας, δηλαδή $U_\alpha \in \mathcal{T}$, με $n \in \mathbb{N} \Rightarrow \bigcap_{\alpha=1}^n U_\alpha \in \mathcal{T}$.
- iii. Η ένωση οποιονδήποτε υποσυνόλων της τοπολογίας \mathcal{T} είναι επίσης υποσύνολο της τοπολογίας \mathcal{T} , δηλαδή $U_\alpha \in \mathcal{T} \Rightarrow \bigcup_{\alpha \in A} U_\alpha \in \mathcal{T}$.

Μια **απεικόνιση** (map) $f : X \rightarrow Y$ είναι **συνεχής** (continuous) όταν για κάθε ανοικτό υποσύνολο $U \subseteq Y$ η αντίστροφη εικόνα $f^{-1}(U)$ είναι ανοικτό υποσύνολο του X . Στην περίπτωση που μια συνεχής απεικόνιση συνδέει οποιοδήποτε μοναδικό σημείο $p \in X$ με μόνο ένα άλλο, μοναδικό σημείο $q \in Y$, τότε η απεικόνιση αυτή ονομάζεται “**ένα προς ένα και επί**” (bijective).

Μια ένα προς ένα και επί απεικόνιση $f : X \rightarrow Y$ με αντίστροφη απεικόνιση η οποία είναι συνεχής $f^{-1} : Y \rightarrow X$, χαρακτηρίζεται ως **ομοιομορφισμός** (homeomorphism) και αυτά τα δύο σύνολα ως ομοιομορφικά ($X \cong Y$). Αυτό πρακτικά σημαίνει ότι ακόμα και αν το σχήμα

τους έχει μεταβληθεί έπειτα από γραμμικές παραμορφώσεις, οι δύο αυτοί Τοπολογικοί Χώροι είναι ίδιοι.

2.1.1 Χώροι Hausdorff

Το αντικείμενο αυτής της διπλωματικής εργασίας εστιάζει στις Πολλαπλότητες. Για την τοπολογία των Πολλαπλοτήτων είναι απαραίτητο να ισχύει επιπλέον η ακόλουθη ιδιότητα, καθώς εγγυάται ότι ένας τοπολογικός χώρος έχει αρκετά πολλά ανοικτά υποσύνολα:

- Ένας τοπολογικός χώρος χαρακτηρίζεται ως **Hausdorff** [48], εάν για κάθε δύο διαφορετικά στοιχεία $p, q \in X$ υπάρχουν ανοικτά υποσύνολα $U, V \subseteq X$ για τα οποία ισχύει $p \in U$ και $q \in V$, τα οποία είναι ξένα μεταξύ τους ($U \cap V = \emptyset$).

Πλεονεκτήματα των Hausdorff χώρων που αφορούν την παρούσα διπλωματική εργασία, είναι ότι τα σημεία αυτών είναι διαχωρίσιμα και ότι οι ακολουθίες τους συγκλίνουν σε ένα και μόνο σημείο. Επίσης, ισχύει ότι όλοι οι μετρικοί χώροι, οι χώροι στους οποίους ορίζεται κάποια μορφή απόστασης μεταξύ των στοιχείων τους, είναι Hausdorff.

2.1.2 Βάση Τοπολογικού Χώρου

Σε πιο σύνθετες περιπτώσεις, επειδή οι τοπολογίες των χώρων αυτών είναι περίπλοκες, περιγράφονται από **βάσεις** (bases) [48]. Μια βάση (basis) είναι μια συλλογή ανοικτών υποσυνόλων, που περιλαμβάνει τα απολύτως απαραίτητα ανοικτά υποσύνολα μιας τοπολογίας. Συνάγεται ότι, από τους κατάλληλους συνδυασμούς (ενώσεις) αυτών, μπορεί να προκύψει η τοπολογία ενός τοπολογικού χώρου. Συγκεκριμένα ισχύουν τα ακόλουθα:

- i. Το σύνολο X προκύπτει από την ένωση των ανοικτών υποσυνόλων B της βάσης \mathcal{B} , δηλαδή $X = \bigcup_{B \in \mathcal{B}} B$.
- ii. Αν δύο υποσύνολα της βάσης $B_1, B_2 \in \mathcal{B}$ και $p \in B_1 \cap B_2$, τότε υπάρχει ένα $B_3 \in \mathcal{B}$ για το οποίο ισχύει ότι $p \in B_3 \subseteq B_1 \cap B_2$.

Όπου B τα στοιχεία της βάσης \mathcal{B} , X ο τοπολογικός χώρος του οποίου η τοπολογία προκύπτει μέσω της βάσης και p ένα στοιχείο του X .

2.2 Τοπολογικές Πολλαπλότητες

Οι **Τοπολογικές Πολλαπλότητες** (Topological Manifolds) [48], αποτελούν τοπικά Ευκλείδειους, Hausdorff χώρους. Με τον όρο τοπικά Ευκλείδειοι εννοείται ότι σε τοπικό επίπεδο, οι χώροι αυτοί μοιάζουν με τον γνωστό σε όλους Ευκλείδειο n -διάστατο χώρο, \mathbb{R}^n . Συγκεκριμένα, για τις τοπολογικές πολλαπλότητες ισχύουν τα εξής:

- i. **Είναι τοπικά Ευκλείδειοι χώροι διάστασης n .** Για κάθε σημείο $p \in \mathcal{M}$ ενός ανοικτού υποσυνόλου $U \subseteq \mathcal{M}$ υπάρχει ομοιομορφισμός φ , ο οποίος μεταφέρει το U σε έναν Ευκλείδειο χώρο \mathbb{R}^n :

$$\forall p \in \mathcal{M} : \exists \varphi : U \rightarrow \hat{U} \subseteq \mathbb{R}^n$$

- ii. **Είναι χώροι Hausdorff**, δηλαδή για οποιαδήποτε δύο σημεία $p, q \in \mathcal{M}$, υπάρχουν ξένα ανοικτά υποσύνολα $U, V \subseteq \mathcal{M}$ για τα οποία ισχύει $p \in U$ και $q \in V$.

iii. **Είναι δεύτεροι-αριθμήσιμοι (second-countable)**, δηλαδή έχουν βάση με αριθμήσιμα ανοικτά υποσύνολα.

Μια σημαντική προϋπόθεση για να είναι δύο τοπολογικές πολλαπλότητες ομοιομορφικές, είναι να έχουν την ίδια διάσταση, δηλαδή μια n -διάστατη πολλαπλότητα δεν μπορεί να είναι ομοιομορφική με μια πολλαπλότητα m διαστάσεων ($m \neq n$). Λόγω των παραπάνω ιδιοτήτων, οι τοπολογικές πολλαπλότητες συμπεριφέρονται τοπικά όπως οι Ευκλείδειοι χώροι, κάτι το οποίο βοηθά τη μελέτη τους, αφού έχουμε καλύτερη αντίληψη των Ευκλείδειων χώρων.

Για να εφοδιάσουμε μια Τοπολογική Πολλαπλότητα με παραπάνω δομή, στη συνέχεια θα ορίσουμε την έννοια της Λειότητας (Smoothness), η οποία είναι βασική προϋπόθεση για την ανάπτυξη μαθηματικών πράξεων μεταξύ στοιχείων της. Ο ορισμός της έννοιας της Λειότητας δεν μπορεί να γίνει μέσω ιδιοτήτων της τοπολογίας (συμπάγεια, σύνδεση μεταξύ υποσυνόλων, ομοιομορφισμός, κ.α.), αφού δεν υπάρχει διαχωρισμός μεταξύ Λείων και μη, ομοιομορφικών τοπολογικών πολλαπλοτήτων. Για παράδειγμα, μπορούμε εύκολα να φανταστούμε το πώς από έναν εύπλαστο κύβο μπορεί να προκύψει μια σφαίρα, χωρίς να χρειαστεί να γίνει κάποια τομή ή ένωση, παρά μόνο μέσω γραμμικών παραμορφώσεων. Αφού δεν έχει γίνει κάποια τομή ή ένωση στο κύβο, πλέον σφαίρα, αυτές οι δύο οντότητες, με βάση την τοπολογία, είναι ομοιομορφικές. Από το πώς αντιλαμβανόμαστε την έννοια της λειότητας όμως, θα λέγαμε ότι η σφαίρα έχει λείο σχήμα ενώ ο κύβος όχι.

Αναζητούμε λοιπόν ένα τρόπο να διαφοροποιήσουμε τέτοιους τοπολογικούς χώρους και αυτό επιτυγχάνεται μέσω των παρακάτω εργαλείων της Τοπολογίας. Στη συνέχεια, θα οριστεί η έννοια της Λειότητας, η οποία θα αποτελέσει το δεύτερο δομικό στοιχείο των πολλαπλοτήτων που αφορούν αυτή τη διπλωματική εργασία.

2.2.1 Χάρτες Συντεταγμένων

Οι **χάρτες συντεταγμένων** (coordinate charts) ή απλά **χάρτες** [48], χρησιμοποιούνται για να μεταφέρουν μια τοπικά Ευκλείδεια περιοχή μιας τοπολογικής πολλαπλότητας στον Ευκλείδειο χώρο. Ένας χάρτης αποτελείται από το ζεύγος (U, φ) , όπου U είναι ένα ανοικτό υποσύνολο (γειτονιά) της πολλαπλότητας \mathcal{M} και $\varphi : U \rightarrow \hat{U}$, ένας ομοιομορφισμός από το $U \subseteq \mathcal{M}$, στο ανοικτό υποσύνολο $\hat{U} = \varphi(U) \subseteq \mathbb{R}^n$.

Το U ονομάζεται πεδίο συντεταγμένων ή απλά **πεδίο του χάρτη** (coordinate domain) (U, φ) . Η απεικόνιση φ ονομάζεται **τοπική απεικόνιση συντεταγμένων** (local coordinate map) και αποτελείται από τις **συνιστώσες απεικόνισης** (local coordinates) x^i .

Οι συνιστώσες απεικόνισης φανερώνουν τη θέση του σημείου p στον x^i άξονά. Επομένως, η απεικόνιση ενός n -διάστατου σημείου p γράφεται και ως $\varphi(p) = (x^1(p), x^2(p), \dots, x^n(p))$. Αξίζει να σημειωθεί ότι παρόλο που ο δείκτης i είναι στον εκθέτη του x , το i δεν είναι δύναμη, αλλά δείκτης. Τα x^i γράφονται με αυτό το τρόπο για να είναι συμβατά με τη σημειογραφία άθροισης του Einstein [49], η οποία χρησιμοποιείται κατά κόρον στη μελέτη των Ομαλών Πολλαπλοτήτων.

2.2.2 Συναρτήσεις Μετάβασης

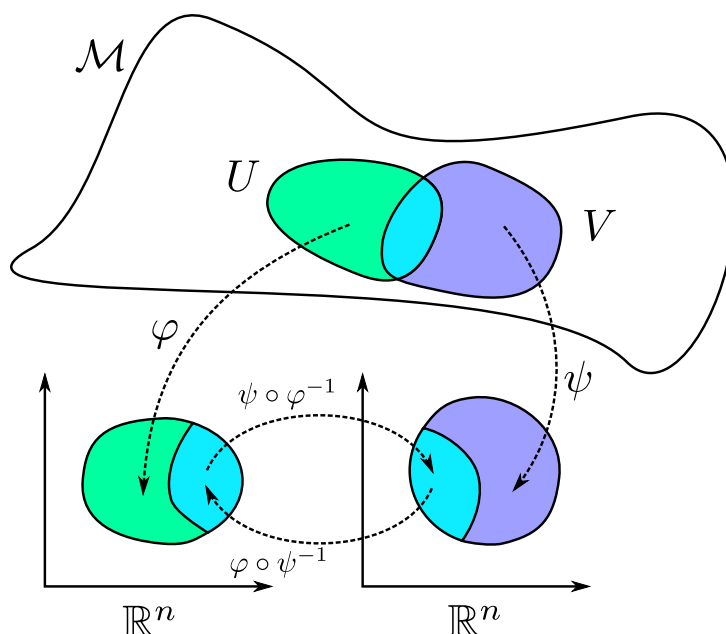
Οι **συναρτήσεις μετάβασης** (transition maps) [48], είναι συναρτήσεις που συνδέουν δύο διαφορετικούς χάρτες συντεταγμένων μεταξύ τους, εφόσον αυτοί καλύπτουν κάποιες κοινές περιοχές. Δηλαδή έστω ότι για τους χάρτες (U, φ) , (V, ψ) ισχύει ότι $U \cap V \neq \emptyset$ και αφού τα φ , ψ είναι ομοιομορφισμοί, άρα έχουν συνεχείς αντίστροφες απεικονίσεις φ^{-1} , ψ^{-1} , τότε μπορεί να μεταβεί κανείς σε περιοχές μεταξύ των δύο αυτών χαρτών, με μια συνάρτηση μετάβασης.

Η μετάβαση από το χάρτη συντεταγμένων (U, φ) στον χάρτη (V, ψ) , θα γίνει μέσω της συνάρτησης μετάβασης $\psi \circ \varphi^{-1} : \varphi(U \cap V) \rightarrow \psi(U \cap V)$. Η μετάβαση θα γίνει δηλαδή από το χάρτη (U, φ) , του οποίου η εικόνα είναι στον Ευκλείδειο χώρο, μέσω της αντίστροφης απεικόνισης του φ^{-1} , πίσω στην τοπολογική πολλαπλότητα και έπειτα μέσω της απεικόνισης ψ στο χάρτη (V, ψ) . Για τις συναρτήσεις μετάβασης ισχύουν οι παρακάτω ιδιότητες:

- Εφόσον η **συνάρτηση μετάβασης** αποτελεί τη σύνθεση δύο ομοιομορφισμών, ισχύει ότι και αυτή είναι **ομοιομορφισμός**.
- Εφόσον είναι ομοιομορφισμός, τότε είναι και **συνεχής, όπως και η αντίστροφη απεικόνισή της**. Οι συνεχείς συναρτήσεις ονομάζονται και C^0 -Συμβατές (C^0 -Compatible), για το λόγο του ότι οι μηδενικές τους παράγωγοι είναι συνεχείς.
- Εφόσον είναι C^0 -Συμβατές, τότε και οι χάρτες (U, φ) και (V, ψ) είναι επίσης C^0 -Συμβατοί.

Γενικότερα, δύο χάρτες είναι C^k -Συμβατοί αν οι συναρτήσεις μετάβασής τους είναι C^k -Συμβατές, το οποίο σημαίνει ότι αυτές είναι k φορές παραγωγίσιμες και συνεχείς. Δύο χάρτες (U, φ) και (V, ψ) είναι C^∞ -**Συμβατοί** ή Συμβατοί κατά Λειότητα (smoothly compatible) αν για αυτούς ισχύει ότι $U \cap V = \emptyset$ ή αν συνάρτηση μετάβασής τους $\psi \circ \varphi^{-1}$ είναι **αμφιδιαφόριση** (diffeomorphism), δηλαδή η ίδια και η αντίστροφη συνάρτηση μετάβασης $\varphi \circ \psi^{-1} = (\psi \circ \varphi^{-1})^{-1}$ είναι παραγωγίσιμες.

Στο πλαίσιο του ορισμού της έννοιας της Λείας Δομής, υπάρχει η απαίτηση οι χάρτες που χρησιμοποιούνται να είναι C^∞ -Συμβατοί. Εδώ, αυτό που αντιλαμβανόμαστε, είναι ότι αφού οι χάρτες αυτοί είναι C^∞ -Συμβατοί και υποσύνολα του \mathbb{R}^n , τότε όλες οι τάξεις παραγώγων τους είναι συνεχείς.



Εικόνα 5: Δύο χάρτες συντεταγμένων και οι συναρτήσεις μετάβασής μεταξύ περιοχών τους¹³.

¹³Atlas (topology)

2.2.3 Άτλαντες

Οι **Άτλαντες** (atlases) [48] είναι μια συλλογή χαρτών συντεταγμένων, οι οποίοι περιγράφουν μεμονωμένες περιοχές μιας πολλαπλότητας. Ένας άτλαντας $\mathcal{A} = \{(U_i, \varphi_i) : i \in I\}$ μπορεί να καλύψει (να χαρτογραφήσει) ολόκληρη μια πολλαπλότητα, αν τα πεδία των χαρτών του U , συνθέτουν το σύνολο της πολλαπλότητας, $\bigcup_{i \in I} U_i = M$.

Ένας άτλαντας ονομάζεται **Λείος Άτλαντας** αν οποιοδήποτε δύο χάρτες συντεταγμένων του συνόλου του, είναι C^∞ -Συμβατοί μεταξύ τους.

2.2.4 Σύνολο Πηλίκο

Το **Σύνολο Πηλίκο** (Quotient Space) [48, 50], ενός τοπολογικού χώρου X , είναι ένας νέος τοπολογικός χώρος, ο οποίος αποτελείται από το σύνολο των κλάσεων ισοδυναμίας, συμβολιζόμενες ως $[x]$, του αρχικού συνόλου και συμβολίζεται ως $X/\sim := \{[x] : x \in X\}$. Μια **σχέση ισοδυναμίας** (equivalence relation), συμβολιζόμενη ως \sim , είναι μια σχέση που καθορίζει ποια στοιχεία ενός συνόλου, είναι ισοδύναμα μεταξύ τους και διαχωρίζει το αρχικό σύνολο, σε **κλάσεις ισοδυναμίας** (equivalence class), οι οποίες είναι ξένα υποσύνολα και περιέχουν στοιχεία ισοδύναμα μεταξύ τους. Μια σχέση είναι σχέση ισοδυναμίας, όταν ισχύουν οι παρακάτω ιδιότητες:

- Ανακλαστική (Reflexive): $x \sim x, \forall x \in X$.
- Συμμετρική (Symmetric): Για $x, y \in X$, αν $x \sim y$ τότε $y \sim x$.
- Μεταβατική (Transitive): Για $x, y, z \in X$, αν $x \sim y$ και $y \sim z$ τότε $x \sim z$.

Μια κλάση ισοδυναμίας ενός στοιχείου x , συμβολίζεται με αγκύλες, ως $[x]$ και περιέχει τα στοιχεία του αρχικού συνόλου $y \in X$, τα οποία είναι ισοδύναμα με το στοιχείο x , δηλαδή: $[x] := \{y \in X : x \sim y\}$. Από τις ιδιότητες των σχέσεων ισοδυναμίας, συμπεραίνεται επίσης ότι δύο κλάσεις ισοδυναμίας, δεν μπορούν να περιέχουν κοινά στοιχεία, αφού κάθε ισοδύναμο στοιχείο με ένα άλλο, ανήκει στην ίδια κλάση. Με άλλα λόγια, αν περιείχαν κοινά στοιχεία, δε θα ήταν δύο διαφορετικές κλάσεις, αλλά μία.

2.2.5 Λεία Δομή

Ένας τρόπος να ορίσουμε την έννοια της Λείας Δομής, είναι μέσω των C^∞ και **Ολοκληρωμένων Άτλάντων** (Smooth and Maximal ή Complete Atlases) [48, 51]. Οι C^∞ και ολοκληρωμένοι άτλαντες, δεν εμπεριέχονται σε άλλους C^∞ και οι χάρτες συντεταγμένων αυτών, είναι C^∞ -Συμβατοί αποκλειστικά και μόνο, με τους χάρτες του ίδιου του Άτλαντα. Γενικότερα, ένας C^k άτλαντας, είναι ολοκληρωμένος (maximal ή complete), όταν αυτός δεν εμπεριέχεται σε άλλους άτλαντες κλάσης k .

Έχοντας ορίσει την έννοια της λειότητας, στη συνέχεια θα γίνει αναφορά στην επόμενη έννοια που πρέπει να οριστεί, αυτή των λείων πολλαπλοτήτων. Η αναφορά στις λείες πολλαπλότητες, είναι απαραίτητη, αφού οι Riemannian Πολλαπλότητες, αποτελούν λείες πολλαπλότητες, με κάποια Riemannian μετρική, για την οποία θα αναφερθούμε σε επόμενο κεφάλαιο.

2.3 Λείες Πολλαπλότητες

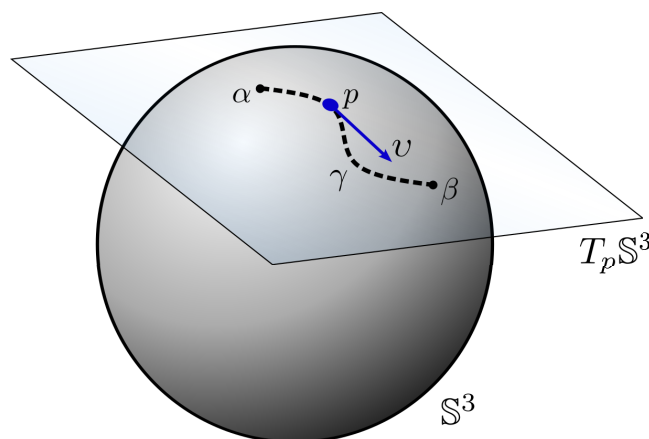
Οι **Λείες ή Παραγωγίσιμες Πολλαπλότητες** (Smooth or Differentiable Manifolds) [48], είναι Τοπολογικές Πολλαπλότητες με λεία δομή και συμβολίζονται ως $(\mathcal{M}, \mathcal{A})$. Λόγω αυτών των δύο ιδιοτήτων τους, στις λείες πολλαπλότητες μπορούν και ορίζονται μαθηματικές πράξεις. Παραδείγματα λείων πολλαπλοτήτων αποτελούν οι Ευκλείδειοι χώροι, τα γεωμετρικά σχήματα όπως αυτά των κύκλων, σφαιρών, ελλειψοειδών και άλλων, καθώς και οι Grassmannian, οι Stiefel και οι Riemannian χώροι.

2.3.1 Εφαπτόμενος Χώρος

Ο **Εφαπτόμενος Χώρος** (Tangent Space) [48, 52, 53], για ένα σημείο αποτελεί το σύνολο των εφαπτόμενων διανυσμάτων που διέρχονται από το σημείο αυτό, στην επιφάνεια μιας λείας πολλαπλότητας. Ο εφαπτόμενος χώρος φανερώνει το πως είναι η επιφάνεια μιας πολλαπλότητας, γύρω από ένα σημείο. Μέσω αυτού ορίζονται οι μερικές παράγωγοι των καμπυλών της επιφάνειας μιας πολλαπλότητας, και βασιζόμενοι σε αυτόν και άλλες πιο απλές μαθηματικές πράξεις. Σημειώνεται επίσης ότι για την παρούσα διπλωματική εργασία, τα διανύσματα καθώς και οι πίνακες συμβολίζονται με έντονη γραφή, για να ξεχωρίζουν από άλλα μεγέθη.

Οι παράγωγοι αυτές έχουν μέτρο και κατεύθυνση, το οποίο τις καθιστά διανύσματα. Τα διανύσματα αυτά ονομάζονται **εφαπτόμενα διανύσματα** (tangent vectors), συμβολίζονται με \mathbf{v} και ορίζονται ως οι παράγωγοι των καμπυλών γ που διέρχονται από ένα σημείο $p \in \mathcal{M}$, της επιφάνειας μιας πολλαπλότητας. Ο εφαπτόμενος χώρος ή εφαπτόμενο επίπεδο (tangent plane), όπως συναντάται αλλιώς, αποτελεί το σύνολο των εφαπτόμενων διανυσμάτων \mathbf{v} , για το σημείο p και συμβολίζεται ως $T_p\mathcal{M}$.

Για τον ορισμό του εφαπτόμενου διανύσματος θα χρειαστεί να εισάγουμε τον ορισμό της λείας καμπύλης μιας πολλαπλότητας. Μια καμπύλη γ μιας πολλαπλότητας είναι η απεικόνιση ενός διαστήματος πραγματικών αριθμών $J = (\alpha, \beta)$, σε μια πολλαπλότητα \mathcal{M} . Γενικότερα, μια απεικόνιση μεταξύ δύο λείων πολλαπλοτήτων, $f : \mathcal{M} \rightarrow \mathcal{N}$ είναι **λεία**, όταν για τους χάρτες των \mathcal{M} και \mathcal{N} , (U, φ) και (V, ψ) αντίστοιχα, με $f(U) \cap V \neq \emptyset$, η συνάρτηση $\psi \circ f \circ \varphi^{-1} : \mathbb{R}^m \rightarrow \mathbb{R}^n$, είναι λεία. Όπου m και n , οι διαστάσεις των πολλαπλοτήτων \mathcal{M} και \mathcal{N} αντίστοιχα. Μια **λεία απεικόνιση** ή **C^∞ -Απεικόνιση**, είναι άπειρες φορές παραγωγίσιμη. Στη συγκεκριμένη περίπτωση, η **λεία καμπύλη** γ , είναι μια λεία απεικόνιση μεταξύ του J , το οποίο είναι μια λεία πολλαπλότητα, με διάσταση $\dim(J) = 1$, και της \mathcal{M} .



Εικόνα 6: Το Εφαπτόμενο Επίπεδο $T_p\mathbb{S}^3$, που ορίζεται στο σημείο p της σφαίρας \mathbb{S}^3 ¹⁴.

¹⁴Tangent space

Για μια λεία απεικόνιση f και μια λεία καμπύλη $\gamma : (\alpha, \beta) \rightarrow \mathcal{M}$, με $\gamma(t_0) = p$, για κάποιο $t_0 \in (\alpha, \beta)$, το εφαπτόμενο διάνυσμα $\mathbf{v} = \gamma'(t_0)(f)$ ορίζεται ως η κατευθυνόμενη παράγωγος της f στο σημείο p κατά την καμπύλη γ , δηλαδή:

$$\mathbf{v} = \gamma'(t_0)(f) = \lim_{t \rightarrow 0} \frac{f(\gamma(t_0 + t)) - f(\gamma(t_0))}{t} = (f \circ \gamma)'(t_0) = \left. \frac{d}{dt} \right|_{t=t_0} (f \circ \gamma)$$

Ο Εφαπτόμενος Χώρος $T_p\mathcal{M}$ που προκύπτει από το σύνολο των εφαπτόμενων διανυσμάτων ενός σημείου $p \in \mathcal{M}$, της επιφάνειας μιας n -διάστατης λείας πολλαπλότητας \mathcal{M} , αποτελεί ένα Διανυσματικό Χώρο n διαστάσεων. Στους διανυσματικούς χώρους, ορίζονται οι μαθηματικές πράξεις της πρόσθεσης μεταξύ διανυσμάτων και του πολλαπλασιασμού μεταξύ διανυσμάτων και αριθμών (βαθμωτών μεγεθών), καθώς και η προσεταιριστική και αντιμεταθετική ιδιότητα, της πρόσθεσης και η επιμεριστική του πολλαπλασιασμού.

2.3.2 Γεωδεσιακή

Η **γεωδεσιακή** (geodesic) [48], είναι η καμπύλη $\gamma : J \rightarrow \mathcal{M}$ με $J = (\alpha, \beta)$ ελάχιστου μήκους, που ενώνει δύο σημεία μιας πολλαπλότητας $p, q \in \mathcal{M}$, με $\gamma(\alpha) = p$ και $\gamma(\beta) = q$. Μια γεωδεσιακή δηλαδή, αποτελεί το συντομότερο μονοπάτι μεταξύ δύο σημείων της πολλαπλότητας. Για τις γεωδεσιακές ισχύει ότι η ταχύτητά τους, δηλαδή η μεταβολή του εφαπτόμενου διανύσματος πάνω σε αυτές, είναι σταθερή $\gamma''(t) = \mathbf{v}' = 0$, για κάθε $t \in (\alpha, \beta)$. Από το παραπάνω συμπεραίνεται ότι οι γεωδεσιακές αποτελούν τη γενίκευση της έννοιας του ευθύγραμμου τμήματος στις πολλαπλότητες. Επιπλέον, μια γεωδεσιακή χαρακτηρίζεται ως **μεγιστική** (maximal), όταν το διάστημα J στο οποίο ορίζεται είναι το μεγαλύτερο, το μέγιστο σε σύγκριση με άλλα, διάστημα στο οποίο μπορεί να οριστεί. Οι μεγιστικές γεωδεσιακές συμβολίζονται ως $\gamma_{\mathbf{v}}$.

2.4 Εφαπτόμενη Δέσμη

Η **Εφαπτόμενη Δέσμη** (Tangent Bundle) [48, 54], συμβολιζόμενη ως $T\mathcal{M}$ μιας πολλαπλότητας \mathcal{M} , αποτελεί το σύνολο των εφαπτόμενων χώρων, για κάθε στοιχείο αυτής της πολλαπλότητας. Συγκεκριμένα, ορίζεται ως η ξένη ένωση (disjoint union), των εφαπτόμενων χώρων $T_p\mathcal{M}$ για κάθε σημείο $p \in \mathcal{M}$ της πολλαπλότητας. Ισοδύναμα, αποτελεί το σύνολο όλων των εφαπτόμενων διανυσμάτων $\mathbf{v} \in T_p\mathcal{M}$, κάθε σημείου p μιας πολλαπλότητας:

$$T\mathcal{M} = \coprod_{p \in \mathcal{M}} T_p\mathcal{M} = \bigcup_{p \in \mathcal{M}} \{(p, \mathbf{v}) : \mathbf{v} \in T_p\mathcal{M}\}$$

Η χρήση της ξένης ένωσης [48, 55, 56] στην παραπάνω σχέση, εγγυάται ότι στο σύνολο $T\mathcal{M}$ θα εμπεριέχονται εφαπτόμενοι χώροι ξένοι μεταξύ τους, οι οποίοι δηλαδή, δε θα εμπεριέχουν κοινά στοιχεία. Παρόλο που μπορεί κάποιος εφαπτόμενοι χώροι να περιέχουν ένα ή περισσότερα εφαπτόμενα διανύσματα ίσα μεταξύ τους, για παράδειγμα $\mathbf{v} \in T_p\mathcal{M}$ και $\mathbf{u} \in T_q\mathcal{M}$ με $\mathbf{v} = \mathbf{u}$, αυτό που διαφοροποιεί τα δύο αυτά εφαπτόμενα διανύσματα, θα είναι το σημείο της πολλαπλότητας από το οποίο προήλθαν. Για να γίνεται αντιληπτό αυτό, τα στοιχεία της Εφαπτόμενης Δέσμης γράφονται ως ο συνδυασμός ενός σημείου της πολλαπλότητας με το εφαπτόμενο διάνυσμά του. Για το παραπάνω παράδειγμα συνεπώς, τα στοιχεία της Εφαπτόμενης Δέσμης θα γράφονταν ως: $T\mathcal{M} = \{\dots, (p, \mathbf{v}), (q, \mathbf{u}), \dots\} \stackrel{\mathbf{v}=\mathbf{u}}{=} \{\dots, (p, \mathbf{v}), (q, \mathbf{v}), \dots\}$.

2.5 Εκθετική Απεικόνιση

Η **Εκθετική Απεικόνιση** (Exponential Map) [48], είναι η απεικόνιση ενός διανύσματος της Εφαπτόμενης Δέσμης $T\mathcal{M}$ σε ένα σετ μεγιστικών γεωδειακών της πολλαπλότητας \mathcal{M} . Συγκεκριμένα, ορίζεται ως ο χάρτης από ένα το πεδίο (domain), που αποτελεί υποσύνολο της Εφαπτόμενης Δέσμης $\mathcal{E} \subseteq T\mathcal{M}$ προς την πολλαπλότητα \mathcal{M} . Το πεδίο αυτό ορίζεται ως:

$$\mathcal{E} = \{\mathbf{v} \in T\mathcal{M} : \gamma_{\mathbf{v}} : [\alpha, \beta] \rightarrow \mathcal{M}\}$$

Τότε, η Εκθετική Απεικόνιση $\exp : \mathcal{E} \rightarrow \mathcal{M}$ ορίζεται ως:

$$\exp(\mathbf{v}) = \gamma_{\mathbf{v}}(\beta)$$

Η Εκθετική Απεικόνιση ενός διανύσματος \mathbf{v} για ένα σημείο $p \in \mathcal{M}$, μιας μεγιστικής γεωδειακής, συμβολίζεται ως $\exp_p : \mathcal{E}_p \rightarrow \mathcal{M}$, με πεδίο ορισμού $\mathcal{E}_p = \mathcal{E} \cap T_p\mathcal{M}$. Μέσω αυτής προσδιορίζεται το σημείο που βρίσκεται στο τέλος της μεγιστικής γεωδειακής ($\gamma_{\mathbf{v}}(\beta)$), μέσω ενός εφαπτόμενου διανύσματος και ορίζεται ως:

$$\exp(p, \mathbf{v}) = \exp_p(\mathbf{v}) = \gamma_{\mathbf{v}}(\beta)$$

2.6 Riemannian Πολλαπλότητες

Οι **Riemannian Πολλαπλότητες** (Riemannian Manifolds) [48], αποτελούνται από μια Λεία Πολλαπλότητα \mathcal{M} και μια Riemannian μετρική g και συμβολίζονται ως (\mathcal{M}, g) . Η **μετρική** (metric) g είναι ένα θετικά ορισμένο εσωτερικό γινόμενο, δύο εφαπτόμενων διανυσμάτων, του εφαπτόμενου χώρου της πολλαπλότητας $T_p\mathcal{M}$, για κάθε σημείο p αυτής της Riemannian πολλαπλότητας \mathcal{M} . Το εσωτερικό γινόμενο μεταξύ δύο εφαπτόμενων διανυσμάτων $\mathbf{v}, \mathbf{w} \in T_p\mathcal{M}$, εκφράζεται ως $\langle \mathbf{v}, \mathbf{w} \rangle_g = g_p(\mathbf{v}, \mathbf{w})$ και για αυτό ισχύουν οι παρακάτω ιδιότητες:

- Συμμετρικό (Symmetric): $g_p(\mathbf{v}, \mathbf{w}) = g_p(\mathbf{w}, \mathbf{v})$.
- Γραμμικό (Linear): $g_p(\alpha\mathbf{v} + \beta\mathbf{w}, \mathbf{u}) = \alpha g_p(\mathbf{v}, \mathbf{u}) + \beta g_p(\mathbf{w}, \mathbf{u})$, με $\mathbf{u} \in T_p\mathcal{M}$.
- Θετικά Ορισμένο (Positive Definite): $g_p(\mathbf{v}, \mathbf{w}) \geq 0$. Αν $\mathbf{v} = \mathbf{w}$, τότε και μόνο τότε $g_p(\mathbf{v}, \mathbf{w}) = 0$.

Το εσωτερικό γινόμενο επιτρέπει τον ορισμό μηκών και γωνιών διανυσμάτων του εφαπτόμενου χώρου, καθώς και αποστάσεων μεταξύ σημείων. Οι παραπάνω ιδιότητες καθιστούν εφικτή την εφαρμογή της Ευκλείδειας γεωμετρίας στους τοπικά ευκλείδειους χώρους των Λείων Πολλαπλοτήτων. Επιπλέον, όπως και για της Τοπολογικές Πολλαπλότητες, τα υποσύνολα μιας Riemannian πολλαπλότητας, είναι και αυτά Riemannian πολλαπλότητες και υιοθετούν τη μετρική αυτής.

Οι Riemannian Πολλαπλότητες αποτελούν το είδος πολλαπλοτήτων στο οποίο ανήκουν στοιχεία αυτής της διπλωματικής εργασίας. Όπως θα αναφερθεί σε επόμενες ενότητες, ο χώρος των Συμμετρικών και Θετικά Ορισμένων Πολλαπλοτήτων αποτελεί το πεδίο στο οποίο ανήκουν τα δεδομένα εισόδου, καθώς και άλλα καίρια δομικά στοιχεία της μεθόδου που αναλύεται.

2.6.1 Πολλαπλότητες των Συμμετρικών και Θετικά Ορισμένων Πινάκων

Οι **Πολλαπλότητες των Συμμετρικών και Θετικά Ορισμένων Πινάκων** (Symmetric Positive Definite, SPD Manifolds) [57], είναι Riemannian Πολλαπλότητες και αποτελούν το σύνολο των συμμετρικών και θετικά ορισμένων πινάκων πραγματικών αριθμών, διαστάσεων $n \times n$. Ένας συμμετρικός πίνακας $\mathbf{X} \in \mathbb{R}^{n \times n}$ είναι θετικά ορισμένος όταν οι ιδιοτιμές του είναι θετικές ($\lambda > 0$), ή ισοδύναμα, ορίζεται ως ο πίνακας \mathbf{X} για τον οποίον ισχύει ότι $\mathbf{v}^T \mathbf{X} \mathbf{v} > 0$, για κάθε μη μηδενικό διάνυσμα ($\mathbf{v} \neq 0$).

$$Sym_n^+ = \{\mathbf{X} \in \mathbb{R}^{n \times n} : \mathbf{v}^T \mathbf{X} \mathbf{v} > 0 \quad \forall \mathbf{v} \in \mathbb{R}^n \setminus \{0\}\}$$

Ένας Συμμετρικός και Θετικά Ορισμένος πίνακας αποτελεί ένα σημείο αυτού χώρου. Το εσωτερικό γινόμενο ενός σημείου μιας πολλαπλότητας, ορίζεται στον εφαπτόμενο χώρο της και για να γίνει χρήση του, καθώς και άλλων αλγεβρικών πράξεων, είναι αναγκαία η μεταφορά του σημείου μεταξύ της πολλαπλότητας και του εφαπτόμενου χώρου. Για τη μεταφορά από τον χώρο των Πολλαπλοτήτων των Συμμετρικών και Θετικά Ορισμένων πινάκων, προς τον Εφαπτόμενο Χώρο και πίσω, χρησιμοποιούνται οι συναρτήσεις του φυσικού λογαρίθμου και η εκθετική αντίστοιχα.

2.6.2 Stiefel Πολλαπλότητες

Οι **Stiefel πολλαπλότητες** (Stiefel Manifolds) [58, 59], είναι Riemannian πολλαπλότητες και αποτελούν το σύνολο των $n \times p$ πινάκων πραγματικών αριθμών, με p ορθοκανονικές στήλες και n γραμμές, με $p \leq n$. Δύο διανύσματα είναι ορθοκανονικά, και κατ' επέκταση γραμμικά ανεξάρτητα, όταν το εσωτερικό τους γινόμενο ισούται με το μηδέν (ορθογώνια) και το μέτρο του κάθε ενός είναι ίσο με τη μονάδα. Σε μορφή πινάκων, το παραπάνω ικανοποιείται από τη σχέση $\mathbf{X}^T \mathbf{X} = \mathbf{I}_p$, όπου \mathbf{X} ο $n \times p$ πίνακας υπό μελέτη και \mathbf{I}_p , ο $p \times p$ μοναδιαίος πίνακας.

$$St(n, p) = \{\mathbf{X} \in \mathbb{R}^{n \times p} : \mathbf{X}^T \mathbf{X} = \mathbf{I}_p\}$$

2.6.3 Grassmannian Πολλαπλότητες

Οι **Grassmannian Πολλαπλότητες** (Grassmannian Manifolds) [60, 61], είναι Riemannian πολλαπλότητες και αποτελούν επίσης το πηλίκο των Stiefel Πολλαπλοτήτων ως προς το σετ των ορθογώνιων πινάκων $p \times p$. Το σύνολο των Grassmannian πολλαπλοτήτων αποτελείται από υποσύνολα p διαστάσεων ενός n -διάστατου συνόλου των πραγματικών αριθμών \mathbb{R}^n .

$$Gr(n, p) = \{\mathcal{U} \subseteq \mathbb{R}^n : \dim(\mathcal{U}) = p\}$$

Όπως αναφέρθηκε, ορίζεται επίσης ως το πηλίκο των Stiefel πολλαπλοτήτων $St(n, p)$ ως προς το σετ των ορθογώνιων πινάκων $O(p)$, διαστάσεων $p \times p$:

$$Gr(n, p) = St(n, p)/O(p)$$

2.7 Χρήσιμες Μαθηματικές Πράξεις Πινάκων

2.7.1 Λογαριθμική Συνάρτηση

Η **λογαριθμική συνάρτηση** (logarithmic function) [62, 63], συμβολιζόμενη ως $\log_b(x)$, είναι μια συνάρτηση με δύο ορίσματα, τη βάση b , έναν θετικό πραγματικό αριθμό, διάφορο της μονάδας και τον αριθμό x . Η λογαριθμική συνάρτηση, προσδιορίζει τη δύναμη στην οποία πρέπει να υψωθεί η βάση b , ώστε να έχει αποτέλεσμα τον αριθμό x . Ο **φυσικός λογάριθμος** (natural logarithm), ο λογάριθμος με βάση τον αριθμό του Euler ($e \approx 2.71828$), αποτελεί την αντίστροφη συνάρτηση της εκθετικής συνάρτησης e^x και συμβολίζεται ως $\ln(x)$, $\log_e(x)$ ή $\log(x)$. Για τις εφαρμογές της διπλωματικής αυτής εργασίας, θα γίνεται χρήση μόνο του φυσικού λογαρίθμου, για τους πραγματικούς αριθμούς ή τετραγωνικούς πίνακες, καθώς θα χρησιμοποιείται, όπως θα δούμε στη συνέχεια, μαζί με την εκθετική συνάρτηση. Η συνάρτηση του φυσικού λογαρίθμου ορίζεται ως το παρακάτω ολοκλήρωμα, αλλά για να γίνει μετέπειτα ο ορισμός της στην περίπτωση που το όρισμά της x είναι πίνακας, θα γραφτεί και στη μορφή δυναμοσειράς:

$$\begin{aligned}\ln(x) &= \int_1^x \frac{1}{t} dt = \int_0^{x-1} \frac{1}{1+u} du = \\ &= \int_0^{x-1} (1 - u + u^2 - u^3 + \dots) du = \\ &= (x-1) - \frac{(x-1)^2}{2} + \frac{(x-1)^3}{3} - \frac{(x-1)^4}{4} + \dots = \\ &= \sum_{k=1}^{\infty} (-1)^{k-1} \frac{(x-1)^k}{k}\end{aligned}$$

Στην περίπτωση που το όρισμα της συνάρτησης του φυσικού λογαρίθμου είναι τετραγωνικός πίνακας \mathbf{X} πραγματικών αριθμών, διαστάσεων $n \times n$, ο λογάριθμος ενός πίνακα ($\log(\mathbf{X})$) ορίζεται όπως και παραπάνω, με την αντικατάσταση της μονάδας του αριθμητή με το μοναδιαίο πίνακα \mathbf{I} , διαστάσεων $n \times n$. Επιπλέον, ορίζεται και ως η εφαρμογή της παραπάνω δυναμοσειράς στις ιδιοτιμές λ_i του πίνακα \mathbf{X} , εφόσον αναλυθεί στο διαγώνιο πίνακα ιδιοτιμών ($\mathbf{\Lambda}$) και το πίνακα ιδιοδιανυσμάτων (\mathbf{U}) του:

$$\log(\mathbf{X}) = \sum_{k=1}^{\infty} (-1)^{k-1} \frac{(\mathbf{X} - \mathbf{I})^k}{k} = (\mathbf{X} - \mathbf{I}) - \frac{(\mathbf{X} - \mathbf{I})^2}{2} + \frac{(\mathbf{X} - \mathbf{I})^3}{3} + \dots$$

$$\mathbf{X} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T, \quad \log(\mathbf{X}) = \mathbf{U}\log(\mathbf{\Lambda})\mathbf{U}^T = \mathbf{U}\text{diag}(\ln(\{\lambda_i\}_{i=1}^n))\mathbf{U}^T$$

Η εφαρμογή της συνάρτησης του φυσικού λογαρίθμου σε έναν θετικό αριθμό μεγαλύτερο της μονάδας, θα επιφέρει θετικό αποτέλεσμα. Η εφαρμογή της όμως σε αριθμό μεταξύ του μηδενός και της μονάδας θα έχει ως αποτέλεσμα αρνητική τιμή. Όσο πιο πολύ προσεγγίζει αυτός ο αριθμός το μηδέν, τόσο μεγάλο το αρνητικό αποτέλεσμα. Αυτό για τους τετραγωνικούς πίνακες σημαίνει ότι, παρόλο που ένας πίνακας μπορεί να είναι θετικά ορισμένος, δηλαδή με ιδιοτιμές μεγαλύτερες του μηδενός ($\lambda > 0$), έπειτα από την εφαρμογή του φυσικού λογαρίθμου, αυτός, ανάλογα με τις αρχικές τιμές των ιδιοτιμών του, μπορεί να πάψει να είναι θετικά ορισμένος ($\lambda \in \mathbb{R}$). Όπως θα δούμε αργότερα, η χρήση της συνάρτησης του φυσικού λογαρίθμου καθώς και της αντίστροφης συνάρτησης αυτού, η εκθετική συνάρτηση, θα χρησιμοποιηθούν με άλλο σκοπό, παρόλα αυτά όμως το παραπάνω έχει και πάλι σημασία.

2.7.2 Εκθετική Συνάρτηση

Η **εκθετική συνάρτηση** (exponential function) [64, 65], συμβολιζόμενη ως e^x ή $\exp(x)$, είναι μια συνάρτηση με ένα όρισμα, το x , η οποία υψώνει τον αριθμό του Euler ($e \approx 2.71828$) στο x . Το όρισμά της, μπορεί να είναι πραγματικός ή μιγαδικός αριθμός, καθώς και τετραγωνικός πίνακας. Η εκθετική συνάρτηση με όρισμα πραγματικό αριθμό ή πραγματικό τετραγωνικό πίνακα, έχει πάντα ως αποτέλεσμα πραγματικό θετικό αριθμό και πίνακα με πραγματικές θετικές ιδιοτιμές αντίστοιχα. Μεταξύ άλλων τρόπων, η εκθετική συνάρτηση ορίζεται ως η παρακάτω δυναμοσειρά:

$$\exp(x) = e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!} = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \dots$$

Στην περίπτωση που το όρισμα της εκθετικής συνάρτησης είναι τετραγωνικός πίνακας \mathbf{X} πραγματικών ή μιγαδικών αριθμών, διαστάσεων $n \times n$, η εκθετική συνάρτηση μπορεί να οριστεί όπως προηγουμένως, αλλά και ως η εφαρμογή της παραπάνω δυναμοσειράς στις ιδιοτιμές λ_i του πίνακα \mathbf{X} , έπειτα από την ανάλυσή του σε διαγώνιο πίνακα ιδιοτιμών ($\mathbf{\Lambda}$) και πίνακα ιδιοδιανυσμάτων (\mathbf{U}):

$$\expm(\mathbf{X}) = e^{\mathbf{X}} = \sum_{k=0}^{\infty} \frac{\mathbf{X}^k}{k!} = 1 + \mathbf{X} + \frac{\mathbf{X}^2}{2!} + \frac{\mathbf{X}^3}{3!} + \dots$$

$$\mathbf{X} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T, \quad \expm(\mathbf{X}) = \mathbf{U}\expm(\mathbf{\Lambda})\mathbf{U}^T = \mathbf{U}\text{diag}(\exp(\{\lambda_i\}_{i=1}^n))\mathbf{U}^T$$

Η εφαρμογή της εκθετικής συνάρτησης σε μία τιμή $x \in \mathbb{R}$ είτε θετική, είτε αρνητική, επιφέρει πάντοτε θετική τιμή. Ειδικότερα, στην περίπτωση της εφαρμογής της σε τετραγωνικούς πίνακες, παρόλο που οι ιδιοτιμές του πίνακα στον οποίο εφαρμόζεται μπορεί να μην είναι απαραίτητα θετικές ($\lambda_i \in \mathbb{R}$), η εφαρμογή της εκθετικής συνάρτησης εγγυάται ότι αυτές θα είναι πάντα θετικές, το οποίο με τη σειρά του καθιστά τον πίνακα $\expm(\mathbf{X}) \in \mathbb{R}_+^{n \times n}$ θετικά ορισμένο.

2.7.3 Συνδιακύμανση

Η **συνδιακύμανση** (covariance) [66, 67], είναι μια μετρική της Στατιστικής, που χρησιμοποιείται για την εξέταση της εξάρτησης τυχαίων μεταβλητών μεταξύ τους. Τυχαία μεταβλητή ονομάζεται ένα σετ τιμών, το οποίο περιέχει παρατηρήσεις ή μετρήσεις μιας ποσότητας. Οι τυχαίες μεταβλητές μπορούν να αναπαρασταθούν ως διανύσματα, όπου κάθε στοιχείο αυτού του σετ, αποτελεί και μια παρατήρηση. Συγκεκριμένα, η συνάρτηση της συνδιακύμανσης προσδιορίζει το πόσο δύο ή περισσότερες τυχαίες μεταβλητές, διαφέρουν από κοινού από τις μέσες τιμές τους. Προϋπόθεση για την εξέταση της συνδιακύμανσης μεταξύ δύο τυχαίων μεταβλητών, είναι αυτές να έχουν τον ίδιο αριθμό παρατηρήσεων, δηλαδή τα διανύσματα που αναπαριστούν τις τυχαίες μεταβλητές, να έχουν την της ίδιας διάστασης. Η **διακύμανση** (variance), μιας τυχαίας μεταβλητής, συμβολίζεται ως $\text{var}(\mathbf{X})$ ή $\sigma_{\mathbf{X}}^2$, ορίζεται ως η διαφορά των στοιχείων της τυχαίας μεταβλητής από τη μέση τιμή τους. Η διακύμανση μπορεί να προκύψει και από τον τύπο της συνδιακύμανσης, εάν χρησιμοποιηθεί η ίδια τυχαία μεταβλητή και για τα δύο όρισματα της συνάρτησης. Η συνδιακύμανση μεταξύ δύο τυχαίων μεταβλητών, \mathbf{X} και \mathbf{Y} , ορίζεται από το παρακάτω τύπο:

$$\mathbf{X} = \{x_i\}_1^n \Leftrightarrow \mathbf{X} = [x_1 x_2 \dots x_n]^T, \quad \mathbf{Y} = \{y_i\}_1^n \Leftrightarrow \mathbf{Y} = [y_1 y_2 \dots y_n]^T$$

$$\text{cov}(\mathbf{X}, \mathbf{Y}) = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

Η διακύμανση μιας τυχαίας μεταβλητής \mathbf{X} ορίζεται από την παρακάτω σχέση:

$$\text{var}(\mathbf{X}) = \text{cov}(\mathbf{X}, \mathbf{X}) = \sigma_{\mathbf{X}}^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

Όπως αναφέρθηκε, οι τυχαίες μεταβλητές αναπαρίστανται ως διανύσματα. Πολλές τυχαίες μεταβλητές μπορούν να συνθέσουν ένα πίνακα τυχαίων μεταβλητών \mathbf{A} , εάν αυτές στοιβαχτούν η μια δίπλα στην άλλη, έτσι ώστε κάθε στήλη να αποτελεί μια σειρά παρατηρήσεων για μια τυχαία μεταβλητή. Οι διαστάσεις αυτού του πίνακα θα ισούται με $n_{\text{obs}} \times n_{\text{var}}$, όπου n_{obs} ο αριθμός των παρατηρήσεων και n_{var} ο αριθμός των τυχαίων μεταβλητών, όπως φαίνεται δίπλα:

$$\mathbf{A} = [\mathbf{X}, \mathbf{Y}, \dots, \mathbf{Z}] = \begin{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \dots, \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{bmatrix} \end{bmatrix} = \begin{bmatrix} x_1 & y_1 & \cdots & z_1 \\ x_2 & y_2 & \cdots & z_2 \\ \vdots & \vdots & \ddots & \vdots \\ x_n & y_n & \cdots & z_n \end{bmatrix}$$

2.8 Μετρικές Πολλαπλοτήτων των SPD Πινάκων

Παρακάτω, αναλύονται οι μετρικές Riemannian Πολλαπλοτήτων, που αναφέρονται στη δημοσίευση των Gao *et al.* [57], καθώς και η Alpha-Beta Divergence, που χρησιμοποιεί ο υπό ανάλυση αλγόριθμος, για την Εκμάθηση της Μετρικής της Riemannian SPD Πολλαπλότητας στην οποία ανήκουν τα δεδομένα της εφαρμογής.

2.8.1 Affine Invariant Riemannian Metric - AIRM

Η **Affine Invariant Riemannian Metric - AIRM**, όπως αυτή παρουσιάζεται στη δημοσίευση των Pennec *et al.* [68], αποτελεί μια μετρική που βασίζεται στη μέτρηση του μήκους της γεωδειακής, που ενώνει δύο σημεία $\mathbf{X}, \mathbf{Y} \in \text{Sym}_n^+$, μιας SPD Πολλαπλότητας Sym_n^+ , αποτυπώνοντας έτσι, την πραγματική απόσταση μεταξύ των σημείων αυτού του χώρου. Το μειονέκτημα αυτής της μετρικής είναι το γεγονός ότι είναι υπολογιστικά δαπανηρή. Η Σχέση 1 εκφράζει την απόσταση μεταξύ σημείων, που ανήκουν στο χώρο των Riemannian SPD Πολλαπλοτήτων κατά την AIRM. Σημειώνεται επίσης ότι, όπου $\|\cdot\|_F$, η Frobenius νόρμα πινάκων.

$$D^{\text{AIRM}}(\mathbf{X}, \mathbf{Y}) = \left\| \logm \left(\mathbf{X}^{-\frac{1}{2}} \mathbf{Y} \mathbf{X}^{-\frac{1}{2}} \right) \right\|_F \quad (1)$$

2.8.2 Log-Euclidean Metric - LEM

Η **Log-Euclidean Metric - LEM**, είναι μια μετρική για το χώρο των Riemannian SPD Πολλαπλοτήτων, που προτάθηκε από τους Arsigny *et al.* στη δημοσίευσή τους [69] και αποτελεί προσέγγιση της AIRM. Κατά τη μετρική αυτή υπολογίζεται η απόσταση μεταξύ των προβολών των σημείων $\mathbf{X}, \mathbf{Y} \in \text{Sym}_n^+$, μιας SPD Πολλαπλότητας Sym_n^+ , στο σημείο $\mathbf{I} \in \mathbb{R}^{n \times n}$ του Εφαπτόμενου Χώρου $T_{\mathbf{I}}\mathcal{M}$. Τα \mathbf{X} και \mathbf{Y} προβάλλονται στον Εφαπτόμενο Χώρο, μέσω της λογαριθμικής συνάρτησης πινάκων $\logm(\cdot)$. Τέλος, η σχέση που εκφράζει την απόσταση μεταξύ των σημείων $\mathbf{X}, \mathbf{Y} \in \text{Sym}_n^+$, σε SPD Πολλαπλότητες, κατά την LEM, περιγράφεται παρακάτω:

$$D^{\text{LEM}}(\mathbf{X}, \mathbf{Y}) = \|\log(\mathbf{X}) - \log(\mathbf{Y})\|_F \quad (2)$$

2.8.3 Alpha-Beta Divergence

Η **Alpha-Beta Divergence** που προτάθηκε από τους Cichocki and Amari [70], είναι μια μετρική με δύο παραμέτρους, τις α και β , οι οποίες μπορούν να βελτιστοποιηθούν, με σκοπό την καλύτερη αποτύπωση της κατανομής των δεδομένων προς μέτρηση. Αποτελεί μια εναλλακτική προσέγγιση από αυτή της μέτρησης αποστάσεων βασιζόμενη σε γεωδειακές και της προβολής στον Εφαπτόμενο Χώρο, χρησιμοποιείται για τη μέτρηση της απόστασης, μεταξύ δύο σημείων στις Πολλαπλότητες των Συμμετρικών και Θετικά Ορισμένων (SPD) Πινάκων.

Η σχέση που εκφράζει την απόσταση μεταξύ των σημείων $\mathbf{X}, \mathbf{Y} \in \text{Sym}_n^+$, σε Πολλαπλότητες αποτελούμενες από Συμμετρικά και Θετικά Ορισμένους (SPD) Πίνακες, κατά την Alpha-Beta Divergence, είναι η:

$$\begin{aligned} D^{(\alpha, \beta)}(\mathbf{X} \parallel \mathbf{Y}) &= \frac{1}{\alpha\beta} \log \left(\det \left(\frac{\alpha(\mathbf{X}\mathbf{Y}^{-1})^\beta + \beta(\mathbf{X}\mathbf{Y}^{-1})^{-\alpha}}{\alpha + \beta} \right) \right) \\ &= \frac{1}{\alpha\beta} \sum_{u=1}^n \log \left(\frac{\alpha\lambda_u^\beta + \beta\lambda_u^{-\alpha}}{\alpha + \beta} \right) \\ &\quad \text{με } \alpha \neq 0, \beta \neq 0 \text{ και } \alpha + \beta \neq 0 \end{aligned} \quad (3)$$

Όπου $\det(\cdot)$ και $\text{Tr}(\cdot)$ είναι η ορίζουσα και το ίχνος ενός πίνακα αντίστοιχα, λ_u είναι η u -οστή ιδιοτιμή του πίνακα $\mathbf{X}\mathbf{Y}^{-1}$. Όπως φαίνεται στη δημοσίευση των Cichocki *et al.* [71], η παραπάνω συνάρτηση είναι λεία. Επιπλέον, μεταβάλλοντας τις τιμές των α και β , η Alpha-Beta Divergence είναι ισοδύναμη με άλλες μετρικές SPD πινάκων που συναντώνται σε αυτή τη δημοσίευση, όπως η AIRM, η Stein, κ.α. Στο Πίνακα 1, παρουσιάζονται οι ισοδύναμες σχέσεις για τον υπολογισμό τους με τη χρήση πινάκων ή ιδιοτιμών, που προκύπτουν από τη μεταβολή των α και β . Η απόδειξη των παρακάτω, βρίσκεται στη δημοσίευση των Cichocki *et al.* [71].

Τιμές των (α, β)	Ισοδύναμη Μετρική	Μορφή Πινάκων	Μορφή Ιδιοτιμών
$(\alpha \rightarrow 0, \beta \rightarrow 0)$	AIRM [68]	$\ \log(\mathbf{X}^{-\frac{1}{2}}\mathbf{Y}\mathbf{X}^{-\frac{1}{2}})\ _F$	$\frac{1}{2} \sum_{u=1}^n \log^2 \lambda_u$
$(\alpha = \pm \frac{1}{2}, \beta = \pm \frac{1}{2})$	Stein Divergence [72]	$4(\log(\det(\frac{\mathbf{X}+\mathbf{Y}}{2})) - \frac{1}{2} \log(\det(\mathbf{X}\mathbf{Y})))$	$4 \sum_{u=1}^n \log \frac{\lambda_u+1}{2\sqrt{\lambda_u}}$
$(\alpha = \pm 1, \beta \rightarrow 0)$	Jeffrey KL Divergence [73]	$\frac{1}{2} \text{Tr}(\mathbf{X}\mathbf{Y}^{-1} + \mathbf{Y}\mathbf{X}^{-1} - 2I)$	$\frac{1}{2} \sum_{u=1}^n (\sqrt{\lambda_u} - \frac{1}{\sqrt{\lambda_u}})^2$
$(\alpha = 1, \beta = 1)$	Burg Matrix Divergence [74]	$\text{Tr}(\mathbf{X}\mathbf{Y}^{-1} - I) - \log(\det(\mathbf{X}\mathbf{Y}^{-1}))$	$\sum_{u=1}^n (\lambda_u - \log(\lambda_u)) - 1$

Πίνακας 1: Η Alpha-Beta Divergence με διάφορες τιμές των (α, β) .

Περαιτέρω, ισχύει ότι η Alpha-Beta Divergence είναι **αμετάβλητη σε ομοπαράλληλικές παραμορφώσεις** (affine transformations), όπως φαίνεται στο Θεώρημα 2.1:

Θεώρημα 2.1. Δεδομένου δύο SPD πινάκων $\mathbf{X}, \mathbf{Y} \in \text{Sym}_n^+$, η Alpha-Beta Divergence είναι αμετάβλητη σε ομοπαράλληλικές παραμορφώσεις (affine transformations). Δηλαδή, για κάθε αντιστρέψιμο πίνακα $\mathbf{C} \in \mathbb{R}^{n \times n}$, η $D^{(\alpha, \beta)}(\mathbf{X} \parallel \mathbf{Y}) = D^{(\alpha, \beta)}(\mathbf{C}^T \mathbf{X} \mathbf{C} \parallel \mathbf{C}^T \mathbf{Y} \mathbf{C})$.

Απόδειξη. Όπως παρουσιάζεται στη Σχέση 3, που προσδιορίζει την απόσταση μεταξύ δύο SPD πινάκων $\mathbf{X}, \mathbf{Y} \in \text{Sym}_n^+$ κατά την Alpha-Beta Divergence, το γινόμενο των πινάκων $\mathbf{X}\mathbf{Y}^{-1}$ μπορεί να εκφραστεί και ως προς τις ιδιοτιμές του, λ_u . Έτσι, ο πίνακας $\mathbf{X}\mathbf{Y}^{-1}$, μπορεί να αναλυθεί στο διαγώνιο πίνακα ιδιοτιμών (Λ) και το πίνακα ιδιοδιανυσμάτων (U) του ως εξής: $\mathbf{X}\mathbf{Y}^{-1} = U\Lambda U^{-1}$. Επομένως, η εφαρμογή μιας ομοπαράλληλικής παραμόρφωσης, δηλαδή

ο πολλαπλασιασμός κάθε πίνακα αμφίπλευρα, με έναν αντιστρέψιμο πίνακα $C \in \mathbb{R}^{n \times n}$, στον XY^{-1} ως εξής: $C^T X C (C^T Y C)^{-1}$, αναλύεται σε:

$$\begin{aligned} C^T X C (C^T Y C)^{-1} &= C^T X C C^{-1} Y^{-1} C^{-T} = C^T X Y^{-1} C^{-T} = \\ &= C^T U \Lambda U^{-1} C^{-T} = C^T U \Lambda (C^T U)^{-1} \end{aligned} \quad (4)$$

Αφού οι πίνακες U και C είναι αντιστρέψιμοι, το γινόμενο $C^T U$ είναι επίσης αντιστρέψιμο και ο $C^T U \Lambda (C^T U)^{-1}$ είναι όμοιος με το διαγώνιο πίνακα ιδιοτιμών Λ . Με βάση αυτό, οι πίνακες XY^{-1} και $C^T X C (C^T Y C)^{-1}$, έχουν τις ίδιες ιδιοτιμές. Άρα ισχύει ότι:

$$D^{(\alpha, \beta)}(X \| Y) = D^{(\alpha, \beta)}(C^T X C \| C^T Y C) \quad (5)$$

□

3 Αλγόριθμος Point-to-Set and Set-to-Set Distance

Ο προτεινόμενος αλγόριθμος που παρουσιάζεται στο πλαίσιο της διπλωματικής αυτής εργασίας ονομάζεται “**Point-to-Set and Set-to-Set Distance**”, εν συντομία “**PSSSD**”, προτάθηκε από τους Gao *et al.* στην αντίστοιχη δημοσίευσή τους [57] και αποτελεί μια εναλλακτική προσέγγιση της Ταξινόμησης δεδομένων στη Μηχανική Μάθηση, κατά την οποία οι αναπαραστάσεις των εικόνων ανήκουν στο Riemannian χώρο των Πολλαπλοτήτων των Συμμετρικών και Θετικά Ορισμένων (SPD) Πινάκων. Η προσέγγιση που ακολουθεί ο συγκεκριμένος αλγόριθμος, που τον καθιστά ξεχωριστό από άλλους παρόμοιους, είναι ότι αξιοποιεί με ένα διαφορετικό τρόπο την πληροφορία που περιέχεται στα δεδομένα. Συγκεκριμένα, τα δεδομένα εισόδου του αλγορίθμου, που είναι πίνακες συνδιακύμανσης και ανήκουν στο χώρο των Πολλαπλοτήτων των SPD Πινάκων, προβάλλονται σε SPD πίνακες μικρότερων διαστάσεων. Η χρήση των υποπινάκων αυτών έναντι των αρχικών αναπαραστάσεων έχει θετική επίδραση στην επίδοση του αλγορίθμου, όπως υποστηρίζεται και πειραματικά.

3.1 Σύγκριση του PSSSD με άλλες μεθόδους

Πολλές είναι οι προσεγγίσεις που έχουν προταθεί από άλλους ερευνητές στο παρελθόν, που αφορούν εφαρμογές Μηχανικής Μάθησης, των οποίων οι αναπαραστάσεις προέρχονται από δεδομένα τύπου εικόνας και βίντεο και ανήκουν στο χώρο των Πολλαπλοτήτων των SPD Πινάκων. Τέτοιες αναπαραστάσεις αποτελούν οι πίνακες συνδιακύμανσης, οι οποίοι συναντώνται σε αυτή τη διπλωματική εργασία, οι πυρήνες των μετασχηματισμών (kernel matrices) και οι μαγνητικές δεσμιδογραφίες (diffusion tensor images, DTI). Κάποιες εφαρμογές στις οποίες χρησιμοποιούνται τέτοιου είδους αναπαραστάσεις αποτελούν η αναγνώριση προσώπου (face recognition) [75], η αναγνώριση κινήσεων (action recognition) [76], η Μεταφορά Μάθησης (Transfer Learning) [77, 78], η αναγνώριση αντικειμένων (object recognition) [79] και η ανάλυση ιατρικών απεικονίσεων (medical image processing) [80].

3.1.1 Μέθοδοι Ταξινόμησης

Είναι κοινός τόπος, ότι η πλειονότητα των μεθόδων που έχουν προταθεί στο παρελθόν για Επεξεργασία Εικόνας, Υπολογιστική Όραση, Αναγνώριση Προτύπων και Μηχανική Μάθηση, χρησιμοποιούν τεχνικές και αλγόριθμους οι οποίοι εφαρμόζονται θεωρώντας ότι τα δεδομένα μας εδράζονται σε διανυσματικούς χώρους. Ενδεικτικά, κάποιες από αυτές επιτυγχάνουν τη διαδικασία της Ταξινόμησης, προσεγγίζοντας την αναπαράσταση μέσω της μετατροπής της στον Εφαπτόμενο Χώρο (tangent approximation) [81, 82], είτε προσδιορίζοντας το πυρήνα μετασχηματισμού (kernel function), για την ταξινόμηση δεδομένων μέσω αλγορίθμων όπως ο SVM, μέσω της λεγόμενης kernel method [83–86], είτε με τη χρήση Αραιής Αναπαράστασης (Sparse Representation) και Εκμάθησης Λεξικού (Dictionary Learning) [87–89]. Συγκεκριμένα, όπως παρουσιάζεται στις δημοσιεύσεις των Dong *et al.* [81] και των Huang και Van Gool [82], οι SPD αναπαραστάσεις μετατρέπονται σε διανύσματα και μεταφέρονται στον Εφαπτόμενο Χώρο, όπου τροφοδοτούνται σε ένα Multilayer Perceptron (MLP), ένα νευρωνικό δίκτυο που χρησιμοποιείται για ταξινόμηση στον Ευκλείδειο χώρο και αποτελεί επέκταση του Perceptron που εφευρέθηκε από τον Frank Rosenblatt [90].

Μια άλλη εναλλακτική αποτελεί η απεικόνιση μιας SPD αναπαράστασης σε Reproducing Kernel Hilbert Spaces (RKHS), μέσω του προσδιορισμού του πυρήνα μετασχηματισμού τύπου Mercer. Οι Wang *et al.* [84] και Vemulapalli *et al.* [85], χρησιμοποίησαν τέτοιους πυρήνες μετασχηματισμού βασισμένους στην Log-Euclidean Metric (LEM), για την εμφύτευση SPD αναπαραστάσεων σε RKHS, ώστε μετέπειτα να εφαρμόσουν ταξινομητές βασισμένους στον

Ευκλείδειο χώρο. Οι Zhang *et al.* [86] ακολούθησαν μια παρόμοια τακτική, προτείνοντας ένα πυρήνα μετασχηματισμού που βασίζεται στην εκμάθηση μετρικής με τη χρήση της απόκλισης Stein. Επιπλέον, οι Harandi *et al.* [87], όπως και οι Harandi και Salzmann [88] πρότειναν μεθόδους στις οποίες οι SPD αναπαραστάσεις εμφυτεύονται σε RHKS ώστε στη συνέχεια να γίνει Ταξινόμηση των δεδομένων, μέσω της χρήσης Αραιής Αναπαράστασης (Sparse Representation) και Εκμάθησης Λεξικού (Dictionary Learning).

Όπως υποστηρίζεται από τον προτεινόμενο αλγόριθμο, η χρήση των παραπάνω μεθόδων διανυσματοποίησης δεν ενδείκνυται, καθώς όλες τους μεταφέρουν την αρχική SPD αναπαράσταση στον Ευκλείδειο διανυσματικό χώρο με αποτέλεσμα να χάνεται η αρχική δομή της. Αντιθέτως, όπως αναφέρθηκε στο τέλος της πρώτης ενότητας της διπλωματικής αυτής εργασίας, μπορεί να υιοθετηθεί μια διαφορετική προσέγγιση για τη διαδικασία της Ταξινόμησης σε χώρους Riemannian γεωμετρίας, η οποία δεν εκτελεί κάποια προβολή ή απεικόνιση των δεδομένων σε άλλο χώρο, παρά μόνο βασίζεται στην κατηγοριοποίηση των SPD σημείων με βάση την απόστασή τους από άλλα γειτονικά, των οποίων η κλάση είναι δεδομένη. Η τεχνική αυτή ονομάζεται Similarity-based Classification και εφαρμόζεται από τον προτεινόμενο αλγόριθμο.

Η ταξινόμηση μέσω της παραπάνω μεθόδου στο χώρο των Πολλαπλοτήτων των SPD Πινάκων, επιτυγχάνεται με τη χρήση μετρικών οι οποίες σέβονται τους γεωμετρικούς περιορισμούς αυτού του χώρου. Κάποιες από αυτές τις μετρικές αποτελούν η Affine-Invariant Riemannian Metric (AIRM), η Log-Euclidean Metric (LEM), που αναλύονται στις υποενότητες 2.8.1 και 2.8.2 αντίστοιχα, καθώς και η Stein Divergence [72], η Burg Matrix Divergence [74], και η Alpha-Beta Divergence [71, 91], όπως αναλύεται στην υποενότητα 2.8.3. Ενώ οι παραπάνω μετρικές σέβονται τους γεωμετρικούς περιορισμούς των Πολλαπλοτήτων των SPD Πινάκων, λόγω των διαφορετικών κατανομών των δεδομένων που συναντώνται από εφαρμογή σε εφαρμογή, υποστηρίζεται πως η Εκμάθηση Μετρικής (Metric Learning) με τη χρήση της Alpha-Beta Divergence, επιφέρει καλύτερη ακρίβεια στην Ταξινόμηση των δεδομένων.

3.1.2 Μέθοδοι Εκμάθησης Μετρικής

Πολλές μέθοδοι, επιχειρούν την Εκμάθηση Μετρικής αφότου μεταφέρουν τις αναπαραστάσεις σε άλλους χώρους, στους οποίους αυτές διαφοροποιούνται ευκολότερα, καθώς κατασκευάζονται έτσι ώστε οι αναπαραστάσεις όμοιας κλάσης να βρίσκονται σε κοντινή απόσταση μεταξύ τους, ενώ ταυτόχρονα κάθε ομάδα σημείων που αποτελεί μια κλάση βρίσκεται μακριά από άλλες. Παραδείγματος χάριν, κάποιες μέθοδοι λειτουργούν μεταφέροντας τις αναπαραστάσεις τους σε χώρους όπως τον Ευκλείδειο Εφαπτόμενο Χώρο (Euclidean Tangent Space) [92–94], είτε μέσω της εμφύτευσής τους σε Reproducing Kernel Hilbert Spaces (RHKS) [95] και μέσω της προβολής τους σε άλλες Πολλαπλότητες SPD Πινάκων μειωμένων διαστάσεων [96]. Στη μέθοδο που προτάθηκε από τους Vemulapalli και Jacobs [92], οι αναπαραστάσεις μεταφέρονται στον Εφαπτόμενο Χώρο, όπου γίνεται η εκμάθηση των παραμέτρων της μετρικής Mahalanobis. Το μειονέκτημα που έχει αυτή η μέθοδος είναι το γεγονός ότι η αναπαράσταση μεταφέρεται στον Ευκλείδειο διανυσματικό χώρο, με αποτέλεσμα να αλλοιώνεται η δομή της.

Οι Huang *et al.* [93] παρουσίασαν τη μέθοδο που ονόμασαν “Log-Euclidean Metric Learning” (LEML), η οποία χρησιμοποιεί την LEM για να μαθαίνει τη συνάρτηση εκείνη, που θα μεταφέρει την αρχική αναπαράσταση σε έναν Εφαπτόμενο Χώρο στον οποίο τα σημεία είναι πιο ευδιάκριτα. Η μέθοδος αυτή έχει μεγάλη αποδοτικότητα, ενώ παράλληλα αποφεύγει την αλλοίωση της δομής της αρχικής αναπαράστασης. Βασιζόμενοι στην LEML, οι Zhou *et al.* [94] παρουσίασαν τη δική τους προσέγγιση, την “ α -based covariance-like metric learning (α -CML)”, κατά την οποία οι ιδιοτιμές των SPD αναπαραστάσεων μεταβάλλονται, επιτυγχάνοντας καλύτερη διαφοροποίηση μεταξύ των δεδομένων. Το μειονέκτημα της μεθόδου αυτής όμως είναι

ότι η εκπαίδευση που λαμβάνει χώρα για τη μεταβολή των ιδιοτιμών των δεδομένων δεν μπορεί να χρησιμοποιηθεί για τη μεταβολή ιδιοτιμών άλλων δεδομένων και επομένως, πρέπει να εκτελεστεί εκ' νέου.

Όπως αναφέρθηκε προηγουμένως, μια άλλη προσέγγιση που ακολουθείται είναι εκείνη που απεικονίζει τις SPD αναπαραστάσεις σε RHKS. Οι Wang *et al.* [95] έθεσαν το μαθηματικό και υπολογιστικό πλαίσιο για τη γενίκευση της απόστασης μεταξύ δύο SPD αναπαραστάσεων, μετρήσιμων διαστάσεων, σε αναπαραστάσεις πινάκων συνδιακύμανσης απεικονισμένων σε RHKS, απείρων διαστάσεων, με τη χρήση Hilbert-Schmidt τελεστών. Με σκοπό τη διατήρηση της δομής της SPD αναπαραστάσεως, η οποία αλλοιώνεται όταν προβάλλεται στον Ευκλείδειο χώρο, οι Harandi *et al.* [96] πρότειναν μια μέθοδο εκμάθησης μετρικής σε ένα μειωμένο σε διαστάσεις χώρο των Πολλαπλοτήτων αποτελούμενο από SPD πίνακες, όπου είναι ευκολότερη η εκμάθηση μετρικής.

3.2 Παρουσίαση της μεθόδου που ακολουθείται

Ένα κοινό στοιχείο των μεθόδων Ταξινόμησης που αναφέρθηκαν παραπάνω, είναι το γεγονός ότι οι SPD πίνακες που χρησιμοποιούνται ως αναπαραστάσεις των αρχικών δεδομένων αντιμετωπίζονται ως ένα στοιχείο. Το περιεχόμενο των SPD Πινάκων αποτελείται από μια συλλογή στατιστικών ποσοτήτων που περιγράφουν τη σχέση μεταξύ χαρακτηριστικών που έχουν εξαχθεί από τα δεδομένα. Στη δημοσίευση που αναλύεται, υποστηρίζεται ότι τα χαρακτηριστικά αυτά έχουν διαφορετικές κατανομές, καθώς και επίδραση στη σύνθεση και στη μέτρηση απόστασης μεταξύ των αναπαραστάσεων. Έτσι, κρίνεται απαραίτητη η μεμονωμένη διαχείρισή τους, για την κατασκευή ενός πιο εύρωστου (robust) ταξινομητή.

Το πρόβλημα της Ταξινόμησης προσεγγίζεται από τον αλγόριθμο PSSSD ως ένα πρόβλημα μέτρησης αποστάσεων μεταξύ των προβολών των αρχικών αναπαραστάσεων σε SPD χώρους μειωμένων διαστάσεων, όπου αυτές διαφοροποιούνται περισσότερο. Στο τέλος, γίνεται η σύνθεση των επιμέρους αποστάσεων σε μία, η οποία χαρακτηρίζει την απόσταση μεταξύ των αρχικών αναπαραστάσεων. Η μέθοδος αυτή, χαρακτηρίζεται ως εκμάθηση πολλαπλών μετρικών (multi-metric learning), αφού για κάθε προβολή ανακαλύπτεται η μετρική που εκφράζει καλύτερα την απόσταση μεταξύ των δεδομένων που βρίσκονται σε αυτή. Η τεχνική αυτή βρίσκει εφαρμογές στην αναγνώριση προσώπων και συγγένειας (facial and kinship verification) [97], αναγνώριση εκφράσεων του προσώπου (facial expression recognition) [98], καθώς και σε εφαρμογές που συνθέτουν μετρήσεις διαφορετικών αισθητήρων (multi-sensor fusion) [99]. Συνοπτικά, τα βήματα που ακολουθούνται από τον αλγόριθμο είναι τα ακόλουθα:

1. **Μετασχηματισμός Point-to-Set (T_S):** Από μια αρχική SPD αναπαράσταση (σημείο), κατασκευάζεται ένα σύνολο (σετ) επιμέρους αναπαραστάσεων ή προβολών (projections), μειωμένων διαστάσεων.
2. **Set-to-Set Μετρική (D_S):**
 - (a) **Υπολογισμός επιμέρους αποστάσεων** (subdistances) μεταξύ των προβολών που προκύπτουν από τις αρχικές SPD αναπαραστάσεις.
 - (b) **Σύνθεση ολικής απόστασης** μέσω της αφομοίωσης των επιμέρους αποστάσεων με κάποιους βαθμούς βαρύτητας.

3.2.1 Μετασχηματισμός Σημείου σε Σύνολο

Η λειτουργία του Μετασχηματισμού Σημείου σε Σύνολο (Point-to-Set) (T_S) επιτυγχάνεται με τη χρήση της **συνάρτησης προβολής** (low-dimensional projection) $f_W(\cdot)$, κατά την

οποία ένας αρχικός $n \times n$ SPD πίνακας $\mathbf{X}_i \in Sym_n^+$, προβάλλεται σε m επιμέρους πίνακες $\mathbf{X}_i^k \in Sym_p^+$ με $k = \{1, 2, \dots, m\}$, μειωμένων διαστάσεων $p \times p$. Οι επιμέρους αυτές αναπαραστάσεις, συνθέτουν το σετ προβολών $\mathcal{X}_i = \{f_{\mathbf{W}}^k(\mathbf{X}_i)\}_{k=1}^m = \{\mathbf{X}_i^k\}_{k=1}^m$ όπως φαίνεται στις παρακάτω ενδεικτικές σχέσεις:

$$\begin{aligned}\mathbf{X}_i^1 &= f_{\mathbf{W}}^1(\mathbf{X}_i) = \mathbf{W}_1^T \mathbf{X}_i \mathbf{W}_1 \\ \mathbf{X}_i^2 &= f_{\mathbf{W}}^2(\mathbf{X}_i) = \mathbf{W}_2^T \mathbf{X}_i \mathbf{W}_2 \\ &\vdots \\ \mathbf{X}_i^m &= f_{\mathbf{W}}^m(\mathbf{X}_i) = \mathbf{W}_m^T \mathbf{X}_i \mathbf{W}_m\end{aligned}\quad (6)$$

Η παράμετρος $\mathbf{W}_k \in \mathbb{R}^{n \times p}$ της συνάρτησης $f_{\mathbf{W}}^k(\cdot)$, αποτελεί το λεγόμενο **πίνακα προβολής** (projection matrix), ο οποίος χρησιμοποιείται για την προβολή της αρχικής αναπαράστασης σε μια μικρότερων διαστάσεων, εκτελώντας πολλαπλασιασμό πινάκων. Η παράμετρος αυτή, αποτελεί μια από τις τρεις τις οποίες βελτιστοποιεί ο αλγόριθμος και ο ρόλος της είναι καθοριστικός στην ικανότητα του αλγορίθμου να ταξινομεί, καθώς οι αναπαραστάσεις που παράγονται μέσω αυτής, αναμένονται να αντικατοπτρίζουν με ακρίβεια τα αρχικά δεδομένα, παρόλο που έχουν μειωθεί σε διαστάσεις, όπως και να διαφέρουν μεταξύ τους όσο γίνεται περισσότερο, το οποίο θα επισημανθεί και αργότερα. Περαιτέρω, απαιτείται κάθε πίνακας \mathbf{W}_k να είναι πλήρους τάξης, ώστε να εγγυάται ότι κάθε προβολή της αρχικής αναπαράστασης $\mathbf{X}_i^k = \mathbf{W}_k^T \mathbf{X}_i \mathbf{W}_k$, θα είναι SPD πίνακας. Η απαίτηση αυτή επιτυγχάνεται όταν τίθεται ως περιορισμός, κάθε πίνακας \mathbf{W}_k , να είναι ορθογώνιος. Ο περιορισμός αυτός, όπως αποδεικνύεται στο Θεώρημα 3.1, εφαρμόζεται χωρίς βλάβη της γενικότητας.

Θεώρημα 3.1. Δεδομένου δύο SPD πινάκων $\mathbf{X}, \mathbf{Y} \in Sym_n^+$ και πίνακα προβολής πλήρους τάξης $\tilde{\mathbf{W}} \in \mathbb{R}^{n \times p}$, με $n \geq p$. Κατά την Alpha-Beta Divergence, η απόσταση μεταξύ των προβολών των αρχικών αναπαραστάσεων δίνεται από τη σχέση: $D^{(\alpha, \beta)}(\tilde{\mathbf{W}}^T \mathbf{X} \tilde{\mathbf{W}} \parallel \tilde{\mathbf{W}}^T \mathbf{Y} \tilde{\mathbf{W}})$. Ισχύει ότι, η βελτιστοποίηση οποιασδήποτε συνάρτησης κόστους $\mathcal{L}(D^{(\alpha, \beta)}(\tilde{\mathbf{W}}^T \mathbf{X} \tilde{\mathbf{W}} \parallel \tilde{\mathbf{W}}^T \mathbf{Y} \tilde{\mathbf{W}}))$, γίνεται χωρίς βλάβη της γενικότητας ακόμα και αν απαιτείται ο πίνακας προβολής $\tilde{\mathbf{W}}$ να είναι ορθογώνιος.

Απόδειξη. Ισχύει ότι η Alpha-Beta Divergence, είναι αμετάβλητη σε ομοπαράλληλες παραμορφώσεις (affine invariant) δηλαδή $D^{(\alpha, \beta)}(\mathbf{X} \parallel \mathbf{Y}) = D^{(\alpha, \beta)}(\mathbf{C}^T \mathbf{X} \mathbf{C} \parallel \mathbf{C}^T \mathbf{Y} \mathbf{C})$, με $\mathbf{X}, \mathbf{Y} \in Sym_n^+$ (Θεώρημα 2.1), για κάθε αναστρέψιμο πίνακα $\mathbf{C} \in \mathbb{R}^{n \times n}$. Επιπλέον, με τη χρήση της Αποσύνθεσης QR (QR Decomposition), κάθε πίνακας πλήρους τάξης, αναλύεται σε έναν ορθογώνιο $\mathbf{W} \in \mathbb{R}^{n \times p}$ και σε έναν άνω τριγωνικό πίνακα $\mathbf{B} \in \mathbb{R}^{p \times p}$, δηλαδή $\tilde{\mathbf{W}} = \mathbf{W} \mathbf{B}$. Δεδομένου του γεγονότος ότι ο άνω τριγωνικός πίνακας \mathbf{B} , είναι αντιστρέψιμος, καθώς και της ιδιότητας της Alpha-Beta Divergence, για τους $\mathbf{X}, \mathbf{Y} \in Sym_n^+$, αυτή γράφεται ως:

$$\begin{aligned}D^{(\alpha, \beta)}(\tilde{\mathbf{W}}^T \mathbf{X} \tilde{\mathbf{W}} \parallel \tilde{\mathbf{W}}^T \mathbf{Y} \tilde{\mathbf{W}}) &= \\ &= D^{(\alpha, \beta)}(\mathbf{B}^T \mathbf{W}^T \mathbf{X} \mathbf{B} \mathbf{W} \parallel \mathbf{B}^T \mathbf{W}^T \mathbf{Y} \mathbf{B} \mathbf{W}) = \\ &= D^{(\alpha, \beta)}(\mathbf{W}^T \mathbf{X} \mathbf{W} \parallel \mathbf{W}^T \mathbf{Y} \mathbf{W})\end{aligned}\quad (7)$$

Επομένως, για κάθε πίνακα πλήρους τάξης $\tilde{\mathbf{W}}$, υπάρχει ορθογώνιος \mathbf{W} για τον οποίο εγγυάται ότι $\mathcal{L}(D^{(\alpha, \beta)}(\tilde{\mathbf{W}}^T \mathbf{X} \tilde{\mathbf{W}} \parallel \tilde{\mathbf{W}}^T \mathbf{Y} \tilde{\mathbf{W}})) = \mathcal{L}(D^{(\alpha, \beta)}(\mathbf{W}^T \mathbf{X} \mathbf{W} \parallel \mathbf{W}^T \mathbf{Y} \mathbf{W}))$ \square

Ο περιορισμός που τίθεται για τον πίνακα προβολής \mathbf{W}_k , η απαίτηση δηλαδή να είναι ορθογώνιος, πληροί τις γεωμετρικές προϋποθέσεις που πρέπει να έχει ένας πίνακας για να ανήκει

στο χώρο των Stiefel Πολλαπλοτήτων (Stiefel Manifolds) [100]. Επιπλέον, λόγω του ότι η συνάρτηση κόστους είναι αμετάβλητη σε ομοπαράλληλες παραμορφώσεις, ο πολλαπλασιασμός του \mathbf{W}_k , με έναν ορθογώνιο (και αντιστρέψιμο) πίνακα $\mathbf{R} \in \mathbb{R}^{p \times p}$ από τα δεξιά του, δε θα μεταβάλλει τη συνάρτηση, δηλαδή θα ισχύει ότι $\mathcal{L}(D^{(\alpha, \beta)}(\mathbf{R}^T \mathbf{W}_k^T \mathbf{X} \mathbf{W}_k \mathbf{R} \| \mathbf{R}^T \mathbf{W}_k^T \mathbf{Y} \mathbf{W}_k \mathbf{R})) = \mathcal{L}(D^{(\alpha, \beta)}(\mathbf{W}_k^T \mathbf{X} \mathbf{W}_k \| \mathbf{W}_k^T \mathbf{Y} \mathbf{W}_k))$. Λόγω αυτού, ο πίνακας προβολής \mathbf{W}_k , τελικά ανήκει στο χώρο των Grassmann Πολλαπλοτήτων [100], ο οποίος αποτελεί επίσης το σύνολο πηλίκου (quotient space) των Stiefel Πολλαπλοτήτων.

Όπως αναφέρθηκε προηγουμένως, κάθε προβολή της αρχικής αναπαράστασης αναμένεται να διαφέρει όσο γίνεται περισσότερο από τις υπόλοιπες προβολές, καθώς έτσι θα παρέχεται στον αλγόριθμο μια πιο ειδικευμένη γνώση για ένα αντικείμενο, αφού κάθε προβολή, θα είναι κατά κάποιο τρόπο, μια διαφορετική πτυχή της αρχικής αναπαράστασης. Αυτό επιτυγχάνεται όταν ο πίνακας προβολής \mathbf{W}_k που χρησιμοποιείται για την εξαγωγή της κάθε προβολής από μια αρχική αναπαράσταση, διαφέρει από τους υπόλοιπους όσο γίνεται περισσότερο, επισημαίνοντας έτσι διαφορετικά χαρακτηριστικά του αρχικού. Το γεγονός ότι επιβάλλεται οι πίνακες προβολής \mathbf{W}_k να είναι ορθογώνιοι, και κατ' επέκταση να ανήκουν στο χώρο των Grassmann Πολλαπλοτήτων, μεγιστοποιεί την απόσταση μεταξύ τους. Ενδεικτικά, για δύο πίνακες προβολής $\mathbf{W}_k, \mathbf{W}_l \in Gr(n, p)$, η μετρική που χαρακτηρίζει την απόσταση μεταξύ τους, μέσω της προβολής τους από τον χώρο των Grassmann Πολλαπλοτήτων στον εφαπτόμενο χώρο, δίνεται από τον τύπο: $p - \|\mathbf{W}_k^T \mathbf{W}_l\|_F$. Δηλαδή, για κάθε προβολή επιβάλλεται ο περιορισμός $\mathbf{W}_k^T \mathbf{W}_l = \mathbf{0}$, με $k \neq l$ και $\mathbf{0} \in \mathbb{R}^{p \times p}$, το μηδενικό πίνακα διαστάσεων $p \times p$.

Όλοι οι πίνακες προβολής \mathbf{W}_k , συνθέτουν έναν **ολικό πίνακα προβολής** \mathbf{W} , ο οποίος περιέχει πίνακες προβολής \mathbf{W}_k , πλήθους m και μεγέθους $p \times p$, $\mathbf{W} = [\mathbf{W}_1, \mathbf{W}_2, \dots, \mathbf{W}_m] \in \mathbb{R}^{n \times mp}$. Ο πίνακας \mathbf{W} , είναι επίσης ορθογώνιος, που σημαίνει ότι για αυτόν ισχύει: $\mathbf{W}^T \mathbf{W} = \mathbf{I}_{mp}$. Λόγω αυτής της ιδιότητας, ανήκει στο χώρο των Stiefel Πολλαπλοτήτων [100].

Παρατήρηση 3.1. *Επισημαίνεται ότι, αντί για το σετ προβολών $\mathcal{X}_i = \{\mathbf{X}_i^k \in Sym_p^+\}_{k=1}^m$, χρησιμοποιείται ο πίνακας \mathbf{Z}_i , του οποίου τα στοιχεία ισοδυναμούν με το περιεχόμενο του \mathcal{X}_i . Ο πίνακας αυτός προκύπτει από τον πολλαπλασιασμό του ολικού πίνακα προβολής \mathbf{W} , και της αρχικής αναπαράστασης \mathbf{X}_i όπως παρουσιάζεται στη Σχέση 8:*

$$\begin{aligned} \mathbf{Z}_i &= \text{mask}(\mathbf{W}^T \mathbf{X}_i \mathbf{W}) = \\ &= \text{mask} \left(\begin{bmatrix} \mathbf{W}_1^T \mathbf{X}_i \mathbf{W}_1 & \dots & \mathbf{W}_1^T \mathbf{X}_i \mathbf{W}_m \\ \vdots & \ddots & \vdots \\ \mathbf{W}_m^T \mathbf{X}_i \mathbf{W}_1 & \dots & \mathbf{W}_m^T \mathbf{X}_i \mathbf{W}_m \end{bmatrix} \right) = \\ &= \begin{bmatrix} \mathbf{X}_i^1 & 0 & \dots & 0 \\ 0 & \mathbf{X}_i^2 & \dots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & \dots & \mathbf{X}_i^m \end{bmatrix} \end{aligned} \quad (8)$$

Ο πίνακας $\mathbf{Z}_i \in \mathbb{R}^{mp \times mp}$, του οποίου τα διαγώνια στοιχεία αποτελούν τις προβολές $\mathbf{X}_i^k \in Sym_p^+$, μιας αρχικής αναπαράστασης $\mathbf{X}_i \in Sym_n^+$, υλοποιείται στον κώδικα και χρησιμοποιείται, γιατί λόγω της απλής δομής του είναι εύκολη η αντιστροφή του και ο υπολογισμός των ιδιοτιμών του.

3.2.2 Set-to-Set Μετρική

Σε αυτό το στάδιο της λειτουργίας του, ο αλγόριθμος υπολογίζει την απόσταση μεταξύ δύο αρχικών αναπαράστασεων $\mathbf{X}_i, \mathbf{X}_j \in Sym_n^+$, που προκύπτει από τη σύνθεση των επιμέρους

αποστάσεων (subdistances) που υπολογίζονται από τα σετ προβολών $\mathcal{X}_i, \mathcal{X}_j$. Συγκεκριμένα, χρησιμοποιούνται η **συνάρτηση μέτρησης επιμέρους απόστασης** (subdistance measure) $g_A(\cdot, \cdot)$, για τη μέτρηση της απόστασης μεταξύ δύο προβολών μέσω της Alpha-Beta Divergence και η **συνάρτηση σύνθεσης αποστάσεων** (integration function) $h_M(\cdot)$, η οποία συνθέτει τις πολλαπλές επιμέρους αποστάσεις $\{g_A^k(\cdot, \cdot)\}_{k=1}^m$ που έχουν υπολογιστεί, σε μια ολική απόσταση. Οι παραπάνω δύο συναρτήσεις, χρησιμοποιούν τις παραμέτρους $\mathbf{A} = [(\alpha_1, \beta_1), (\alpha_2, \beta_2), \dots, (\alpha_m, \beta_m)] \in \mathbb{R}^{m \times 2}$ και $\mathbf{M} \in \mathbb{R}^{m \times m}$ αντίστοιχα, οι οποίες αποτελούν τις υπόλοιπες δύο, οι οποίες βελτιστοποιούνται από τον παρών αλγόριθμο. Η παράμετρος \mathbf{A} , περιέχει τα ζεύγη τιμών (α, β) που χρησιμοποιούνται από την Alpha-Beta Divergence, των m υπο-χώρων στους οποίους βρίσκονται οι προβολές των αρχικών δεδομένων. Η παράμετρος \mathbf{M} , η οποία ονομάζεται **παράμετρος σύνθεσης** (integration parameter), περιέχει τα βάρη με τα οποία θα πολλαπλασιαστούν οι επιμέρους αποστάσεις που έχουν υπολογιστεί από τις $\{g_A^k(\cdot, \cdot)\}_{k=1}^m$. Η σωστή εκμάθησή τους θα δώσει ακριβέστερες μετρήσεις, καθώς με τα ζεύγη (α, β) της παραμέτρου \mathbf{A} αποτυπώνεται η δομή του χώρου στην οποία θα βρίσκεται κάθε προβολή, όπως επίσης και ο βαθμός συμμετοχής της κάθε επιμέρους απόστασης που περιέχεται στην \mathbf{M} , ο οποίος θα χρησιμοποιηθεί για να συνθέσει την ολική.

Η συνάρτηση μέτρησης επιμέρους αποστάσεων ($g_A(\cdot, \cdot)$), υλοποιεί την Alpha-Beta Divergence και με τη χρήση των ζευγών (α_k, β_k) , που έχουν υπολογιστεί από τον αλγόριθμο μέσω βελτιστοποίησης [71, 91]. Η απόσταση d_{ij}^k μεταξύ δύο προβολών $\mathbf{X}_i^k, \mathbf{X}_j^k$ δίνεται από την παρακάτω σχέση:

$$\begin{aligned} d_{ij}^k &= g_A^k(\mathbf{X}_i^k, \mathbf{X}_j^k) = D^{(\alpha_k, \beta_k)}(\mathbf{X}_i^k \| \mathbf{X}_j^k) = \\ &= \frac{1}{\alpha_k \beta_k} \log \left(\det \left(\frac{\alpha_k (\mathbf{X}_i^k (\mathbf{X}_j^k)^{-1})^{\beta_k} + \beta_k (\mathbf{X}_j^k (\mathbf{X}_i^k)^{-1})^{-\alpha_k}}{\alpha_k + \beta_k} \right) \right) \\ &= \frac{1}{\alpha_k \beta_k} \sum_{u=1}^p \log \left(\frac{\alpha_k (\lambda_{iju}^k)^{\beta_k} + \beta_k (\lambda_{iju}^k)^{-\alpha_k}}{\alpha_k + \beta_k} \right) \end{aligned} \quad (9)$$

Όπου λ_{iju}^k η u -οστή ιδιοτιμή του πίνακα $\mathbf{X}_i^k (\mathbf{X}_j^k)^{-1}$, (α_k, β_k) το εκάστοτε ζεύγος τιμών που χρησιμοποιούνται από την Alpha-Beta Divergence και $\mathbf{d}_{ij} = [d_{ij}^1, d_{ij}^2, \dots, d_{ij}^m]^T \in \mathbb{R}^{m \times 1}$ το διάνυσμα απόστασης (distance vector) μεταξύ δύο σετ προβολών $\mathcal{X}_i, \mathcal{X}_j$. Για τη σύνθεση της απόστασης μεταξύ των αρχικών αναπαραστάσεων $D^\Theta(\mathbf{X}_i, \mathbf{X}_j)$, γίνεται χρήση της παραμέτρου \mathbf{M} , όπως φαίνεται στη Σχέση 10:

$$\begin{aligned} D^\Theta(\mathbf{X}_i, \mathbf{X}_j) &= \\ &= D_S(\mathcal{X}_i, \mathcal{X}_j) = h_M(d_{ij}^1, d_{ij}^2, \dots, d_{ij}^m) = \mathbf{d}_{ij}^T \mathbf{M} \mathbf{d}_{ij} = \\ &= \sum_{k=1}^m \sum_{l=1}^m (d_{ij}^k \cdot \mathbf{M}_{kl} \cdot d_{ij}^l) \end{aligned} \quad (10)$$

Όπου \mathbf{M}_{kl} το στοιχείο της k γραμμής και l στήλης του πίνακα \mathbf{M} , το οποίο περιέχει το συντελεστή βαρύτητας που αντικατοπτρίζει τη συνεισφορά των $\mathbf{X}_i^k, \mathbf{X}_j^l$ προβολών στη σύνθεση της απόστασης μεταξύ των αρχικών αναπαραστάσεων $\mathbf{X}_i, \mathbf{X}_j$. Από τη Σχέση 10 διαπιστώνεται ότι, η Set-to-Set Μετρική (D_S) είναι εξίσωση δευτέρου βαθμού. Επιπλέον, το διάνυσμα απόστασης \mathbf{d}_{ij} δύο αναπαραστάσεων $\mathbf{X}_i, \mathbf{X}_j$, θα είναι μηδενικό ($\mathbf{d}_{ij} = \mathbf{0}_m$) αν και μόνο αν συγκρίνονται οι ίδιες αναπαραστάσεις $\mathbf{X}_i = \mathbf{X}_j$. Επομένως, είναι αναμενόμενο το διάνυσμα απόστασης \mathbf{d}_{ij} και κατ' επέκταση η ολική απόσταση μεταξύ δύο αναπαραστάσεων $D^\Theta(\mathbf{X}_i, \mathbf{X}_j)$, να είναι

πάντα θετικές ποσότητες. Για το λόγω του ότι η απόσταση θα είναι πάντα θετική, ο πίνακας M εγγυάται να είναι Συμμετρικός και Θετικά Ορισμένος [101–103].

Τα παραπάνω γνωρίσματα καθιστούν την Point-to-Set and Set-to-Set απόσταση παρόμοια με τη Mahalanobis μετρική, κατά την οποία κάθε διάσταση του χώρου στον οποίο χρησιμοποιείται θεωρείται πως έχει τη δική της κατανομή δεδομένων και συνεισφορά στην εξαγωγή μιας τελικής τιμής, όπως γίνεται στην προκειμένη περίπτωση με την παράμετρο M και τις προβολές των αρχικών αναπαραστάσεων. Περαιτέρω, όπως αποδεικνύεται στο Θεώρημα 3.2, η Σχέση 10 που περιγράφει την τελική απόσταση μεταξύ δύο αρχικών αναπαραστάσεων, μπορεί να προκύψει και από τον πολλαπλασιασμό βαρών στην Alpha-Beta Divergence υψωμένη στο τετράγωνο (squared Alpha-Beta Divergence) ως ακολούθως:

Θεώρημα 3.2. Η Set-to-Set Distance $D_S(\mathcal{X}_i, \mathcal{X}_j)$ που προσδιορίζεται στη Σχέση 10, μπορεί να προκύψει και από τον πολλαπλασιασμό βαρών στην Alpha-Beta Divergence υψωμένη στο τετράγωνο.

Απόδειξη. Όπως έχει αναφερθεί στην Παρατήρηση 3.1 ένα σετ προβολών $\mathcal{X}_i = \{\mathbf{X}_i^k\}_{k=1}^m$, αναπαρίσταται αλλιώς από το διαγώνιο πίνακα $\mathbf{Z}_i \in \mathbb{R}^{mp \times mp}$, του οποίου τα στοιχεία είναι οι προβολές $\mathbf{X}_i^k \in Sym_p^+$ της αρχικής αναπαράστασης \mathbf{X}_i , όπως φαίνεται παρακάτω:

$$\mathbf{Z}_i = \begin{bmatrix} \mathbf{X}_i^1 & 0 & \cdots & 0 \\ 0 & \mathbf{X}_i^2 & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \cdots & \mathbf{X}_i^m \end{bmatrix} \quad (11)$$

Για τη μέτρηση της απόστασης μεταξύ δύο διαγώνιων πινάκων $\mathbf{Z}_i, \mathbf{Z}_j$, γίνεται χρήση της Alpha-Beta Divergence υψωμένης στο τετράγωνο, ως ακολούθως:

$$D^2 = (D^{(\alpha, \beta)}(\mathbf{Z}_i \| \mathbf{Z}_j))^2 = \left(\frac{1}{\alpha\beta} \sum_{u=1}^{mp} \log \left(\frac{\alpha\lambda_u^\beta + \beta\lambda_u^{-\alpha}}{\alpha + \beta} \right) \right)^2 \quad (12)$$

Όπου λ_u η u -οστή ιδιοτιμή του πίνακα $\mathbf{Z}_i \mathbf{Z}_j^{-1}$, ο οποίος είναι ο ακόλουθος:

$$\mathbf{Z}_i \mathbf{Z}_j^{-1} = \begin{bmatrix} \mathbf{X}_i^1 (\mathbf{X}_j^1)^{-1} & 0 & \cdots & 0 \\ 0 & \mathbf{X}_i^2 (\mathbf{X}_j^2)^{-1} & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \cdots & \mathbf{X}_i^m (\mathbf{X}_j^m)^{-1} \end{bmatrix} \quad (13)$$

Οι ιδιοτιμές του πίνακα $\mathbf{Z}_i \mathbf{Z}_j^{-1}$ είναι οι $\{\lambda_u\}_{u=(k-1)p+1}^{k \cdot p}$, όπου $k = \{1, 2, \dots, m\}$, m και p το πλήθος και οι διαστάσεις των προβολών αντίστοιχα. Η συνάρτηση 12, μπορεί να αναλυθεί στην παρακάτω μορφή (Σχέση 14), στην οποία απομονώνεται η ποσότητα που εκφράζει την απόσταση για κάθε m προβολή της αρχικής αναπαράστασης:

$$D^2 = \left(\sum_{k=1}^m \left(\frac{1}{\alpha\beta} \sum_{u=(k-1)p+1}^{k \cdot p} \log \left(\frac{\alpha\lambda_u^\beta + \beta\lambda_u^{-\alpha}}{\alpha + \beta} \right) \right) \right)^2 \quad (14)$$

Ο όρος: $(1/\alpha\beta) \sum_{u=(k-1)p+1}^{k \cdot p} \log ((\alpha\lambda_u^\beta + \beta\lambda_u^{-\alpha})/(\alpha + \beta))$ αποτελεί την απόσταση μεταξύ των προβολών $\mathbf{X}_i^k, \mathbf{X}_j^k$, ο οποίος συμβολίζεται ως d^k :

$$D^2 = \left(\sum_{k=1}^m d^k \right)^2 = \sum_{k=1}^m \sum_{l=1}^m d^k d^l \quad (15)$$

Δεδομένου ότι οι χώροι στους οποίους βρίσκονται οι διαφορετικές προβολές $\mathbf{X}_i^k, \mathbf{X}_j^k$, έχουν ξεχωριστή δομή και για αυτούς χρησιμοποιούνται διαφορετικά ζεύγη τιμών (α, β) , η Σχέση 15 φτάνει σε μορφή η οποία ισοδυναμεί με τη Σχέση 9, η οποία είναι η συνάρτηση μέτρησης επιμέρους αποστάσεων g_A , βασιζόμενη στην Alpha-Beta Divergence:

$$d^k = \frac{1}{\alpha_k \beta_k} \sum_{u=(k-1) \cdot p+1}^{k \cdot p} \log \left(\frac{\alpha_k \lambda_u^{\beta_k} + \beta_k \lambda_u^{-\alpha_k}}{\alpha_k + \beta_k} \right) \quad (16)$$

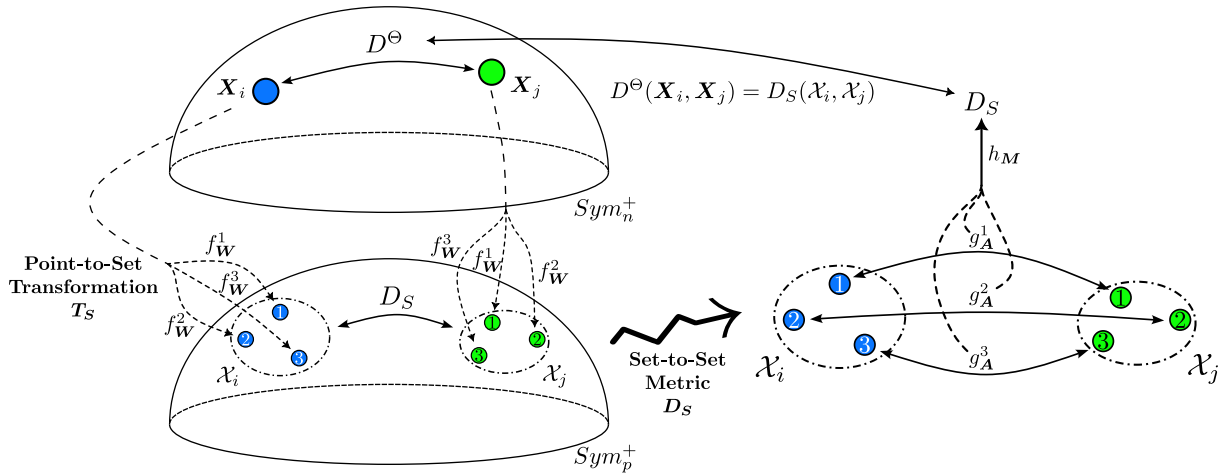
Τέλος, χρησιμοποιώντας διαφορετικά βάρη w_k , τα οποία αναλογούν στον βαθμό συμμετοχής της κάθε απόστασης d^k στη σύνθεση μιας ολικής, η απόσταση D^2 , που περιγράφεται από τη Σχέση 17, τελικά παίρνει τη μορφή της Set-to-Set Μετρικής, που περιγράφεται από τη Σχέση 10:

$$D^2 = \left(\sum_{k=1}^m w_k d^k \right)^2 = \sum_{k=1}^m \sum_{l=1}^m w^k w^l d^k d^l = \sum_{k=1}^m \sum_{l=1}^m M_{kl} d^k d^l \quad (17)$$

Συμπεραίνεται ότι, η Set-to-Set απόσταση ισοδυναμεί με εκείνη που συντίθεται από τη χρήση της Alpha-Beta Divergence υψωμένης στο τετράγωνο, για τον υπολογισμό αποστάσεων μεταξύ προβολών, για τις οποίες λαμβάνει ξεχωριστά ζεύγη (α, β) και βαθμούς συμμετοχής w_k . \square

Συνοψίζοντας, η μέτρηση της απόστασης μεταξύ δύο αρχικών αναπαραστάσεων $\mathbf{X}_i, \mathbf{X}_j \in Sym_n^+$, κατά την Point-to-Set and Set-to-Set Απόσταση, προκύπτει από τη σύνθεση των αποστάσεων μεταξύ των προβολών που προκύπτουν από τις αρχικές αναπαραστάσεις, σε μια. Στο παρακάτω διάγραμμα ροής (Εικόνα 7) παρουσιάζεται η λειτουργία του αλγορίθμου PSSSD, καθώς και οι συναρτήσεις που χρησιμοποιούνται για την επίτευξη κάθε διαδικασίας του. Στο αριστερό μέρος του διαγράμματος, εικονίζεται μια Πολλαπλότητα Sym_n^+ , της οποίας τα στοιχεία είναι δύο SPD πίνακες $\mathbf{X}_i, \mathbf{X}_j$, που αναπαρίστανται ως τον μπλε και πράσινο κύκλο αντίστοιχα. Κατά το στάδιο Point-to-Set Transformation T_S , οι αρχικές αναπαραστάσεις προβάλλονται σε $m = 3$ επιμέρους αναπαραστάσεις (προβολές) Sym_p^+ , μέσω της συνάρτησης f_W και συμβολίζονται από άλλους κύκλους με διαφορετικούς αριθμούς στο εσωτερικό τους, ανάλογα με την προβολή από την οποία προέρχονται. Στο δεξιό μέρος του διαγράμματος, παρουσιάζεται η διαδικασία που ακολουθείται για την εξαγωγή της Set-to-Set Μετρικής D_S . Αρχικά, γίνεται η μέτρηση της απόστασης μεταξύ των προβολών των \mathbf{X}_i και \mathbf{X}_j , μέσω της συνάρτησης g_A και έπειτα η σύνθεσή τους σε μια τιμή, από τη συνάρτηση h_M . Το αποτέλεσμα που προκύπτει από αυτή τη διαδικασία, αποτελεί την απόσταση μεταξύ των δύο αρχικών αναπαραστάσεων $D^\Theta(\mathbf{X}_i, \mathbf{X}_j)$, το οποίο παρουσιάζεται και με τη μορφή συναρτήσεων από τη Σχέση 18:

$$\begin{aligned} D^\Theta(\mathbf{X}_i, \mathbf{X}_j) &= \\ &= D_S(T_S(\mathbf{X}_i), T_S(\mathbf{X}_j)) = D_S(\mathcal{X}_i, \mathcal{X}_j) = \\ &= D_S(\{f_W^1(\mathbf{X}_i), \dots, f_W^m(\mathbf{X}_i)\}, \{f_W^1(\mathbf{X}_j), \dots, f_W^m(\mathbf{X}_j)\}) = \\ &= h_M(g_A^1(f_W^1(\mathbf{X}_i), f_W^1(\mathbf{X}_j)), \dots, g_A^m(f_W^m(\mathbf{X}_i), f_W^m(\mathbf{X}_j))) \end{aligned} \quad (18)$$



Εικόνα 7: Διάγραμμα Ροής του αλγορίθμου PSSSD, στηριζόμενο σε αυτό της δημοσίευσης των Gao et al.¹⁵.

3.3 Μέθοδος Βελτιστοποίησης

Οι τρεις παράμετροι που βελτιστοποιεί ο αλγόριθμος, όπως έχει προαναφερθεί, είναι οι $\mathbf{W} \in St(mp, n)$, $\mathbf{A} \in \mathbb{R}^{m \times 2}$, $\mathbf{M} \in Sym_m^+$ και αναπαρίστανται από το σετ $\Theta = \{\mathbf{W}, \mathbf{A}, \mathbf{M}\}$. Η συνάρτηση κόστους (loss function) ελαχιστοποιείται σταδιακά από τον αλγόριθμο Minibatch Stochastic Gradient Descent, ο οποίος για λόγους ταχύτητας χρησιμοποιεί υποσύνολα των δεδομένων εκπαίδευσης αντί για ολόκληρο το σύνολο. Η συνάρτηση κόστους του αλγορίθμου, δίνεται από τη Σχέση 19, όπως ορίζεται παρακάτω:

$$\begin{aligned} \mathcal{L}(\Theta, \mathcal{S}, \mathcal{D}, Y) &= \\ &= \frac{1}{|\mathcal{S}|} \sum_{i,j \in \mathcal{S}} y_{ij} \cdot \max(D^\Theta(\mathbf{X}_i, \mathbf{X}_j) - \zeta_s, 0)^2 \\ &\quad + \frac{1}{|\mathcal{D}|} \sum_{i,j \in \mathcal{D}} (1 - y_{ij}) \cdot \max(\zeta_d - D^\Theta(\mathbf{X}_i, \mathbf{X}_j), 0)^2 \\ &\quad + \xi \cdot \gamma(\mathbf{M}), \text{ s.t. } \mathbf{M} \in Sym_m^+, \mathbf{W} \in St(mp, n) \end{aligned} \quad (19)$$

Όπου \mathcal{S}, \mathcal{D} , τα σετ των ζευγών των αναπαραστάσεων που ανήκουν σε ίδια και σε διαφορετική κλάση αντίστοιχα, τα οποία επιλέχθηκαν στην τύχη από τον αλγόριθμο Minibatch Stochastic Gradient Descent για την ελαχιστοποίηση της συνάρτησης κόστους. Για τα ζεύγη αυτά, η παράμετρος Y περιέχει τιμές οι οποίες, με βάση τις κλάσεις των αναπαραστάσεων, υποδεικνύουν το αν το κάθε ζεύγος ανήκει ($y_{ij} = 1$) ή όχι ($y_{ij} = 0$) στην ίδια κλάση. Οι σταθερές ζ_s και ζ_d , αποτελούν τις τιμές κατωφλίου με τις οποίες συγκρίνεται η απόσταση που υπολογίζεται μεταξύ δύο όμοιων και ανόμοιων αναπαραστάσεων αντίστοιχα. Οι συναρτήσεις $\max(\cdot, \cdot)$, επιλέγουν τη μεγαλύτερη τιμή μεταξύ των δύο ορισμάτων τους. Εφόσον έχει ακολουθήσει σωστή ρύθμισή τους ($\zeta_s < \zeta_d$), η επιθυμητή τιμή που θα προκύπτει από τη συνάρτηση $\max(\cdot, \cdot)$, είναι το μηδέν, καθώς θα υποδηλώνει ότι η απόσταση μεταξύ αναπαραστάσεων όμοιας κλάσης ήταν μικρότερη από το κατώφλι ζ_s , άρα αυτές ήταν σε κοντινή απόσταση και αντίστοιχα, η απόσταση

¹⁵Z. Gao, Y. Wu, M. Harandi, and Y. Jia, "A Robust Distance Measure for Similarity-Based Classification on the SPD Manifold," IEEE Transactions on Neural Networks and Learning Systems, vol. 31, pp. 3230-3244, 2020.

αναπαραστάσεων ανόμοιας κλάσης, ήταν μεγαλύτερη της ζ_d , άρα οι ανόμοιες είναι απομακρυσμένες. Οι όροι $|\mathcal{S}|$ και $|\mathcal{D}|$ αποτελούν το πλήθος (cardinality) των ζευγών όμοιας και ανόμοιας κλάσης αντίστοιχα, ενώ οι όροι $(1/|\mathcal{S}|)$ και $(1/|\mathcal{D}|)$ χρησιμοποιούνται για την εξισορρόπηση της συνάρτησης στην περίπτωση που δεν υπάρχει ισοβάθμιση μεταξύ όμοιων και ανόμοιων ζευγών.

Για την αποφυγή της υπερπροσαρμογής (overfitting) των παραμέτρων προς βελτιστοποίηση στα δεδομένα εκπαίδευσης και για τη συμπερίληψη προγενέστερης γνώσης της Set-to-Set απόστασης (M_0) [101, 104], δημιουργείται η συνάρτηση $\gamma(\cdot)$ η οποία, υλοποιώντας την Burg Matrix Divergence, την ενσωματώνει στην τρέχουσα κατάσταση της παραμέτρου M ως εξής: $\gamma(M, M_0) = \text{Tr}(MM_0^{-1}) - \log \det(MM_0^{-1}) - m$. Η προγενέστερη πληροφορία που εισάγεται από τη συνάρτηση $\gamma(\cdot)$, ρυθμίζεται από τη πολλαπλασιαστική σταθερά ξ . Λόγω των γεωμετρικών περιορισμών των παραμέτρων W, M , η βελτιστοποίησή τους γίνεται με τη χρήση της Riemannian παραγώγου που ορίζεται στο χώρο στον οποίο ανήκει η κάθε μία, σε αντίθεση με την παράμετρο A για την οποία χρησιμοποιείται η Ευκλείδεια παράγωγος, αφού ανήκει στον Ευκλείδειο χώρο. Συγκεκριμένα, για τις παραμέτρους W, M , θα χρησιμοποιηθούν παράγωγοι που προκύπτουν από το εσωτερικό γινόμενο που ορίζεται για τις Stiefel και SPD Πολλαπλότητες αντίστοιχα. Η διόρθωση των παραμέτρων, γίνεται με την ακόλουθη σειρά: M, A, W , όπως φαίνεται παρακάτω:

3.3.1 Παράγωγος της Συνάρτησης Κόστους ως προς την παράμετρο M

Η μερική παράγωγος της συνάρτησης κόστους ως προς την παράμετρο σύνθεσης M , υπολογίζεται από τη Σχέση 20:

$$\frac{\partial \mathcal{L}}{\partial M} = \frac{1}{|\mathcal{S}|} \sum_{i,j \in \mathcal{S}} d_{ij} \frac{\partial \mathcal{L}}{\partial D_{ij}^{\ominus}} d_{ij}^T + \frac{1}{|\mathcal{D}|} \sum_{i,j \in \mathcal{D}} d_{ij} \frac{\partial \mathcal{L}}{\partial D_{ij}^{\ominus}} d_{ij}^T + \xi \cdot \frac{\partial \gamma(M, M_0)}{\partial M} \quad (20)$$

Όπου $(\partial \mathcal{L} / \partial D_{ij}^{\ominus})$, είναι η μερική παράγωγος της συνάρτησης κόστους ως προς την Point-to-Set και Set-to-Set απόσταση $D_{ij}^{\ominus}(X_i, X_j)$, η οποία υπολογίζεται από τη Σχέση 21:

$$\frac{\partial \mathcal{L}}{\partial D_{ij}^{\ominus}} = 2 \cdot y_{ij} \cdot \max(D_{ij}^{\ominus} - \zeta_s, 0) + 2 \cdot (y_{ij} - 1) \cdot \max(\zeta_d - D_{ij}^{\ominus}, 0) \quad (21)$$

Όπου $(\partial \gamma(M, M_0) / \partial M)$, η μερική παράγωγος της συνάρτησης $\gamma(\cdot, \cdot)$, ως προς την παράμετρο M :

$$\frac{\partial \gamma(M, M_0)}{\partial M} = M_0^{-1} - M^{-1} \quad (22)$$

3.3.2 Παράγωγος της Συνάρτησης Κόστους ως προς την παράμετρο A

Η μερική παράγωγος της συνάρτησης κόστους ως προς το πίνακα A , που περιέχει τα ζεύγη τιμών που χρησιμοποιούνται για τον υπολογισμό της Alpha-Beta Divergence για κάθε k προβολή, υπολογίζεται ξεχωριστά ως προς α_k και ως προς β_k , από τις Σχέσεις 23 και 24 αντίστοιχα:

$$\frac{\partial \mathcal{L}}{\partial \alpha_k} = \frac{1}{|\mathcal{S}|} \sum_{i,j \in \mathcal{S}} \frac{\partial \mathcal{L}}{\partial d_{ij}^k} \cdot \frac{\partial d_{ij}^k}{\partial \alpha_k} + \frac{1}{|\mathcal{D}|} \sum_{i,j \in \mathcal{D}} \frac{\partial \mathcal{L}}{\partial d_{ij}^k} \cdot \frac{\partial d_{ij}^k}{\partial \alpha_k} \quad (23)$$

$$\frac{\partial \mathcal{L}}{\partial \beta_k} = \frac{1}{|\mathcal{S}|} \sum_{i,j \in \mathcal{S}} \frac{\partial \mathcal{L}}{\partial d_{ij}^k} \cdot \frac{\partial d_{ij}^k}{\partial \beta_k} + \frac{1}{|\mathcal{D}|} \sum_{i,j \in \mathcal{D}} \frac{\partial \mathcal{L}}{\partial d_{ij}^k} \cdot \frac{\partial d_{ij}^k}{\partial \beta_k} \quad (24)$$

Όπου $(\partial\mathcal{L}/\partial d_{ij}^k)$, είναι το k -οστό στοιχείο της μερικής παραγώγου της συνάρτησης κόστους \mathcal{L} ως προς το διάνυσμα απόστασης \mathbf{d}_{ij} η οποία ορίζεται παρακάτω:

$$\frac{\partial\mathcal{L}}{\partial\mathbf{d}_{ij}} = \frac{\partial\mathcal{L}}{\partial D_{ij}^{\ominus}} \cdot \frac{\partial D_{ij}^{\ominus}}{\partial\mathbf{d}_{ij}} = \frac{\partial\mathcal{L}}{\partial D_{ij}^{\ominus}} \mathbf{d}_{ij}^T (\mathbf{M}^T + \mathbf{M}) \quad (25)$$

Οι όροι $(\partial d_{ij}^k/\partial\alpha_k)$ και $(\partial d_{ij}^k/\partial\beta_k)$, αποτελούν τις μερικές παραγώγους της απόστασης d_{ij}^k ως προς α και ως προς β αντίστοιχα. Οι όροι αυτοί, αναλύονται από τις Σχέσεις 26 και 27 αντίστοιχα:

$$\frac{\partial d_{ij}^k}{\partial\alpha_k} = \frac{1}{\alpha_k^2\beta_k} \sum_{u=1}^p \left(\frac{\alpha_k (\lambda_{iju}^k)^{\beta_k} - \alpha_k\beta_k (\lambda_{iju}^k)^{-\alpha_k} \log \lambda_{iju}^k}{\alpha_k (\lambda_{iju}^k)^{\beta_k} + \beta_k (\lambda_{iju}^k)^{-\alpha_k}} - \frac{\alpha_k}{\alpha_k + \beta_k} - \log \frac{\alpha_k (\lambda_{iju}^k)^{\beta_k} + \beta_k (\lambda_{iju}^k)^{-\alpha_k}}{\alpha_k + \beta_k} \right) \quad (26)$$

$$\frac{\partial d_{ij}^k}{\partial\beta_k} = \frac{1}{\alpha_k\beta_k^2} \sum_{u=1}^p \left(\frac{\beta_k (\lambda_{iju}^k)^{-\alpha_k} - \alpha_k\beta_k (\lambda_{iju}^k)^{\beta_k} \log \lambda_{iju}^k}{\alpha_k (\lambda_{iju}^k)^{\beta_k} + \beta_k (\lambda_{iju}^k)^{-\alpha_k}} - \frac{\beta_k}{\alpha_k + \beta_k} - \log \frac{\alpha_k (\lambda_{iju}^k)^{\beta_k} + \beta_k (\lambda_{iju}^k)^{-\alpha_k}}{\alpha_k + \beta_k} \right) \quad (27)$$

3.3.3 Παράγωγος της Συνάρτησης Κόστους ως προς την παράμετρο \mathbf{W}

Η μερική παράγωγος της συνάρτησης κόστους ως προς τον κάθε πίνακα προβολής \mathbf{W}_k , υπολογίζεται από τη Σχέση 28:

$$\frac{\partial\mathcal{L}}{\partial\mathbf{W}_k} = \sum_{i,j \in (SU\mathcal{D})} \left(\mathbf{X}_i^T \mathbf{W}_k \frac{\partial\mathcal{L}}{\partial\mathbf{X}_i^k} + \mathbf{X}_i \mathbf{W}_k \frac{\partial\mathcal{L}}{\partial\mathbf{X}_i^k}{}^T + \mathbf{X}_j^T \mathbf{W}_k \frac{\partial\mathcal{L}}{\partial\mathbf{X}_j^k} + \mathbf{X}_j \mathbf{W}_k \frac{\partial\mathcal{L}}{\partial\mathbf{X}_j^k}{}^T \right) \quad (28)$$

Όπου $(\partial\mathcal{L}/\partial\mathbf{X}_i^k)$ και $(\partial\mathcal{L}/\partial\mathbf{X}_j^k)$, η μερική παράγωγος της συνάρτησης κόστους \mathcal{L} ως προς τις k προβολές \mathbf{X}_i^k και \mathbf{X}_j^k των αρχικών αναπαραστάσεων \mathbf{X}_i και \mathbf{X}_j αντίστοιχα. Οι σχέσεις που τις περιγράφουν, είναι οι 29 και 30 αντίστοιχα:

$$\frac{\partial\mathcal{L}}{\partial\mathbf{X}_i^k} = \mathbf{U}_{ij}^k \frac{\partial\mathcal{L}}{\partial\Lambda_{ij}^k} (\mathbf{U}_{ij}^k)^T (\mathbf{X}_i^k)^{-T} \quad (29)$$

$$\frac{\partial\mathcal{L}}{\partial\mathbf{X}_j^k} = (-1) \cdot (\mathbf{X}_j^k)^{-T} (\mathbf{X}_i^k)^T \mathbf{U}_{ij}^k \frac{\partial\mathcal{L}}{\partial\Lambda_{ij}^k} (\mathbf{U}_{ij}^k)^T (\mathbf{X}_j^k)^{-T} \quad (30)$$

Όπου $(\partial\mathcal{L}/\partial\Lambda_{ij}^k)$, η μερική παράγωγος των ιδιοτιμών λ_{iju}^k του πίνακα $\mathbf{X}_i^k (\mathbf{X}_j^k)^{-1}$, που έχει προκύψει από την ανάλυσή του στο διαγώνιο πίνακα ιδιοτιμών (Λ_{ij}^k) και το πίνακα ιδιοδιανυσμάτων (\mathbf{U}_{ij}^k) του. Επιπλέον, τα στοιχεία του διαγώνιου πίνακα Λ_{ij}^k η μερική παράγωγος $(\partial\mathcal{L}/\partial\Lambda_{ij}^k)$, ορίζεται στην παρακάτω σχέση:

$$\frac{\partial\mathcal{L}}{\partial\lambda_{iju}^k} = \frac{\partial\mathcal{L}}{\partial d_{ij}^k} \cdot \frac{\partial d_{ij}^k}{\partial\lambda_{iju}^k} = \frac{\partial\mathcal{L}}{\partial d_{ij}^k} \cdot \frac{1}{\alpha_k\beta_k} \frac{\alpha_k\beta_k (\lambda_{iju}^k)^{\beta_k-1} - \alpha_k\beta_k (\lambda_{iju}^k)^{-\alpha_k-1}}{\alpha_k (\lambda_{iju}^k)^{\beta_k} + \beta_k (\lambda_{iju}^k)^{-\alpha_k}} \quad (31)$$

3.3.4 Ανανέωση των παραμέτρων

Όπως αναφέρθηκε προηγουμένως, η παράμετρος \mathbf{A} ανήκει στον Ευκλείδειο χώρο, ενώ οι \mathbf{W} , \mathbf{M} στο χώρο των Stiefel και των SPD Πολλαπλοτήτων αντίστοιχα. Η ανανέωση της \mathbf{A} γίνεται όπως ορίζει ο Ευκλείδειος Gradient Descent, δηλαδή για το βήμα $t + 1$ η παράμετρος \mathbf{A} λαμβάνει την τιμή:

$$\mathbf{A}_{t+1} = \mathbf{A}_t - \eta \left(\frac{\partial \mathcal{L}}{\partial \mathbf{A}_t} \right) \quad (32)$$

Όπου η , ο ρυθμός μάθησης (learning rate), ο οποίος καθορίζει πόσο μεγάλη αλλαγή θα γίνεται κατά την ανανέωση της μεταβλητής προς διόρθωση. Όσο μεγαλύτερη η τιμή του ρυθμού μάθησης τόσο περισσότερη προγενέστερη πληροφορία αντικαθίσταται κατά τη διόρθωση της παραμέτρου. Σε αντίθεση με την \mathbf{A} , οι παράμετροι \mathbf{W} και \mathbf{M} χρήζουν διαφορετικό χειρισμό, καθώς η ανανέωσή τους πρέπει να γίνει σεβόμενη τη γεωμετρία του χώρου στον οποίο κατατάσσονται. Συγκεκριμένα, για τη διόρθωσή τους πρώτα υπολογίζονται οι Riemannian μερικοί παράγωγοί [100, 105, 106] της συνάρτησης κόστους ως προς την κάθε παράμετρο, οι οποίες βρίσκονται στον εφαπτόμενο χώρο των Stiefel και SPD Πολλαπλοτήτων για τις \mathbf{W} και \mathbf{M} αντίστοιχα. Έπειτα, για βήμα $t + 1$ η κάθε παράμετρος διορθώνεται με βάση την παράγωγο που έχει υπολογιστεί, αφού γίνει η απεικόνισή της πίσω στο χώρο της εκάστοτε Πολλαπλότητας. Οι μερικές παράγωγοι καθώς και οι σχέσεις που προσδιορίζουν τον τρόπο με τον οποίο γίνεται η διόρθωση των παραμέτρων \mathbf{W} και \mathbf{M} , δίνονται από τις Σχέσεις 33 και 34 παρακάτω:

$$\begin{cases} \frac{\partial \mathcal{L}}{\partial \mathbf{W}_t^R} = \frac{\partial \mathcal{L}}{\partial \mathbf{W}_t} - \mathbf{W}_t \frac{1}{2} \left(\mathbf{W}_t^T \frac{\partial \mathcal{L}}{\partial \mathbf{W}_t} + \frac{\partial \mathcal{L}}{\partial \mathbf{W}_t}^T \mathbf{W}_t \right) \\ \mathbf{W}_{t+1} = \mathfrak{q} \left(\mathbf{W}_t - \eta \frac{\partial \mathcal{L}}{\partial \mathbf{W}_t^R} \right) \end{cases} \quad (33)$$

$$\begin{cases} \frac{\partial \mathcal{L}}{\partial \mathbf{M}_t^R} = \mathbf{M}_t \frac{1}{2} \left(\frac{\partial \mathcal{L}}{\partial \mathbf{M}_t} + \frac{\partial \mathcal{L}}{\partial \mathbf{M}_t}^T \right) \mathbf{M}_t \\ \mathbf{M}_{t+1} = \mathbf{M}_t^{\frac{1}{2}} \expm \left(-\eta \mathbf{M}_t^{-\frac{1}{2}} \frac{\partial \mathcal{L}}{\partial \mathbf{M}_t^R} \mathbf{M}_t^{-\frac{1}{2}} \right) \mathbf{M}_t^{\frac{1}{2}} \end{cases} \quad (34)$$

Όπου $(\partial \mathcal{L} / \partial \mathbf{W}_t^R)$ και $(\partial \mathcal{L} / \partial \mathbf{M}_t^R)$, είναι οι Riemannian μερικές παράγωγοι που υπολογίζονται για τις παραμέτρους \mathbf{W} και \mathbf{M} αντίστοιχα για βήμα t , δηλαδή με τις τρέχουσες τιμές των παραμέτρων. Το πως προκύπτουν οι παράγωγοι αυτοί, αναλύεται στο βιβλίο των Absil *et al.* [100]. Οι νέες τιμές των \mathbf{W} και \mathbf{M} , για βήμα $t + 1$, είναι οι \mathbf{W}_{t+1} και \mathbf{M}_{t+1} . Οι συναρτήσεις $\mathfrak{q}(\cdot)$ και $\expm(\cdot)$, αποτελούν τις **συναρτήσεις ανάκλησης** (retraction operators), οι οποίες γενικότερα, μεταφέρουν ένα σημείο από τον εφαπτόμενο χώρο της Πολλαπλότητας στην οποία ανήκει, πίσω στην επιφάνεια της Πολλαπλότητας. Οι συγκεκριμένες συναρτήσεις, $\mathfrak{q}(\cdot)$ και $\expm(\cdot)$, μεταφέρουν τους πίνακες \mathbf{W} και \mathbf{M} πίσω στο χώρο των Stiefel και SPD Πολλαπλοτήτων αντίστοιχα. Η συνάρτηση $\mathfrak{q}(\cdot)$ εκφράζει τον ορθογώνιο πίνακα \mathbf{Q} που προκύπτει από την Αποσύνθεσης QR (QR Decomposition) του πίνακα $\mathbf{W} = \mathbf{Q}\mathbf{R}$, με $\mathbf{W} \in \mathbb{R}^{n \times p}$, $\mathbf{Q} \in \mathbb{R}^{n \times p}$ και άνω τριγωνικό πίνακα $\mathbf{R} \in \mathbb{R}^{p \times p}$, ενώ η συνάρτηση $\expm(\cdot)$ αποτελεί την Εκθετική Απεικόνιση (Exponential Map), η οποία συμπίπτει με την εκθετική συνάρτηση πινάκων.

Η ανανέωση των παραμέτρων σε κάθε βήμα t , μεταφράζεται στη μετακίνηση των πινάκων ως σημεία πάνω στο χώρο που ανήκουν. Κατά την εφαρμογή των συναρτήσεων ανάκλησης (retraction operators), η μετακίνηση των σημείων αυτών εγγυάται ότι θα γίνει προς τη σωστή κατεύθυνση, η οποία είναι αυτή που προσδιορίζεται από τις μερικές παραγώγους κατά τη διόρθωση των παραμέτρων, αλλά και ότι το σημείο αυτό θα βρίσκεται στο πάνω στο χώρο της Πολλαπλότητας στην οποία ανήκει [77]. Η χρήση των Riemannian μερικών παραγώγων για

την υλοποίηση του αλγορίθμου Minibatch Stochastic Gradient Descend για τη βελτιστοποίηση παραμέτρων που ανήκουν σε Riemannian Πολλαπλότητες είναι αποτελεσματική, καθώς, όπως αποδεικνύεται στη δημοσίευση του Silvère Bonnabel [107], οι παράμετροι αυτές συγκλίνουν σε μια ελάχιστη τιμή όπως γίνεται αντίστοιχα από την εφαρμογή του ίδιου αλγορίθμου σε Ευκλείδειες ποσότητες.

3.4 Υπολογιστική Πολυπλοκότητα

Η Υπολογιστική Πολυπλοκότητα (Computational Complexity) ενός αλγορίθμου, συμβολίζεται ως $\mathcal{O}(\cdot)$ και προσδιορίζει το ποσό των πόρων που απαιτούνται για την εκτέλεσή του. Ένας συνήθης τρόπος περιγραφής της υπολογιστικής πολυπλοκότητας ενός αλγορίθμου, είναι μέσω της μέτρησης του ποσού των αριθμητικών πράξεων κινητής υποδιαστολής (floating point operations, FLOPS) που εκτελεί ανά δευτερόλεπτο, των λεγόμενων flops/s ή απλά flops. Για τις αναλύσεις που γίνονται παρακάτω, ισχύει ότι n είναι οι διαστάσεις των αρχικών SPD αναπαραστάσεων, m και p είναι το πλήθος και οι διαστάσεις των προβολών που προκύπτουν από τις αρχικές αναπαραστάσεις αντίστοιχα και N είναι το πλήθος των δειγμάτων εκπαίδευσης.

3.4.1 Υπολογιστική Πολυπλοκότητα της Συνάρτησης Κόστους

Για την προσέγγιση του υπολογιστικού κόστους της αντικειμενικής συνάρτησης (Σχέση 19), απαιτείται ο προσδιορισμός του κόστους της Point-to-Set και Set-to-Set Απόστασης $D^\ominus(\mathbf{X}_i, \mathbf{X}_j)$ μεταξύ δύο αρχικών αναπαραστάσεων $\mathbf{X}_i, \mathbf{X}_j \in Sym_n^+$, ο οποίος αποτελείται από τον Point-to-Set Μετασχηματισμό T_S και τη Set-to-Set Απόσταση D_S . Η υπολογιστική πολυπλοκότητα του Point-to-Set Μετασχηματισμού T_S , όπως περιγράφεται από τη Σχέση 6, ισούται με $\mathcal{O}(mn^2p + mnp^2)$ flops. Για τον υπολογισμό της Set-to-Set απόστασης D_S , είναι περαιτέρω απαραίτητος ο υπολογισμός της αντιστροφής πινάκων, εφόσον χρησιμοποιείται κατά τον υπολογισμό της απόστασης μεταξύ δύο προβολών από τη Σχέση 9. Η αντιστροφή ενός πίνακα, διαστάσεων $p \times p$, με βάση τον αλγόριθμο Coppersmith–Winograd, απαιτεί $\mathcal{O}(p^{2.376})$ flops και η Set-to-Set Απόσταση D_S που περιγράφεται από τη Σχέση 10, κοστίζει $\mathcal{O}(3mp^3 + mp^{2.376} + m^2 + m)$ flops. Επομένως, το συνολικό υπολογιστικό κόστος της συνάρτησης κόστους, όπως αυτή περιγράφεται από τη Σχέση 19, προσδιορίζεται ότι είναι $\mathcal{O}(N^2 + (8/3)m^3)$ flops.

3.4.2 Υπολογιστική Πολυπλοκότητα των μερικών παραγώγων

Κατά τη διαδικασία της βελτιστοποίησης των παραμέτρων \mathbf{M} , \mathbf{A} και \mathbf{W} , μέσω του αλγορίθμου Minibatch Stochastic Gradient Descend, υπολογίζονται οι μερικές παράγωγοι αυτών των ποσοτήτων. Το κόστος των υπολογισμών της μερικής παραγώγου της συνάρτησης κόστους \mathcal{L} ως προς την παράμετρο \mathbf{M} όπως αυτή ορίζεται από τη Σχέση 20, είναι $\mathcal{O}(N^2m + N^2m^2 + 2m^{2.376})$ flops. Για τη βελτιστοποίηση της παραμέτρου \mathbf{A} , υπολογίζεται η μερική παράγωγος της συνάρτησης κόστους \mathcal{L} , ως προς τις τιμές α και β ξεχωριστά, από τις Σχέσεις 23 και 24 αντίστοιχα. Το υπολογιστικό κόστος των παραγώγων αυτών, είναι $\mathcal{O}(N^2m^2 + 2N^2m^3 + N^2m^2p)$ και $\mathcal{O}(N^2m^2p)$ flops αντίστοιχα. Τέλος, η μερική παράγωγος της συνάρτησης κόστους \mathcal{L} ως προς την παράμετρο \mathbf{W} , που προσδιορίζεται με βάση τη Σχέση 28, κοστίζει $\mathcal{O}(m^2 + 8mp^3 + 2Nmn^2p + 2Nmn^2p^2)$ flops.

3.4.3 Υπολογιστική Πολυπλοκότητα της ανανέωσης των Riemannian παραμέτρων

Για την ανανέωση των παραμέτρων \mathbf{M} και \mathbf{W} , όπως γίνεται από τις Σχέσεις 34 και 33, το υπολογιστικό κόστος ισούται με $\mathcal{O}(3m^3)$ και $\mathcal{O}(4nm^2p^2 + 11m^3p^3)$ flops αντίστοιχα.

4 Γενικά περί Βιομετρικής Ταυτοποίησης

Η **βιομετρία** (biometrics) είναι ο κλάδος της στατιστικής που αναπτύσσει μαθηματικά εργαλεία, με σκοπό την ποσοτικοποίηση βιολογικών στοιχείων σε ανθρώπους και άλλους ζωντανούς οργανισμούς [108]. Ενώ τα αντικείμενα εφαρμογών του κλάδου ποικίλουν, εκείνα που αφορούν την παρούσα διπλωματική εργασία σχετίζονται με τη βιομετρική ταυτοποίηση, δηλαδή τη χρήση των ανθρώπινων χαρακτηριστικών και ιδιαίτερα εκείνων που είναι ξεχωριστά για κάθε άτομο. Παραδείγματα τέτοιων χαρακτηριστικών αποτελούν οι ίριδες των ματιών, τα δακτυλικά αποτυπώματα, καθώς και χαρακτηριστικά που αφορούν τη συμπεριφορά ενός ανθρώπου, όπως η βόδιση, η φωνή και η υπογραφή του. Ακριβώς επειδή τα χαρακτηριστικά αυτά είναι ξεχωριστά για κάθε άνθρωπο, πολλές είναι οι εφαρμογές που βασίζονται στη χρήση τους, για την ταυτοποίηση ατόμων. Λόγω της αυξημένης ασφάλειας που προσφέρει αυτό το μέσο αναγνώρισης, υιοθετείται από τις τράπεζες για την παροχή πρόσβασης σε ΑΤΜ και για διαδικτυακές τραπεζικές συναλλαγές (online banking), αλλά και από πολλούς κατασκευαστές προϊόντων πληροφορικής, που ενσωματώνουν συστήματα βιομετρικής αναγνώρισης σε laptops και smartphones.

4.1 Κλασικές μέθοδοι Ταυτοποίησης της Ανθρώπινης Παρουσίας

Η χρήση των βιομετρικών χαρακτηριστικών ενός ατόμου για την αναγνώρισή του, πλεονεκτεί έναντι των μέσων που βασίζονται στην **απομνημόνευση** ή τη χρήση **φυσικών μέσων** [108]. Ένα από τα πιο διαδεδομένα μέσα απόκτησης πρόσβασης που βασίζονται στην απομνημόνευση, είναι η χρήση **κωδικών πρόσβασης** (passwords). Η μέθοδος αυτή συναντάται καθημερινά σε πολλές υπηρεσίες που απαιτούν ταυτοποίηση και λειτουργούν μέσω του διαδικτύου, αφού ένας χρήστης μπορεί να εισέλθει με οποιοδήποτε μέσο συνδεδεμένο στο διαδίκτυο χωρίς να απαιτείται κάποιος εξειδικευμένος εξοπλισμός (hardware). Σε αυτή την περίπτωση, η ισχύς του κωδικού πρόσβασης του χρήστη έχει καθοριστικό ρόλο στην ασφάλεια των δεδομένων του. Ένας ισχυρός κωδικός πρέπει να συμπεριλαμβάνει μια πληθώρα χαρακτήρων, όπως γράμματα, αριθμούς και σύμβολα, καθώς και να μη σχηματίζει λέξεις και αριθμούς που εύκολα μαντεύονται, παραδείγματος χάριν ονόματα και ημερομηνίες. Το γεγονός αυτό, συνήθως καθιστά τους κωδικούς πρόσβασης δύσκολους στην απομνημόνευση, όπως και επιρρεπείς στο να ξεχαστούν από το χρήστη.

Τα φυσικά μέσα ταυτοποίησης, επειδή απαιτούν εξειδικευμένο εξοπλισμό για τη χρήση τους, συνήθως συναντώνται σε υπηρεσίες που παρέχουν φυσική πρόσβαση σε μια φυλασσόμενη υποδομή, αλλά και σε άλλες μονάδες, όπως ΑΤΜ, που παρέχουν απομακρυσμένη πρόσβαση στο τραπεζικό λογαριασμό ενός ατόμου. Ένα από τα συνηθέστερα φυσικά μέσα ταυτοποίησης, αποτελούν οι **μαγνητικές κάρτες** (magnetic stripe cards), οι οποίες φυλάσσουν την πληροφορία που ταυτοποιεί το χρήστη στις μαγνητικές τους ταινίες. Οι μαγνητικές κάρτες, πέρα από το γεγονός ότι κινδυνεύουν να χαθούν ή να καταστραφούν, είναι επιρρεπείς στην αντιγραφή των δεδομένων τους, αφού η πληροφορία που φυλάσσουν είναι απλά κωδικοποιημένη (encoded) και όχι κρυπτογραφημένη (encrypted). Μια αναβάθμιση των μαγνητικών καρτών αποτελούν οι **έξυπνες κάρτες** (smart cards), οι οποίες δεν είναι επιρρεπείς στην αντιγραφή της πληροφορίας τους, διότι έχουν τη δυνατότητα της κρυπτογράφησης της.

Πέρα από τα προαναφερόμενα μειονεκτήματά τους, οι κωδικοί πρόσβασης όπως και οι μαγνητικές και έξυπνες κάρτες, είναι επιρρεπείς στην κλοπή από επιτήδειους που στοχεύουν στην απόκτηση πρόσβασης εκ μέρους του εξουσιοδοτημένου χρήστη, με σκοπό την επίθεση στον ίδιο ή την εν λόγω υπηρεσία. Ενώ υπάρχουν τεχνικές επαλήθευσης της ταυτότητας του χρήστη, όπως ο έλεγχος ταυτότητας δύο παραγόντων (Two-Factor Authentication, 2FA), έχουν

και εκείνες τα μειονεκτήματά τους. Η βιομετρική αναγνώριση από την άλλη, είναι ασφαλέστερη, αφού τα βιομετρικά χαρακτηριστικά ενός ανθρώπου δεν μπορούν να ξεχαστούν από τον ίδιο, ενώ δύσκολα μιμούνται από επιτήδειους, καθώς είναι περίπλοκα και ξεχωριστά.

4.2 Βιομετρικές μέθοδοι Ταυτοποίησης

Τα βιομετρικά χαρακτηριστικά τα οποία χρησιμοποιούνται για την ταυτοποίηση ενός ανθρώπου, χωρίζονται στα λεγόμενα **στατικά** (static) και **δυναμικά** (dynamic) χαρακτηριστικά [108]. Τα στατικά είναι εκείνα που αφορούν τα βιολογικά γνωρίσματα ενός ανθρώπου, τα οποία δε μεταβάλλονται, όπως τη δομή της ίριδας των ματιών του, τα δακτυλικά αποτυπώματα του, το σχήμα των χεριών του, και άλλα. Τα δυναμικά από την άλλη, είναι τα χαρακτηριστικά εκείνα που αναπτύσσει ο άνθρωπος κατά τη διάρκεια της ζωής του. Κάποια από αυτά είναι ο τρόπος με τον οποίο περπατά, η χροιά της φωνής του και ο γραφικός του χαρακτήρας.

4.2.1 Στατικά Βιομετρικά Χαρακτηριστικά

Μια από τις πρώτες και πιο διαδεδομένες μεθόδους βιομετρικής αναγνώρισης ανθρώπων, είναι μέσω του **δακτυλικού αποτυπώματος** (fingerprint recognition) [108, 109]. Το μέσο αυτό είναι ιδιαίτερα διαδεδομένο σε προϊόντα πληροφορικής προσωπικής χρήσης, όπως σε κινητά και laptops, καθώς η ενσωμάτωσή του σε τέτοιες συσκευές, προσφέρει ασφάλεια στα δεδομένα του χρήστη, όπως και διευκόλυνση στην πρόσβασή του σε αυτή, αφού η αναγνώρισή του θα πραγματοποιηθεί σε λιγότερο χρονικό διάστημα, συγκριτικά με το χρόνο που χρειάζεται ο χρήστης για να πληκτρολογήσει το κωδικό πρόσβασής του. Η ταυτοποίηση ανθρώπων μέσω των δακτυλικών τους αποτυπωμάτων αρχικά βρήκε εφαρμογή στο χώρο της εγκληματολογίας, όπου συνεχίζει και σήμερα να χρησιμοποιείται ευρέως για την ταυτοποίηση υπόπτων. Αυτή η μέθοδος ταυτοποίησης, αποτελεί μια από τις πιο εύρωστες, καθώς, όντας στατικό βιομετρικό χαρακτηριστικό, δε μεταβάλλεται κατά τη διάρκεια της ζωής του ανθρώπου, πέρα από ειδικές παθολογικές περιπτώσεις.

Τα χαρακτηριστικά που προσδίδουν την ιδιαιτερότητα στα δακτυλικά αποτυπώματα ονομάζονται μικρολεπτομέρειες (minutiae) και αποτελούν τις κορυφές και τις κοιλάδες του δέρματος, καθώς και τις απολήξεις και τις διακλαδώσεις των κορυφών που σχηματίζονται στην την επιδερμίδα του δακτύλου (Εικόνα 8a). Τα χαρακτηριστικά αυτά διαφέρουν για κάθε δάκτυλο, ακόμα και αυτών που ανήκουν στον ίδιο άνθρωπο, ενώ είναι τόσο ξεχωριστά, όπου σύμφωνα με το Αμερικανικό Γραφείο Ομοσπονδιακών Ερευνών (FBI), οι πιθανότητες δύο άνθρωποι να έχουν ακριβώς τα ίδια δακτυλικά αποτυπώματα είναι μια στις 64,000,000,000. Ενώ υπάρχουν πολλές μέθοδοι αντιστοίχισης δακτυλικών αποτυπωμάτων, οι πιο διαδεδομένες είναι εκείνες που χρησιμοποιούν την πληροφορία που παρέχεται από τις μικρολεπτομέρειες. Η σύλληψη της δομής των μικρολεπτομερειών μπορεί να γίνει μέσω διάφορων αισθητήρων, των οποίων η λειτουργία βασίζεται σε διαφορετικά φαινόμενα. Για την αναγνώριση δακτυλικού αποτυπώματος με τη χρήση οπτικού αισθητήρα, απαιτείται η προεπεξεργασία (preprocessing) της εικόνας που λαμβάνεται, ώστε να γίνει ανάδειξη της ωφέλιμης πληροφορίας. Η προεπεξεργασία αυτή, περιλαμβάνει διαδικασίες όπως την ευθυγράμμιση και την κατάτμηση της εικόνας, τη βελτίωση της ποιότητας της εικόνας και τέλος, την ανάδειξη και την εξαγωγή των μικρολεπτομερειών. Τα προβλήματα που αναδύονται κατά τη βιομετρική αναγνώριση ανθρώπων μέσω των δακτυλικών τους αποτυπωμάτων αφορούν την ποιότητα της εικόνας του δακτύλου που λαμβάνεται, η οποία εξαρτάται από το πόσο καθαρή και στεγνή είναι η επιδερμίδα του δακτύλου, καθώς και η επιφάνεια του αισθητήρα. Η διαδικασία ταυτοποίησης γίνεται πιο σύνθετη όταν η εικόνα είναι αλλοιωμένη ή πλήττεται από θόρυβο.

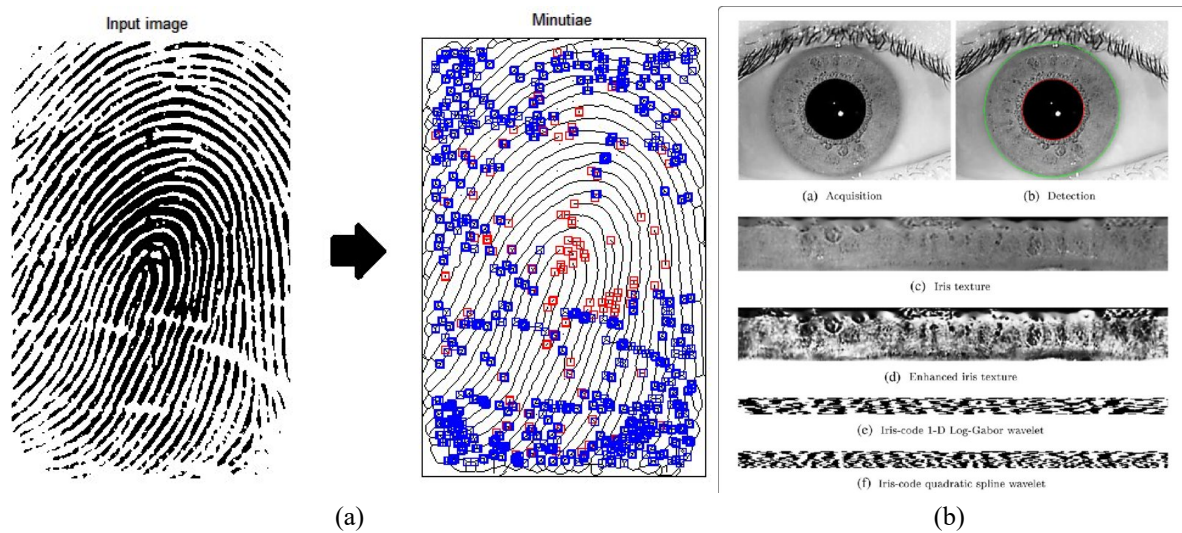
Μια άλλη μέθοδος βιομετρικής ταυτοποίησης, είναι μέσω της αναγνώρισης της **ίριδας** του οφθαλμού των ανθρώπων (iris recognition) [108]. Η ίριδα αποτελεί τη χρωματιστή δισκοειδή μεμβράνη του ματιού, που περιβάλλει την κόρη. Οι ίριδες των ματιών, όπως συμβαίνει και με τα δακτυλικά αποτυπώματα, είναι από τα χαρακτηριστικά τα οποία αναπτύσσονται κατά την κύηση του ανθρώπου, ενώ όπως υποστηρίζεται από πολλαπλές μετρήσεις που έχουν διεξαχθεί, μεταβάλλονται ελάχιστα μετά το πέρας της παιδικής ηλικίας, δεδομένου της απουσίας κάποιου τραυματισμού ή άλλης πάθησης. Το είδος αυτό της βιομετρικής ταυτοποίησης, σε αντίθεση με τη χρήση του δακτυλικού αποτυπώματος, δεν απαιτεί ο χρήστης του να έρθει σε επαφή με κάποιο εξοπλισμό, αφού η πληροφορία για τη δομή της ίριδας λαμβάνεται ως φωτογραφία από το σύστημα αναγνώρισης (Εικόνα 8b). Οι προκλήσεις που προκύπτουν από αυτό το σύστημα βιομετρικής αναγνώρισης, όπως και πριν, έχουν να κάνουν με την ποιότητα της φωτογραφίας που λαμβάνεται, το θόρυβο, ο οποίος μπορεί να είναι αντανακλάσεις, ακόμα και αντικείμενα που καλύπτουν την ίριδα, όπως βλεφαρίδες και τέλος, το γεγονός ότι ο χρήστης δε βρίσκεται σε ευθυγράμμιση με τη συσκευή λήψης της εικόνας της ίριδας, με αποτέλεσμα να αποκόβεται κομμάτι της στην τελική φωτογραφία.

Μία ακόμα μέθοδος που βασίζεται στη χρήση οπτικών αισθητήρων για τη λήψη των δεδομένων προς αναγνώριση, αποτελεί η **ταυτοποίηση μέσω προσώπου** (face recognition) [108]. Η τεχνική αυτή έχει γίνει αρκετά διαδεδομένη τα τελευταία χρόνια, λόγω της εξέλιξης των πεδίων της Επεξεργασίας Εικόνας και της Υπολογιστικής Όρασης, καθώς και της ύπαρξης, φτηνού πλέον ηλεκτρονικού υλικού, το οποίο χρησιμοποιείται κατά κόρον από το μέσο άνθρωπο. Σε σύγκριση με τις προηγούμενες δύο μεθόδους βιομετρικής αναγνώρισης, προσφέρει δύο σημαντικά πλεονεκτήματα. Το πρώτο, αποτελεί το γεγονός ότι για την υλοποίησή της μπορεί να χρησιμοποιηθεί ένας οπτικός αισθητήρας που λειτουργεί στο ορατό φάσμα, σε αντίθεση με τις παραπάνω δύο τεχνικές που απαιτούν εξειδικευμένο εξοπλισμό, ο οποίος χρησιμοποιείται μόνο για αυτό τον σκοπό. Η αναγνώριση ενός προσώπου μπορεί να γίνει είτε από μια μεμονωμένη εικόνα, είτε από μια σειρά εικόνων (ροή βίντεο). Το δεύτερο πλεονεκτήματα, αφορά τη διευκόλυνση του χρήστη, καθώς δεν υπάρχει η απαίτηση ο χρήστης να βρίσκεται ευθυγραμμισμένος με τη συσκευή λήψης της φωτογραφίας σε βαθμό όμοιο με αυτό που απαιτεί ένας αισθητήρας αναγνώρισης ίριδας, καθώς το πρόσωπο, για το μέγεθός του, δεν είναι τόσο πυκνό σε πληροφορία όπως είναι η ίριδα. Τα παραπάνω δύο αποτελούν τους κύριους λόγους για τους οποίους η αναγνώριση μέσω προσώπου έχει υιοθετηθεί σε τόσο μεγάλο βαθμό τα τελευταία χρόνια, ενώ σε πολλές περιπτώσεις έχει αντικαταστήσει τις παραπάνω δύο μεθόδους βιομετρικής ταυτοποίησης. Στη σημερινή εποχή, πολλοί κατασκευαστές συσκευών πληροφορικής προσωπικής χρήσης, όπως αυτών των laptops και των έξυπνων κινητών τηλεφώνων, επιλέγουν την υλοποίηση αυτής της μεθόδου βιομετρικής αναγνώρισης, λόγω του γεγονότος ότι παρέχει ασφάλεια και ευκολία στην υλοποίησή της, ενώ παράλληλα αποτελεί μια λύση ανταγωνιστική από άποψη κόστους, καθώς ο ίδιος οπτικός αισθητήρας μπορεί να χρησιμοποιηθεί και για άλλους σκοπούς, όπως τη λήψη φωτογραφιών και βίντεο.

4.2.2 Δυναμικά Βιομετρικά Χαρακτηριστικά

Μια μέθοδος ταυτοποίησης που βασίζεται σε ένα δυναμικό βιομετρικό χαρακτηριστικό, αποτελεί η αναγνώριση ενός ανθρώπου μέσω της **βάδισης** του (gait identification) [108]. Όπως έχει μελετηθεί στο παρελθόν, η βάδιση ενός ανθρώπου είναι ξεχωριστή για το κάθε ένα, ενώ παράλληλα αναπαράγεται από τον ίδιο με ελάχιστες διαφορές κάθε φορά. Ένα πλεονέκτημα που έχει αυτή η μέθοδος, βασίζεται στο ότι το ανθρώπινο σώμα είναι μεγάλο σε μέγεθος και η κίνηση που εκτελεί καθώς βαδίζει γίνεται εύκολα αντιληπτή. Το γεγονός αυτό, δίνει τη δυνατότητα στο να γίνει αναγνώριση κάποιου από μακρινή απόσταση (Εικόνα 9a), ακόμα και αν

η σειρά των εικόνων που παρέχονται δεν είναι υψηλής ανάλυσης, ενώ δεν απαιτείται να είναι ευδιάκριτα άλλα χαρακτηριστικά του, όπως το πρόσωπο του, το χρώμα του δέρματός του, ή τα ρούχα που φοράει. Εφόσον είναι αντιληπτή η σιλουέτα του, μπορεί να καταγραφεί η κίνηση που εκτελείται κατά τη βάδισή του και στη συνέχεια να πραγματοποιηθεί η ταυτοποίησή του. Λόγω του γεγονότος ότι αυτή η μέθοδος δεν απαιτεί εγγύτητα με τον άνθρωπο προς αναγνώριση, βρίσκει εφαρμογή από τις αστυνομικές αρχές για την αναγνώριση υπόπτων με βάση τη βάδισή τους, εφόσον αυτή είναι γνωστή.



Εικόνα 8: (a) Εφαρμογή ανάλυσης ανθρώπινου δακτυλικού αποτυπώματος, με τη χρήση του MATLAB¹⁶, (b) Ανάλυση ίριδας και αναπαράσταση διαφορετικών χαρακτηριστικών της¹⁷.

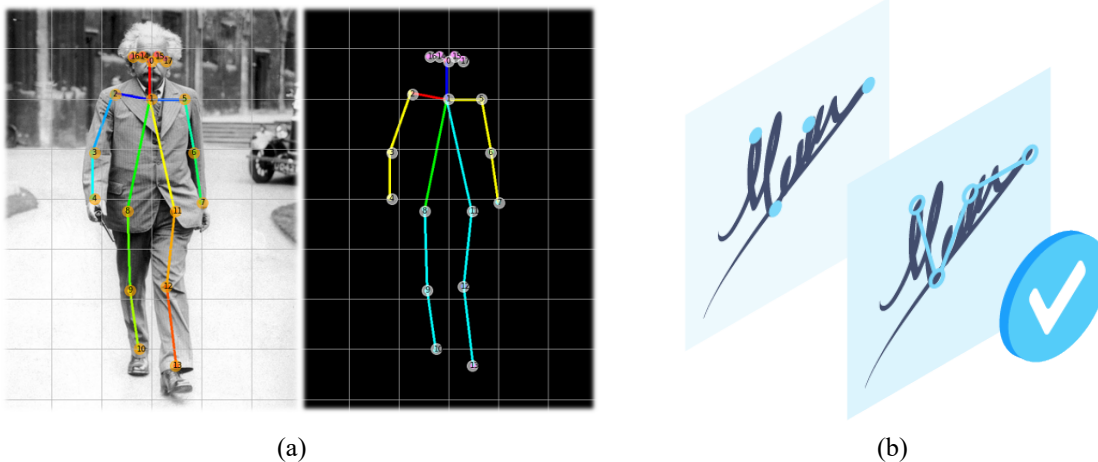
Η φωνή του ανθρώπου αποτελεί ακόμα ένα γνώρισμά του, το οποίο αναπτύσσει μέχρι την αρχή της ενήλικης ζωής του. Η βιομετρική αναγνώριση ενός ατόμου από τη **χροιά της φωνής** του (voice recognition) [108], χρησιμοποιείται για απλές εφαρμογές, όπως την αυτόματη αναγνώριση ενός ομιλητή σε μια συνεδρίαση, ενώ παράλληλα αποτελεί ένα ασφαλές μέσο ταυτοποίησης για εφαρμογές όπως αυτών των τραπεζικών συναλλαγών. Η υλοποίηση αυτού του είδους βιομετρικής αναγνώρισης γίνεται εύκολα με τη χρήση λογισμικού, το οποίο μαθαίνει τη χροιά της φωνής ενός ανθρώπου από παρεχόμενα δείγματα ομιλίας του. Η φωνητική αναγνώριση ενός ανθρώπου γίνεται δυσκολότερη όταν υπάρχει θόρυβος στο υπόβαθρο του ομιλητή ή άλλα στοιχεία που μειώνουν την ποιότητα του ήχου που καταγράφεται.

Ένα ακόμα δυναμικό βιομετρικό χαρακτηριστικό που χρησιμοποιείται για την ασφαλή ταυτοποίηση ενός χρήστη αποτελεί η **υπογραφή** του (signature verification) [108]. Η υπογραφή ενός ανθρώπου είναι συνήθως ένα καλλιτεχνικό, χειρόγραφο σχέδιο που αναπτύσσει, το οποίο συχνά είναι εμπνευσμένο από το όνομα, ένα ψευδώνυμο ή τα αρχικά του ονόματός του. Η ύπαρξη της υπογραφής του, σε ένα έγγραφο για παράδειγμα, χαρακτηρίζει ότι ο ίδιος γνωρίζει για την ύπαρξη και το περιεχόμενό του, ενώ δηλώνει μια συμφωνία ή συμμόρφωση με αυτό. Η υπογραφή έχει νομική βαρύτητα και αποτελεί ένα από τα συνηθέστερα μέσα ταυτοποίησης που χρησιμοποιούνται σε επίσημα έγγραφα από τις τράπεζες και άλλους επίσημους φορείς. Η

¹⁶Dr, Amjad & Hindi, Amjad & Al Qadi, Ziad. (2020). Analysis of Procedures used to build an Optimal Fingerprint Recognition System. 21-37.

¹⁷Rathgeb, Christian & Tams, Benjamin & Wagner, Johannes & Busch, Christoph. (2016). Unlinkable improved multi-biometric iris fuzzy vault. EURASIP Journal on Information Security. 2016. 10.1186/s13635-016-0049-9.

χρήση των υπογραφών ως μέσο ταυτοποίησης, πραγματοποιείται εδώ και αιώνες ενώ δε φαίνεται να εκλείπει, καθώς αποτελεί ένα από τα πιο εύκολα και διαδεδομένα, προς τον καθημερινό άνθρωπο, μέσα ταυτοποίησης. Περαιτέρω, τις τελευταίες δεκαετίες έχουν αναπτυχθεί ηλεκτρονικά συστήματα που τις ψηφιοποιούν, εντάσσοντάς τες με αυτό τον τρόπο στο τομέα της πληροφορικής, συντηρώντας έτσι τη χρήση τους. Η ταυτοποίηση ενός εξουσιοδοτημένου χρήστη μέσω τις υπογραφής του (Εικόνα 9b) γίνεται με δύο κυρίως τρόπους, οι οποίοι θα αναλυθούν στη συνέχεια. Αυτό όμως που ισχύει και για τις δύο μεθόδους ταυτοποίησης, σε διαφορετικό βαθμό για την κάθε μια, είναι το γεγονός ότι έχει σημασία η πολυπλοκότητα της υπογραφής του χρήστη, καθώς, όπως και με τους κωδικούς πρόσβασης, όσο πιο περίπλοκη η υπογραφή, τόσο δυσκολότερη είναι στη μίμησή της.



Εικόνα 9: (a) Εφαρμογή ανάλυσης ανθρώπινου βηματισμού¹⁸, (b) Ενδεικτική λειτουργία συστήματος αναγνώρισης χειρόγραφης υπογραφής¹⁹.

4.3 Προκλήσεις των Συστημάτων Βιομετρικής Αναγνώρισης

Όπως κάθε σύστημα ασφαλείας, έτσι και τα συστήματα βιομετρικής αναγνώρισης, ενώ θεωρούνται ασφαλέστερα, συνοδεύονται από κάποιες προκλήσεις. Όπως αναφέρθηκαν προηγουμένως, για κάθε βιομετρική μέθοδο ταυτοποίησης δίνεται ιδιαίτερη βαρύτητα στην ποιότητα των δεδομένων προς ταυτοποίηση. Εξωτερικοί παράγοντες όπως ο θόρυβος, αποτελούν πρόκληση κατά τη διαδικασία της ταυτοποίησης και για αυτό το λόγο δίνεται έμφαση στη μείωση της επίδρασής τους. Πέρα από την ποιότητα των δεδομένων όμως απαραίτητο είναι να δοθεί έμφαση και σε παράγοντες που αφορούν την ευαλωτότητα του ίδιου του συστήματος ταυτοποίησης, που υφίσταται λόγω της φύσεως των βιομετρικών χαρακτηριστικών του ανθρώπου.

4.3.1 Ασφάλεια

Ένα πρόβλημα που αναδύεται στα συστήματα βιομετρικής αναγνώρισης είναι το γεγονός ότι κανένας δεν μπορεί να αλλάξει τα βιομετρικά του χαρακτηριστικά. Περαιτέρω, τα βιομετρικά στοιχεία ενός ανθρώπου μπορούν να κλαπούν γιατί πολλές φορές αποτελούν δημόσια

¹⁸Deep Learning-Based Pose Estimation - dzone.com

¹⁹Progressoft - [Intelligent Signature Recognition](https://progressoft.com)

πληροφορία. Για παράδειγμα, το πρόσωπο, ο βηματισμός και η φωνή ενός ανθρώπου, που μπορούν να καταγραφούν, όπως και η υπογραφή του, που μπορεί να ληφθεί από έγγραφα τα οποία ο ίδιος έχει υπογράψει. Επομένως, εάν ένας κακόβουλος θελήσει να εισέλθει σε ένα ασφαλές σύστημα που χρησιμοποιεί βιομετρική ταυτοποίηση υποδύομενος έναν εξουσιοδοτημένο χρήστη, θα μπορέσει να το κάνει εάν κλέψει τα βιομετρικά στοιχεία του. Το πρόβλημα αυτό απουσιάζει στα συστήματα που βασίζονται στη χρήση κλασικών μεθόδων ταυτοποίησης, όπως των κωδικών πρόσβασης και των μαγνητικών/έξυπνων καρτών, στα οποία ο χρήστης, σε περίπτωση που αποδειχθεί ή υπάρξει η υποψία πως τα στοιχεία εισόδου του έχουν διαρρεύσει, μπορεί εύκολα να τα αλλάξει.

4.3.2 Πλαστογράφιση

Η βιομετρική αναγνώριση βασίζεται στο γεγονός ότι κάθε άνθρωπος έχει στην κατοχή του ένα μοναδικό χαρακτηριστικό, το οποίο δε μαθαίνεται ούτε αντιγράφεται από κανέναν άλλο. Επομένως, εάν κάποιος άνθρωπος ζητήσει πρόσβαση σε ένα ασφαλές σύστημα φυλασσόμενο από βιομετρική αναγνώριση, παρουσιάζοντας το βιομετρικό χαρακτηριστικό ενός εξουσιοδοτημένου χρήστη, δεν μπορεί να είναι άλλος από τον ίδιο τον εξουσιοδοτημένο χρήστη. Δυστυχώς, αυτό δεν ισχύει στην πράξη, καθώς πολλά χαρακτηριστικά, ιδιαίτερα εκείνα που βασίζονται στη φυσιολογία του ανθρώπου, όπως τα δακτυλικά αποτυπώματα, οι ίριδες των ματιών, το πρόσωπο και άλλα, είναι επιρρεπή στην πλαστογράφιση, η οποία αποτελεί έναν από τους μεγαλύτερους κινδύνους στο τομέα της βιομετρικής αναγνώρισης.

Η πλαστογράφιση ενός βιομετρικού γνωρίσματος πραγματοποιείται σε δύο κύρια στάδια. Το πρώτο, είναι η λήψη του βιομετρικού χαρακτηριστικού του εξουσιοδοτημένου χρήστη που είναι στο στόχαστρο της διαδικασίας και το δεύτερο αφορά την αναπαραγωγή του χαρακτηριστικού εκείνου σε ένα μέσο το οποίο μπορεί να χρησιμοποιηθεί για την απόκτηση πρόσβασης στο ασφαλές σύστημα ενδιαφέροντος. Οι δύο παράγοντες που δυσκολεύουν την επιτυχία της πλαστογράφισης, αφορούν τα δύο αυτά στάδια και είναι το γεγονός ότι κάποια βιομετρικά χαρακτηριστικά είναι δυσκολότερα στην πρόσκτησή τους, καθώς κάποια συστήματα χρησιμοποιούν τεχνικές που ανιχνεύουν την αυθεντικότητα των βιομετρικών στοιχείων με τα οποία τροφοδοτούνται.

4.4 Αυτόματη Επιβεβαίωση της Ανθρώπινης Ταυτότητας μέσω Υπογραφής

Η αυτόματη ταυτοποίηση ενός ατόμου μέσω της χειρόγραφης υπογραφής του (Automatic Handwritten Signature Verification, AHSV) [110], αποτελεί μια πρόκληση, η οποία ερευνάται εκτενώς τις τελευταίες δεκαετίες, καθώς με τη βελτίωσή της θα γίνει διευκόλυνση πολλών ανθρώπων σε μια πληθώρα διαδικασιών, με τις οποίες βρίσκονται αντιμέτωποι καθημερινά. Η κύρια δυσκολία που εμποδίζει τη βελτίωση της αυτόματης ταυτοποίησης ατόμων μέσω των υπογραφών τους, πηγάζει από το γεγονός ότι η υπογραφή ενός ανθρώπου δε γράφεται ποτέ με ακριβώς τον ίδιο τρόπο, καθώς αποτελεί ένα συμπεριφορικό βιομετρικό χαρακτηριστικό (behavioral biometric trait), το οποίο επηρεάζεται από πολλαπλούς εσωτερικούς και εξωτερικούς παράγοντες. Κάποιοι από τους εσωτερικούς παράγοντες που επηρεάζουν την ικανότητα ενός ανθρώπου να χαράξει την ίδια υπογραφή κάθε φορά, αφορούν την ψυχολογία του, το αν είναι κουρασμένος ή στρεσαρισμένος, αν έχει καταναλώσει αλκοόλ ή αν λαμβάνει κάποια φαρμακευτική αγωγή, που επηρεάζει την κινησιολογία του. Από την άλλη, οι εξωτερικοί παράγοντες (perturbations) που έχουν αποδειχθεί πως επηρεάζουν αυτή τη διαδικασία, μπορεί να είναι το υλικό με το οποίο υπογράφει ο χρήστης, η στάση του σώματός του, το αν φοράει κοσμήματα, καθώς και περιβαλλοντικοί παράγοντες, όπως ο θόρυβος και η θερμοκρασία και άλλοι.

Σημαντικό είναι επίσης το γεγονός, ότι αν ένας άνθρωπος χαράζει την υπογραφή του πολλές φορές σε ένα ελεγχόμενο περιβάλλον, κατά το οποίο οι εξωτερικοί παράγοντες είναι σταθεροί και δεν τον επηρεάζουν σε αυτή τη διαδικασία, παρατηρείται ότι και πάλι καμία από τις υπογραφές του δεν είναι ακριβώς ίδια με τις υπόλοιπες. Το γεγονός αυτό οδηγεί στον χαρακτηρισμό των εσωτερικών παραγόντων ως **ενδοπροσωπική μεταβλητότητα** (intrapersonal variability) [110], η οποία προσδίδει τη διαφορετικότητα των υπογραφών του χρήστη κάθε φορά που εκείνος υπογράφει, ανεξαρτήτως των συνθηκών που επικρατούν στο περιβάλλον στο οποίο βρίσκεται (Εικόνα 10b).

Η ποσότητα που περιγράφει τη διακύμανση μεταξύ δύο συνόλων υπογραφών που παράγονται από διαφορετικούς ανθρώπους, ονομάζεται **διαπροσωπική μεταβλητότητα** (interpersonal variability) [110] (Εικόνα 10b). Στη βιβλιογραφία των υπογραφών, όταν συγκρίνεται η αυθεντική υπογραφή ενός εξουσιοδοτημένου χρήστη με την υπογραφή ενός άγνωστου, εξουσιοδοτημένου και μη, η υπογραφή αυτή χαρακτηρίζεται ως πλαστή (forgery). Η διαπροσωπική μεταβλητότητα, αναλύεται στις παρακάτω τρεις κατηγορίες ζευγών υπογραφών, ανάλογα με το είδος της πλαστής υπογραφής που συγκρίνεται με την αυθεντική:

- **Τυχαίες Πλαστογραφίες** (Random Forgeries, RF), αποτελούν οι υπογραφές που χαράζονται από ανθρώπους που δεν έχουν κάποια γνώση για τη συγκρινόμενη αυθεντική. Οι λεγόμενες πλαστές υπογραφές αυτής της κατηγορίας, μπορεί να αποτελούν είτε πραγματικές απόπειρες μίμησης, της αυθεντικής υπογραφής ενός χρήστη από επιτήδειους, είτε ακόμα και άλλες αυθεντικές υπογραφές, άλλων εξουσιοδοτημένων χρηστών.
- **Μη-Επιτηδευμένες Πλαστογραφίες** (Simple Forgeries, SF), είναι οι υπογραφές εκείνες που προέρχονται από ανθρώπους με την πρόθεση πλαστογράφησης της αυθεντικής υπογραφής ενός εξουσιοδοτημένου χρήστη, και ενώ γνωρίζουν το όνομά του, δεν έχουν γνώση της υπογραφής του. Οι πλαστογράφοι αυτής της κατηγορίας βασίζονται στο γεγονός ότι, κάποιοι άνθρωποι υπογράφουν χαράζοντας απλά το ονοματεπώνυμό τους, χωρίς να κάνουν κάποιο καλλιτεχνικό σχέδιο όπως συχνά παρατηρείται σε κάποιες χώρες.
- **Επιτηδευμένες Πλαστογραφίες** (Skilled Forgeries, SK), αποτελούν οι υπογραφές που χαράζονται από πλαστογράφους επιδέξιους να μιμηθούν την υπογραφή και τον τρόπο με τον οποίο τη χαράζει ο χρήστης που επιθυμούν να υποδυθούν. Στόχος των πλαστογράφων αυτής της κατηγορίας είναι να προσεγγίσουν την ενδοπροσωπική μεταβλητότητα του εξουσιοδοτημένου χρήστη, έτσι ώστε να ξεγελάσουν το σύστημα ταυτοποίησης.

Ένα σύστημα βιομετρικής ταυτοποίησης μέσω υπογραφών έχει δύο λειτουργίες. Αυτή της **καταχώρησης** μιας νέας υπογραφής και εκείνη της **αναγνώρισης** μιας εισαγόμενης. Η λειτουργία της καταχώρησης ενός νέου χρήστη, απαιτεί την υποβολή πολλαπλών υπογραφών του για τον υπολογισμό της ενδοπροσωπικής του μεταβλητότητας. Κατά την αναγνώριση μιας εισαγόμενης στο σύστημα υπογραφής, για την ταυτοποίηση ενός ισχυριζόμενου εξουσιοδοτημένου χρήστη η υπογραφή αυτή αναλύεται, έτσι ώστε να συγκριθεί με την ενδοπροσωπική μεταβλητότητα που ανήκει στον ισχυριζόμενο. Πρόσβαση παρέχεται σε αυτόν, εφόσον από την ανάλυση της εισαγόμενης υπογραφής κρίνεται ότι αυτή ανήκει στον ισχυριζόμενο χρήστη. Εναλλακτικά, εάν η ενδοπροσωπική μεταβλητότητα της εισαγόμενης υπογραφής ξεπερνά αυτή του ισχυριζόμενου χρήστη, κρίνεται ότι η εισαγόμενη υπογραφή δεν ανήκει σε αυτόν και η πρόσβασή του απορρίπτεται. Συνοψίζοντας, ένα επιτυχές βιομετρικό σύστημα ταυτοποίησης ανθρώπων μέσω των υπογραφών τους, καλείται να ξεχωρίσει τις υπογραφές διαφορετικών χρηστών, οι οποίες δε γράφονται πάντα με ακριβώς τον ίδιο τρόπο από αυτούς, διακρίνοντας πάντα, αν τις γράφει ο χρήστης στον οποίο ανήκουν ή κάποιος που τον υποδύεται, ξεχωρίζοντας με άλλα λόγια, μεταξύ ενδοπροσωπικής και διαπροσωπικής μεταβλητότητας.

4.5 Μέθοδοι Ανάλυσης Υπογραφών

Οι αλγόριθμοι που πραγματοποιούν την ταξινόμηση των υπογραφών, κατηγοριοποιούνται με βάση δύο λειτουργιών τους. Η μία είναι εκείνη που ορίζει το είδος της ανάλυσης που υφίστανται οι υπογραφές, ενώ η δεύτερη αφορά το είδος του ταξινομητή που χρησιμοποιείται. Η ανάλυση των υπογραφών προς καταχώρηση ή προς αναγνώριση, πραγματοποιείται με δύο τρόπους, ανάλογα με το πως αυτές ψηφιοποιούνται από το μέσο εισαγωγής που χρησιμοποιεί ο άνθρωπος κατά τη χάραξη της υπογραφής του. Κατά τη **δυναμική ή online ανάλυση** (dynamic/online analysis), η υπογραφή συλλέγεται ως ένα σύνολο μετρήσεων που προέρχονται από την ειδική ηλεκτρονική συσκευή, την οποία χρησιμοποιεί η χρήστης για την εισαγωγή της, ενώ κατά τη **στατική ή offline ανάλυση** (static/offline analysis), το μόνο που είναι γνωστό είναι το σχήμα της υπογραφής, αφού αυτή λαμβάνεται ως εικόνα με διάφορες διαβαθμίσεις του γκριζου (grayscale). Οι βασικές αρχές λειτουργίας των δύο αυτών μεθόδων ανάλυσης, παρουσιάζονται παρακάτω.

4.5.1 Δυναμική Ανάλυση Υπογραφών

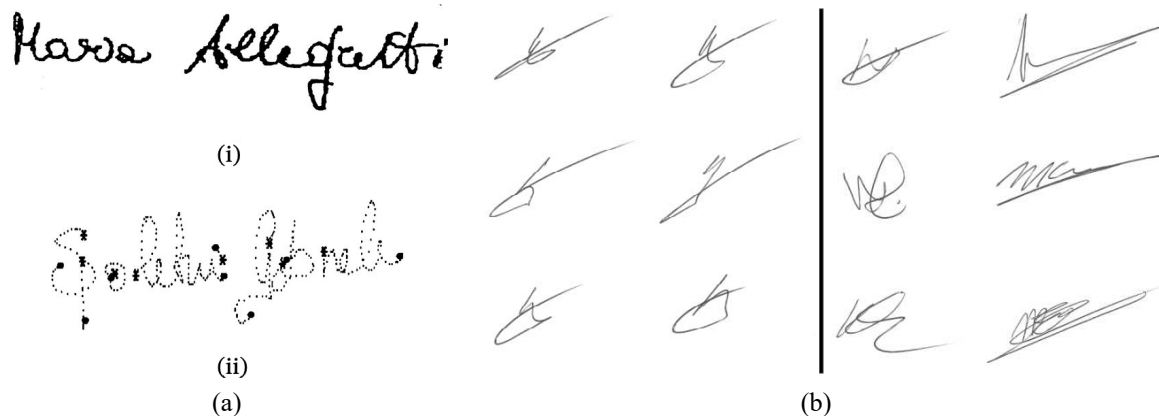
Η δυναμική ανάλυση υπογραφής (dynamic signature analysis) [108], δίνει βάση στον τρόπο με τον οποίο τη γράφει ο χρήστης, γεγονός το οποίο την καθιστά δυσκολότερη στη μίμηση από κάποιον επιτήδειο. Συγκεκριμένα, χρησιμοποιείται μια ηλεκτρονική γραφίδα ή κάποια άλλη συσκευή γραφής ευαίσθητη στην πίεση, όπως tablet ή κινητό τηλέφωνο, για την εισαγωγή της υπογραφής του χρήστη. Η συσκευή αυτή καταγράφει τις αλλαγές στην πίεση και την ταχύτητα με την οποία ο χρήστης χαράζει την υπογραφή του στην επιφάνεια γραφής της (Εικόνα 10a). Στο τέλος, πέρα από την όψη της υπογραφής του χρήστη συλλέγονται και τα παραπάνω δεδομένα, τα οποία αναμένεται να επαναλάβει ο χρήστης με μια μικρή αποδεκτή διακύμανση την επόμενη φορά που θα χαράζει την υπογραφή του. Το γεγονός αυτό, προσδίδει ένα βαθμό ασφάλειας στη μέθοδο αυτή διότι ενώ υπάρχουν επιτήδαιοι που καταφέρνουν, έπειτα από εξάσκηση, να μιμηθούν την υπογραφή ενός εξουσιοδοτημένου χρήστη, το ότι απαιτείται να μιμηθούν και τον ακριβές τρόπο με τον οποίο αυτός τη χαράζει, είναι μια πολύ πιο δύσκολη διαδικασία.

Παρόλο που καθ' αυτή τη διαδικασία επαλήθευσης της υπογραφής ενός χρήστη δίνεται περισσότερη βαρύτητα στον τρόπο γραφής του χρήστη παρά στην ίδια την υπογραφή, η πολυπλοκότητα της υπογραφής παραμένει ένας σημαντικός παράγοντας στην ασφάλεια της, καθώς σχετίζεται άμεσα με τη δυσκολία μίμησής της. Επιπλέον, λόγω του ότι δίνεται τόσο μεγάλη βάση στο τρόπο γραφής του χρήστη, όταν ο τρόπος με τον οποίο αυτός τη χαράζει διαφέρει περισσότερο από ότι συνήθως, η πρόσβασή του απορρίπτεται. Αυτό το είδος λανθασμένης κατηγοριοποίησης ονομάζεται Εσφαλμένη Απόρριψη (False Negative, FN) και ενώ αποτελεί λιγότερο σημαντικό λάθος από την Εσφαλμένη Έγκριση (False Positive, FP), στην περίπτωση της παροχής πρόσβασης αποτελεί ένα ανεπιθύμητο λάθος, το οποίο υπάρχει σε μεγάλο βαθμό με αυτή τη μέθοδο ανάλυσης.

4.5.2 Στατική Ανάλυση Υπογραφών

Κατά τη στατική ανάλυση μιας υπογραφής (static signature analysis), η υπογραφή του χρήστη προέρχεται από την ψηφιοποίηση ενός εγγράφου στο οποίο ο χρήστης έχει πραγματοποιήσει τη χάραξη της υπογραφής του (Εικόνα 10a). Σε αυτή την περίπτωση ανάλυσης, η υπογραφή αποτελεί στοιχείο μιας εικόνας και δεν παρέχεται κάποια πληροφορία για το τρόπο γραφής της. Επομένως, αποκλειστική βάση δίνεται στην όψη της ίδιας της υπογραφής. Λόγω αυτού, η ισχύς της υπογραφής, καθώς και η δυσκολία στη μίμησή της, βασίζεται εξ ολοκλήρου στην πολυπλοκότητά της. Προτού μπορέσει να πραγματοποιηθεί η ανάλυση της υπογραφής, θα πρέπει αυτή

να εξαχθεί από την εικόνα. Συγκεκριμένα, για τη σωστή καταχώρηση και την καλύτερη αναγνώρισή της, θα πρέπει το υπόβαθρο της υπογραφής να αφαιρεθεί μαζί με κάθε ανεπιθύμητο στίγμα (image spots/blobs) ή μουτζούρα που μπορεί να βρίσκεται κοντά σε εκείνη, το οποίο δεν αποτελεί μέρος της. Η διαδικασία αυτή ονομάζεται προεπεξεργασία (preprocessing) και οι ενέργειες που εκτελούνται αποτελούν τεχνικές από το κλάδο της Επεξεργασίας Εικόνας. Άλλες μορφές προεπεξεργασίας που χρησιμοποιούνται, είναι η αφαίρεση το θορύβου της εικόνας μέσω φίλτρων [111–113], η κανονικοποίηση του μεγέθους όλων των υπογραφών [112, 113], η δυαδικοποίηση (binarization) της εικόνας [113] και η λέπτυνση (thinning) της υπογραφής [112, 114].



Εικόνα 10: (a) Υπογραφές που προέρχονται από (i) στατική και (ii) δυναμική ανάλυση²⁰, (b) Παραδείγματα υπογραφών που παρουσιάζουν την ενδοπροσωπική και τη διαπροσωπική μεταβλητότητα, στο αριστερό και δεξί τμήμα της εικόνας αντίστοιχα²¹.

4.6 Μέθοδοι Ταξινόμησης Υπογραφών

Το στάδιο μετά την ανάλυση μιας υπογραφής, είναι η ταξινόμησή της, για την οποία ακολουθούνται κατά κύριο λόγο δύο μέθοδοι. Η επιλογή ανάμεσα σε αυτές τις δύο, καθορίζει το πλήθος των ταξινομητών που χρησιμοποιούνται για το σύστημα ταυτοποίησης. Κατά τη **Writer Dependent** (WD) μέθοδο, το πλήθος των ταξινομητών που κατασκευάζονται είναι τόσο, όσο και οι χρήστες του συστήματος, ενώ κατά τη **Writer Independent** (WI) μέθοδο, χρησιμοποιείται ένας ολικός ταξινομητής για την αναγνώριση οποιουδήποτε καινούριου συμμετέχοντα.

4.6.1 Writer Dependent Μέθοδος

Η Writer Dependent (WD) μέθοδος αποτελεί την πιο διαδεδομένη από τις δύο, καθώς η υλοποίησή της είναι απλή, ενώ παράλληλα επιτυγχάνει μεγάλη ακρίβεια στην αναγνώριση υπογραφών. Κατά τη μέθοδο αυτή, αναπτύσσεται ένας ταξινομητής για κάθε εξουσιοδοτημένο χρήστη, ο οποίος εκπαιδεύεται με δείγματα υπογραφών του, ώστε να ξεχωρίσει το αν μια εισαγόμενη υπογραφή ανήκει σε αυτόν. Το μειονέκτημα αυτής της μεθόδου, είναι το γεγονός ότι για

²⁰Impedovo, Donato & Pirlo, Giuseppe. (2008). Automatic Signature Verification: The State of the Art. Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on. 38. 609 - 635. 10.1109/TSMCC.2008.923866.

²¹Alewijnse, L. C. (2008). Analysis of signature complexity, master thesis. University of Amsterdam.

την επαρκή εκπαίδευση του ταξινομητή κάθε χρήστη, απαιτείται η υποβολή πολλών δειγμάτων των υπογραφών του, κάτι το οποίο δεν είναι βολικό για τον ίδιο.

4.6.2 Writer Independent Μέθοδος

Κατά τη Writer Independent (WI) μέθοδο, εκπαιδεύεται ένας ολικός ταξινομητής, ο οποίος θα έχει την ικανότητα να ταξινομεί τις υπογραφές όλων των χρηστών του συστήματος. Στην περίπτωση αυτή, τα δείγματα των υπογραφών αναπαρίστανται σε ένα νέο χώρο χαρακτηριστικών, στον οποίο όταν ανήκουν στον ίδιο χρήστη βρίσκονται σε κοντινή απόσταση, σχηματίζοντας ομάδες. Ο χώρος αυτός, ονομάζεται **χώρος διαφορών** (dissimilarity space) [115], παράδειγμα του οποίου παρουσιάζεται στην Εικόνα 11, και για τη μετατροπή των αναπαραστάσεων των υπογραφών σε σημεία αυτού του χώρου, χρησιμοποιείται ο λεγόμενος **μετασχηματισμός της διχοτόμησης** (dichotomy transformation) [115, 116]. Σε αυτή την περίπτωση, το πρόβλημα της ταυτοποίησης αντιμετωπίζεται ως μια εφαρμογή δυαδικής ταξινόμησης (binary classification), όπου ο αλγόριθμος εκπαιδεύεται να αναγνωρίζει τις υπογραφές του ενός χρήστη (γνήσιες) από τις υπογραφές των υπολοίπων (πλαστές), διακρίνοντας τη διαφορά τους από το πόσο απέχουν μεταξύ τους. Λόγω αυτού, για τη λειτουργία του αλγόριθμου απαιτείται το λιγότερο μια ήδη υπάρχουσα υπογραφή από τον κάθε εξουσιοδοτημένο, γεγονός το οποίο καθιστά τη μέθοδο αυτή πιο βολική για το χρήστη.

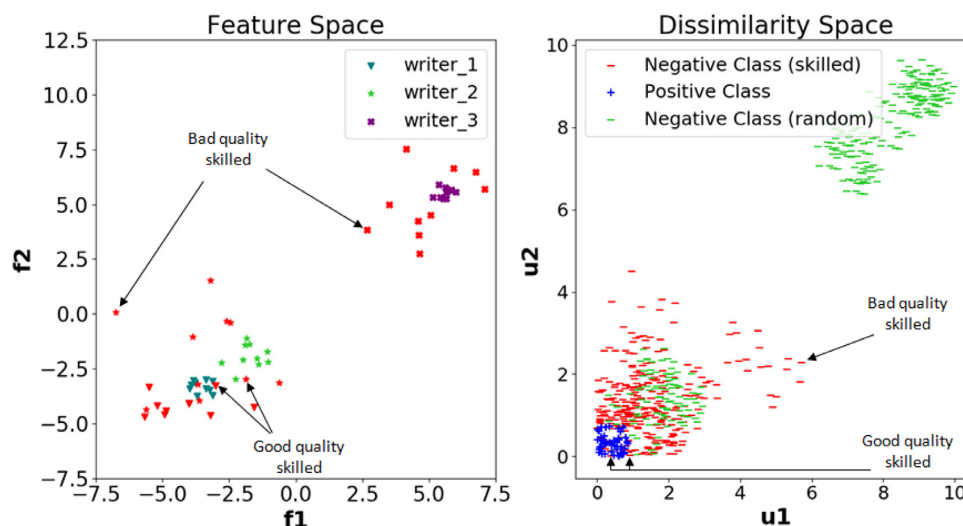
Όπως θα αιτιολογηθεί παρακάτω, η χρήση του μετασχηματισμού της διχοτόμησης δίνει λύση σε πολλά προβλήματα που παρουσιάζονται κατά τη στατική ανάλυση υπογραφών, τα οποία πηγάζουν κυρίως από το πλήθος των υπογραφών που υποβάλλονται από κάθε χρήστη κατά τη διαδικασία της εκπαίδευσης [117], όπως παρουσιάζονται παρακάτω:

1. Ο **αυξημένος αριθμός εξουσιοδοτημένων χρηστών** (κλάσεων) του συστήματος σε σύγκριση με το πλήθος των υπογραφών που υποβάλλουν. Όσες περισσότερες είναι οι κλάσεις σε ένα πρόβλημα ταξινόμησης, τόσο πιο δύσκολη γίνεται η διαφοροποίησή τους, ειδικότερα όταν δεν υπάρχουν αρκετά δεδομένα ή χαρακτηριστικά ώστε να διευκολύνουν τη διαδικασία.
2. Η συνήθως **αυξημένη διαστατικότητα** των αναπαραστάσεων των υπογραφών. Σε αντίθεση με την περίπτωση της δυναμικής ανάλυσης των υπογραφών, όπου κάθε υπογραφή συνοδεύεται από μετρικές για το τρόπο γραφής του χρήστη, κατά τη στατική ανάλυση χρειάζεται να εξάγονται πολλαπλά χαρακτηριστικά από κάθε υπογραφή, όπου το κάθε ένα αποτελεί μια διάσταση στο χώρο χαρακτηριστικών.
3. Η **αυξημένη ενδοπροσωπική μεταβλητότητα** του κάθε χρήστη **σε συνδυασμό με το μικρό αριθμό δειγμάτων που παρέχονται**. Παρόλο που η Writer Independent μέθοδος λειτουργεί χωρίς την απαίτηση υποβολής πολλαπλών δειγμάτων υπογραφών κάθε χρήστη, η ακρίβειά της αυξάνεται όταν παρέχονται περισσότερα από ένα δείγματα, καθώς έτσι μπορεί να γίνει αντιληπτή η ενδοπροσωπική μεταβλητότητα κάθε χρήστη, ώστε να διαφοροποιηθεί από τη διαπροσωπική.
4. Ο **μη-ισορροπημένος αριθμός δειγμάτων υπογραφών που υποβάλλονται από κάθε χρήστη**. Όταν δεν παρέχεται επαρκής γνώση για μια κλάση, γίνεται πιθανότερη η εσφαλμένη ταξινόμηση νέων δεδομένων. Το πρόβλημα αυτό παρουσιάζεται κατά τη διαδικασία της Ταξινόμησης, είτε αυτή γίνεται με την υλοποίηση υπερ επιπέδων (hyperplanes), είτε με βάση την απόσταση γειτνιαζόντων σημείων, όπως υλοποιείται από τους αλγόριθμους SVM και k-NN αντίστοιχα.

5. Η **απουσία δειγμάτων Επιτηδευμένων Πλαστογραφιών (Skilled Forgeries) προς εκπαίδευση**, σε πραγματικές υλοποιήσεις συστημάτων AHSV. Το πιο σύνηθες είναι, ότι υπό πραγματικές συνθήκες, η πρώτη φορά που το σύστημα ταξινόμησης θα κληθεί να αναγνωρίσει μια επιτηδευμένη πλαστή υπογραφή, είναι όταν θα γίνει πραγματική απόπειρα πλαστογράφησης της υπογραφής ενός χρήστη. Το γεγονός αυτό καθιστά εφικτή την εκπαίδευση ενός συστήματος AHSV στην αναγνώριση μόνο μεταξύ ζευγών γνησίων υπογραφών, που ανήκουν σε άλλους χρήστες, δηλαδή Τυχαίων Πλαστογραφιών (Random Forgeries). Συνεπώς, το σύστημα AHSV σε πραγματικές υλοποιήσεις, θα εκπαιδεύεται πάντα με ελλιπή γνώση, το οποίο αποτελεί και τη μεγαλύτερη πρόκληση στις υλοποιήσεις στατικών αναλύσεων υπογραφών.

Η χρήση του μετασχηματισμού της διχοτόμησης κατά Writer Independent μέθοδο ταξινόμησης, βοηθά στην αντιμετώπιση πολλών από των παραπάνω προκλήσεων. Λόγω του γεγονότος ότι ο χώρος διαφορών που αποφέρει, παράγεται με τη σύγκριση ζευγών αποτελούμενων από όμοιων και ανόμοιων υπογραφών, εκείνο που έχει μεγαλύτερη βαρύτητα, είναι η σύνθεση των χαρακτηριστικών εκείνων που διαφοροποιούν καλύτερα τα δείγματα των υπογραφών, ακόμα και αν τα δείγματα που παρέχονται είναι περιορισμένα και ανισόρροπα. Δεύτερο και κυριότερο, λόγω του γεγονότος ότι, η ταξινόμηση βασίζεται στις διαφορές μεταξύ των αναπαραστάσεων των υπογραφών, το σύστημα μπορεί άμεσα να αναγνωρίσει νέα δείγματα για τα οποία δεν έχει εκπαιδευτεί, διευκολύνοντας έτσι την ένταξη νέων χρηστών σε αυτό.

Κατά συνέπεια, τα AHSV συστήματα που υλοποιούν στατική ανάλυση υπογραφών και τη Writer Independent μέθοδο ταξινόμησης, σε συνδυασμό με τον μετασχηματισμό της διχοτόμησης, προσαρμόζονται και επεκτείνονται πιο εύκολα, ενώ το γεγονός ότι βασίζονται στις διαφορές μεταξύ των αναπαραστάσεων με τις οποίες εκπαιδεύονται, δίνει τη δυνατότητα χρήσης τους για την κατάταξη αντικειμένων άλλων εφαρμογών, διαδικασία γνωστή ως Μεταφορά Μάθησης (Transfer Learning).



Εικόνα 11: Η παρουσίαση του χώρου διαφορών (δεξιά), που προέρχεται από την εφαρμογή του μετασχηματισμού της διχοτόμησης, στον αρχικό χώρο χαρακτηριστικών (αριστερά).²²

²²Souza, Victor & Oliveira, Adriano & Cruz, Rafael. (2020). A white-box analysis on the writer-independent dichotomy transformation applied to offline handwritten signature verification. *Expert Systems with Applications*. 154. 10.1016/j.eswa.2020.113397.

5 Πειράματα

Στο κεφάλαιο αυτό παρουσιάζεται η πειραματική διαδικασία που ακολουθήθηκε για την εφαρμογή του υπό ανάλυση αλγορίθμου σε δεδομένα που αφορούν χειρόγραφες υπογραφές διαφόρων χρηστών. Συγκεκριμένα, με τη χρήση του προτεινόμενου αλγορίθμου θα κατασκευαστεί ένα Writer Independent σύστημα αυτόματης ταυτοποίησης ανθρώπων μέσω της στατικής ανάλυσης των υπογραφών τους (Writer Independent Offline AHSV). Αρχικά, γίνεται μια απαραίτητη εισαγωγή σε έννοιες που αφορούν την πειραματική διαδικασία, καθώς και σε διεργασίες που προηγούνται της εξαγωγής των αποτελεσμάτων.

5.1 Βάσεις Δεδομένων

Οι **Βάσεις Δεδομένων** (Datasets) στον τομέα της Υπολογιστικής Νοημοσύνης, αποτελούν συλλογές ακατέργαστης (raw) πληροφορίας, οι οποίες παρέχονται για τη χρήση τους ως δεδομένα εισόδου σε διάφορες εφαρμογές για την αξιολόγηση της λειτουργίας τους. Τα δεδομένα των συνόλων αυτών είναι κάποιας συγκεκριμένης μορφής (κείμενο, αριθμοί, σήματα, εικόνες, βίντεο, κ.α) και συνήθως ανήκουν σε ένα γενικότερο θεματικό πλαίσιο, το οποίο μπορεί να αποτελεί κάτι συγκεκριμένο, όπως στατιστικά αγορών και πωλήσεων, εικόνες που απεικονίζουν είδη λουλουδιών, ζώων και υπογραφών ή κάτι πιο γενικό, όπως συλλογές πολλαπλών αντικειμένων, διαφόρων τύπων.

Δεδομένου των αυξημένων δυνατοτήτων του αλγορίθμου Point-to-Set and Set-to-Set Distance (PSSSD) συγκριτικά με άλλους, καθώς και του γεγονότος ότι η επιτυχής εφαρμογή της αυτόματης ταυτοποίησης ανθρώπων μέσω των χειρόγραφων υπογραφών τους (AHSV) θα αποτελέσει μεγάλη διευκόλυνση στην καθημερινότητά μας, αποφασίστηκε πως θα ήταν ωφέλιμη η αξιοποίησή του σε αυτό το πρόβλημα. Έτσι, στην προκειμένη περίπτωση, για την αξιολόγηση της λειτουργίας του υπό ανάλυση αλγορίθμου, επιλέχθηκαν δύο διάσημες Βάσεις Δεδομένων του χώρου της Υπολογιστικής Όρασης και των υπογραφών, οι οποίες είναι οι CEDAR [118] και η Bangla [119], που αποτελεί υποσύνολο της BHSig260. Κάθε μια από αυτές, περιέχει αυθεντικές και επιτηδευμένες πλαστές υπογραφές (Skilled Forgeries, SF) για κάθε γραφέα τους, όπως αναλύεται στον Πίνακα 2 παρακάτω.

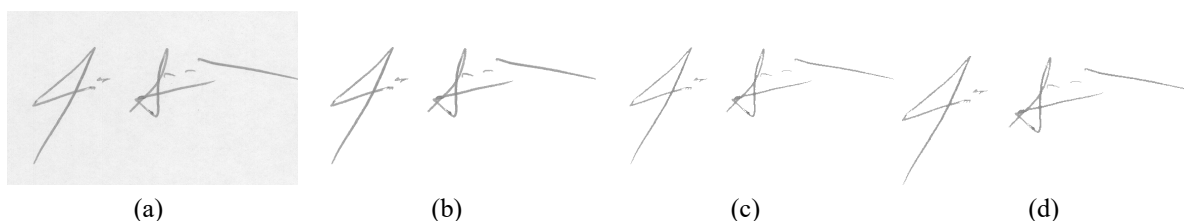
Βάση Δεδομένων (Dataset)	Γραφικός Χαρακτήρας	# Υπογραφότων (Signatories)	# Αυθεντικών ανά Γραφέα	# Πλαστών ανά Γραφέα
CEDAR	Λατινικός	55	24	24
Bangla	Βεγγάλης	100	24	30

Πίνακας 2: Τα χαρακτηριστικά των Βάσεων CEDAR και Bangla.

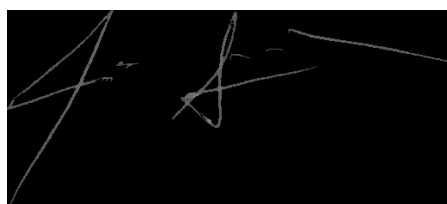
5.2 Προεπεξεργασία Εικόνων

Σε εφαρμογές Υπολογιστικής Όρασης, όπου τα δεδομένα εισόδου προέρχονται από οπτικούς αισθητήρες, συνήθως απαιτείται η απομόνωση και η ανάδειξη των χαρακτηριστικών της ακατέργαστης (raw) εικόνας, από τα οποία επωφελείται η εφαρμογή για την οποία προορίζονται. Η διαδικασία αυτή ονομάζεται **προεπεξεργασία** (preprocessing) και επιτυγχάνεται χρησιμοποιώντας τεχνικές από το κλάδο της Επεξεργασίας Εικόνας. Για τη συγκεκριμένη εφαρμογή το ζητούμενο είναι η διατήρηση του ίχνους της υπογραφής, απομονώνοντάς το από την υπόλοιπη αρχική εικόνα (Εικόνα 12a). Η σειρά των βημάτων που ακολουθούνται για την επίτευξη αυτής της διαδικασίας, είναι η ακόλουθη:

1. **Αφαίρεση στιγμάτων** (spot/blob removal) και όποιας άλλης μορφής θορύβου από την εικόνα, που δεν αποτελούν μέρος της υπογραφής (Εικόνα 12b).
2. **Λέπτυνση** (Thinning) του ίχνους της υπογραφής χρησιμοποιώντας τον αλγόριθμο καταωφλοιοποίησης (thresholding) του Otsu [120] (Εικόνα 12c). Η διαδικασία αυτή είναι μια μορφή κανονικοποίησης της όψης των υπογραφών, καθώς έτσι διατηρείται το σχήμα του ίχνους της υπογραφής και όχι οι ανεπιθύμητες αλλαγές στην πίεση κατά τη χάραξή της. Για το σύνολο των υπογραφών της κάθε Βάσης Δεδομένων έχει υπολογιστεί το βέλτιστο ποσοστό της λέπτυνσης που θα υποστεί κάθε υπογραφή, το οποίο εάν αφαιρεθεί από την κάθε μία δε θα υπάρξει διακοπή στο ίχνος της, ακόμα και στα πιο λεπτά της σημεία. Η τεχνική αυτή περιγράφεται στη δημοσίευση των Zois *et al.* [121] και ονομάζεται Optimal Thinning Level (OTL).
3. **Αποκοπή** της υπογραφής, με σκοπό τον περιορισμό της περιττής ποσότητας υποβάθρου της εικόνας (Εικόνα 12d). Για αυτή τη διαδικασία εντοπίζεται το πλαίσιο που περιβάλλει την υπογραφή, και ότι είναι εκτός αυτού, αφαιρείται από την εικόνα.
4. **Αντιστροφή του χρώματος** της εικόνας (Εικόνα 13). Επειδή συνήθως οι υπογραφές απεικονίζονται σε χαρτί ανοικτού χρώματος, είναι απαραίτητο αυτό που αποτελεί το υπόβαθρο της υπογραφής να είναι μαύρο. Το γεγονός αυτό, συμβολίζει ότι στην περιοχή εκείνη δεν υπάρχει ίχνος υπογραφής. Εν αντιθέσει, το ίχνος της ίδιας της υπογραφής, το οποίο απεικονίζεται στο χαρτί με μελάνι σκούρου χρώματος, είναι απαραίτητο να τείνει προς το λευκό χρώμα. Έτσι, τα λευκότερα σημεία της χαρακτηρίζουν την παραπάνω απόθεση μελανιού από το χρήστη, ενώ τα πιο σκούρα, τη λιγότερη. Αυτό το στάδιο προεπεξεργασίας, όντας το τελευταίο, είναι το τελικό αποτέλεσμα που προκύπτει από μια αρχική εικόνα και προορίζεται για το επόμενο στάδιο, αυτό της εξαγωγής χαρακτηριστικών.



Εικόνα 12: Τα στάδια της προεπεξεργασίας, οπτικοποιημένα. (a) Η αρχική εικόνα, που απεικονίζει την υπογραφή ενός χρήστη, (b) Η ίδια εικόνα μετά την αφαίρεση των στιγμάτων της, (c) Το αποτέλεσμα της λέπτυνσης της υπογραφής, (d) Περικοπή της υπογραφής.



Εικόνα 13: Η τελική εικόνα που απεικονίζει την υπογραφή ενός χρήστη, μετά το πέρας όλων των σταδίων προεπεξεργασίας.

5.3 Εξαγωγή Χαρακτηριστικών

Αφού γίνει η προεπεξεργασία κάθε υπογραφής, ακολουθεί η διαδικασία εξαγωγής των χαρακτηριστικών, που θεωρούνται από την εκάστοτε εφαρμογή, ότι είναι αυτά που διαφοροποιούν καλύτερα το αντικείμενο ενδιαφέροντος. Το στάδιο της επιλογής των χαρακτηριστικών είναι καθοριστικό στην ολική επίδοση του αλγόριθμου ταξινόμησης, καθώς όσο πιο κατάλληλα είναι τα χαρακτηριστικά που θα επιλεχθούν για να περιγράψουν το αντικείμενο ενδιαφέροντος, τόσο πιο εύκολη θα γίνει η ταξινόμησή του. Ενώ υπάρχουν πολλές μέθοδοι χαρακτηρισμού αντικειμένων στην Υπολογιστική Όραση, όπως μέσω της εξαγωγής των λεγόμενων Key Points, σαν αυτών που προσδιορίζονται από τους αλγόριθμους SIFT [122, 123], SURF [124], KAZE [125] και άλλων, η μέθοδος που χρησιμοποιείται σε αυτή την περίπτωση βασίζεται στη χρήση φίλτρων από τον κλάδο της Επεξεργασίας Εικόνας. Οι αποκρίσεις που προκύπτουν από την εφαρμογή φίλτρων σε κάθε εικόνα υπογραμμίζουν διαφορετικά στοιχεία που απεικονίζονται, όπως σχήματα, χρώματα και άλλα. Στην προκειμένη περίπτωση τα διάφορα φίλτρα που επιλέγονται στοχεύουν να επισημάνουν, το κάθε ένα με διαφορετικό τρόπο, το σχήμα της υπογραφής του χρήστη. Συλλέγοντας όλες τις αποκρίσεις των εικόνων από την εφαρμογή φίλτρων σε αυτές, διαπιστώνονται διαφορές μεταξύ τους, και έτσι γίνεται εφικτή η διαφοροποίησή τους. Τα φίλτρα αυτά αποτελούν τα **χαρακτηριστικά** (features), που χρησιμοποιούνται για να γίνει ο διαχωρισμός μεταξύ των αντικειμένων της εφαρμογής.

Το **διάνυσμα χαρακτηριστικών** (feature vector) $\Phi(\mathbf{I}, x, y)$, που περιγράφει το σύνολο των n φίλτρων (χαρακτηριστικών) $\Phi_i(\mathbf{I}, x, y)$, με $i = \{1, 2, \dots, n\}$, που θα εφαρμοστούν στις ασπρόμαυρες (grayscale) εικόνες $\mathbf{I}(x, y) \equiv \mathbf{I}$ των υπογραφών των χρηστών, παρουσιάζεται στη Σχέση 35. Το πλήθος των φίλτρων κατά τη συγκεκριμένη εφαρμογή, ισούται με $n = 12$. Όπου x και y , είναι δείκτες των οριζόντιων και κατακόρυφων pixel της εκάστοτε εικόνας αντίστοιχα. Οι διαφορετικές αποκρίσεις της κάθε εικόνας, που προκύπτουν από την εφαρμογή των φίλτρων σε αυτές, συντελούν μια στοίβα εικόνων (image stack) $\mathbf{F}(x, y, i)$, διαστάσεων ίσων με αυτών της εικόνας από την οποία προκύπτουν.

$$\Phi(\mathbf{I}, x, y) = [\mathbf{I}, \mathbf{I}_x, \mathbf{I}_y, \mathbf{I}_{xx}, \mathbf{I}_{xy}, \mathbf{I}_{yy}, |\nabla_{\mathbf{I}}|, \theta_{\mathbf{I}}, \nabla^2(\mathbf{I}), \text{LoG}(\mathbf{I}), x_n, y_n]^T \quad (35)$$

Τα στοιχεία $\mathbf{I}_x = \frac{\partial \mathbf{I}}{\partial x}$ και $\mathbf{I}_y = \frac{\partial \mathbf{I}}{\partial y}$, αποτελούν τις πρώτες μερικές παραγώγους της έντασης (intensity) της ασπρόμαυρης εικόνας, ως προς την οριζόντια και κατακόρυφη κατεύθυνση αντίστοιχα. Παρομοίως, οι όροι $\mathbf{I}_{xx} = \frac{\partial^2 \mathbf{I}}{\partial x^2}$ και $\mathbf{I}_{yy} = \frac{\partial^2 \mathbf{I}}{\partial y^2}$, αντιπροσωπεύουν τις δεύτερες μερικές παραγώγους της έντασης της εικόνας ως προς την οριζόντια και κατακόρυφη κατεύθυνση. Επιπλέον, ο όρος $\mathbf{I}_{xy} = \frac{\partial^2 \mathbf{I}}{\partial x \partial y}$, αποτελεί τη μικτή παράγωγο δεύτερης τάξης της έντασης της εικόνας, δηλαδή τη μερική παράγωγο ως προς τη μια κατεύθυνση και μετά ως προς την άλλη, χωρίς να έχει σημασία ποια κατεύθυνση επιλέγεται πρώτη. Οι όροι $|\nabla_{\mathbf{I}}| = \sqrt{\mathbf{I}_x^2 + \mathbf{I}_y^2}$ και $\theta_{\mathbf{I}} = \tan^{-1}(\mathbf{I}_y/\mathbf{I}_x)$, εκφράζουν το μέτρο και την κατεύθυνση της πρώτης παραγώγου της έντασης της ασπρόμαυρης εικόνας. Η κατεύθυνση $\theta_{\mathbf{I}}$ κανονικοποιείται σε ακτίνια (radians) στο εύρος $[-\pi, \pi)$. Περαιτέρω, οι ποσότητες x_n και y_n , αποτελούν τις κανονικοποιημένες συντεταγμένες του ίχνους της υπογραφής. Τέλος, οι όροι $\nabla^2(\mathbf{I})$ και $\text{LoG}(\mathbf{I})$, αποτελούν τις αποκρίσεις της ασπρόμαυρης εικόνας, από την εφαρμογή του α) Laplacian και β) του Laplacian of Gaussian (LoG) φίλτρου αντίστοιχα. Ενώ και τα δύο φίλτρα χρησιμοποιούνται για την ανίχνευση ακμών (edge detection), το Laplacian of Gaussian εξομαλύνει την εικόνα με τη χρήση του Gaussian smoothing φίλτρου, επιτυγχάνοντας με αυτό το τρόπο ανοχή στο θόρυβο, αφού οι ακμές δε θα είναι τόσο έντονες. Τα φίλτρα αυτά προσδιορίζονται είτε μέσω τύπων, είτε μέσω της συνέλιξης της αρχικής εικόνας με τους παρακάτω πυρήνες μετασχηματισμού (kernel matrices), όπως ορίζονται στην παρακάτω σχέση. Η πράξη της συνέλιξης μεταξύ δύο ποσοτήτων συμβολίζεται ως $*$.

$$\nabla^2 (\mathbf{I}(x, y)) = \frac{\partial^2 \mathbf{I}(x, y)}{\partial x^2} + \frac{\partial^2 \mathbf{I}(x, y)}{\partial y^2} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix} * (\mathbf{I}(x, y)) \quad (36)$$

$$\text{LoG} (\mathbf{I}(x, y)) = -\frac{1}{\pi\sigma^4} \left(1 - \frac{x^2 + y^2}{2\sigma^2} \right) e^{-\frac{x^2 + y^2}{2\sigma^2}} =$$

$$\underset{\text{για } \sigma=1.4}{=} \begin{bmatrix} 0 & 1 & 1 & 2 & 2 & 2 & 1 & 1 & 0 \\ 1 & 2 & 4 & 5 & 5 & 5 & 4 & 2 & 1 \\ 1 & 4 & 5 & 3 & 0 & 3 & 5 & 4 & 1 \\ 2 & 5 & 3 & -12 & -24 & -12 & 3 & 5 & 2 \\ 2 & 5 & 0 & -24 & -40 & -24 & 0 & 5 & 2 \\ 2 & 5 & 3 & -12 & -24 & -12 & 3 & 5 & 2 \\ 1 & 4 & 5 & 3 & 0 & 3 & 5 & 4 & 1 \\ 1 & 2 & 4 & 5 & 5 & 5 & 4 & 2 & 1 \\ 0 & 1 & 1 & 2 & 2 & 2 & 1 & 1 & 0 \end{bmatrix} * (\mathbf{I}(x, y)) \quad (37)$$

5.4 Εξαγωγή Πινάκων Συνδιακύμανσης

Οι στοίβες εικόνων $\mathbf{F}(x, y, i)$ που προκύπτουν από το προηγούμενο στάδιο, χαρακτηρίζουν μια εικόνα ως προς κάποια στοιχεία της. Στο παρών στάδιο, θα εξαχθεί μια μοναδική ποσότητα που ονομάζεται **περιγραφέας** (descriptor), η οποία χαρακτηρίζει ξεχωριστά την εικόνα. Για τον υπολογισμό του περιγραφέα στην παρούσα εφαρμογή χρησιμοποιούνται οι στοίβες εικόνων και η μαθηματική ποσότητα της συνδιακύμανσης. Όπως προαναφέρθηκε, η συνάρτηση συνδιακύμανσης φανερώνει το πόσο δύο ή περισσότερες τυχαίες μεταβλητές, δηλαδή δύο σύνολα μετρήσεων, διαφέρουν από κοινού από τις μέσες τιμές τους. Στην προκειμένη περίπτωση, τυχαίες μεταβλητές θεωρούνται οι n ξεχωριστές αποκρίσεις που βρίσκονται στη στοίβα εικόνων $\mathbf{F}(x, y, i)$. Για τον υπολογισμό της συνδιακύμανσης μεταξύ δύο εικόνων της στοίβας $\mathbf{A} = \Phi_{\mathbf{A}}(\mathbf{I}, x, y)$ και $\mathbf{B} = \Phi_{\mathbf{B}}(\mathbf{I}, x, y)$, διαστάσεων $w \times h$ υπολογίζεται η από κοινού διαφορά των pixel τους, από τη μέση τιμή της κάθε απόκρισης, όπως παρουσιάζεται στο παρακάτω τύπο:

$$\text{cov}(\mathbf{A}, \mathbf{B}) = \frac{1}{(w \times h) - 1} \sum_{x=1}^w \sum_{y=1}^h (\mathbf{A}(x, y) - \bar{\mathbf{A}})(\mathbf{B}(x, y) - \bar{\mathbf{B}}) \quad (38)$$

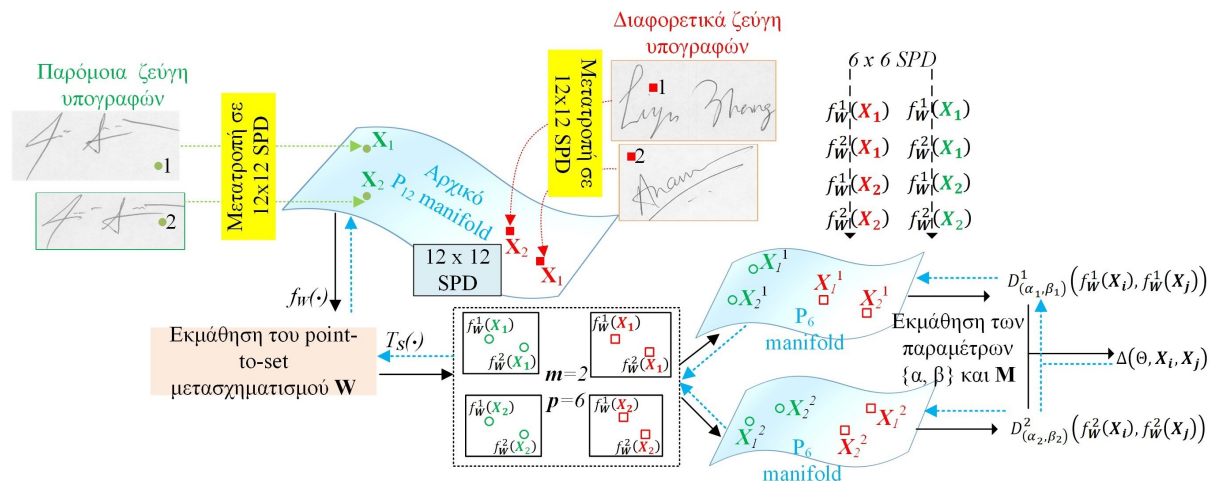
Η αναπαράσταση εικόνων μέσω πινάκων συνδιακύμανσης υπερτερεί έναντι άλλων μορφών [126], όπως των αναπαραστάσεων εικόνων μέσω των ιστογραμμάτων τους (histogram representation), μέσω της απευθείας σύγκρισης των τιμών των pixel τους, με τη χρήση Bag-of-Features (BoF), κ.α, λόγω των παρακάτω:

- Οι διαστάσεις των πινάκων συνδιακύμανσης είναι **ανεξάρτητες των διαστάσεων της εικόνας την οποία περιγράφουν**. Συγκεκριμένα, οι διαστάσεις ενός πίνακα συνδιακύμανσης ισούται με το πλήθος των χαρακτηριστικών που περιγράφουν την εικόνα, δηλαδή με $n \times n$. Συνεπώς, ακόμα και μεγάλες σε μέγεθος εικόνες, μπορούν να αναπαρασταθούν από ένα πίνακα συνδιακύμανσης πιθανότατα μικρότερων διαστάσεων.
- Η αναπαράσταση μέσω πινάκων συνδιακύμανσης είναι **ανεξάρτητη από τον προσανατολισμό της εικόνας** (view/pose invariance), λόγω του γεγονότος ότι η ποσότητα που υπολογίζεται δε συγκρίνει απευθείας τη διαφορά μεταξύ των pixel των αποκρίσεων, αλλά αφορά την εξάρτηση μεταξύ τους.

- Λόγω της χρήσης του μέσου όρου στον υπολογισμό της συνδιακύμανσης, η αναπαράσταση που υπολογίζεται είναι **απαλλαγμένη από θόρυβο**.

5.5 Πειραματική Διαδικασία

Σε αυτό το στάδιο αναλύεται η διαδικασία που ακολουθείται για την υλοποίηση του Writer Independent Offline ASHV συστήματος, κάνοντας χρήση του αλγόριθμου PSSSD. Συνοπτικά, ο αλγόριθμος τροφοδοτείται από ζεύγη αυθεντικών και πλαστών υπογραφών με σκοπό να δημιουργηθεί ένα μοντέλο, το οποίο θα κατασκευάσει ένα Χώρο Διαφορών (Dissimilarity Space), στον οποίο οι υπογραφές του κάθε χρήστη θα είναι ομαδοποιημένες και αποστασιοποιηθέντες από τις υπόλοιπες. Το Σχήμα 14, παρουσιάζει ενδεικτικά τη διαδικασία μάθησης και αξιολόγησης που ακολουθείται από τον αλγόριθμο, για την εκμάθηση των παραμέτρων $\Theta = \{W, A, M\}$, που προσδιορίζουν τη σχέση της απόστασης στο Χώρο Διαφορών. Σημειώνεται ότι η μεθοδολογία των πειραμάτων, ακολουθεί σε κάποιο βαθμό τη δουλειά που προτάθηκε από τους Zois *et al.*, στη δημοσίευσή τους [127].



Εικόνα 14: Ενδεικτική παρουσίαση της διαδικασίας που ακολουθείται από τον αλγόριθμο PSSSD, για την κατασκευή συστήματος WI Offline ASHV. Αρχικά, στην επιφάνεια της αρχικής SPD Πολλαπλότητας P_{12} τοποθετούνται οι αναπαραστάσεις των υπογραφών, που προκύπτουν από τη μετατροπή των εικόνων σε (SPD) πίνακες συνδιακύμανσης. Στη συνέχεια, συλλέγονται ζεύγη όμοιων και ανόμοιων υπογραφών, που απεικονίζονται ως πράσινοι κύκλοι και κόκκινα τετράγωνα αντίστοιχα. Έπειτα, γίνεται η προβολή των αναπαραστάσεων σε SPD Πολλαπλότητες μειωμένων διαστάσεων P_6 δια μέσω του μοντέλου $\Theta = \{W, A, M\}$, στις οποίες υπολογίζεται η επιμέρους απόσταση μεταξύ των υπογραφών. Τέλος, πραγματοποιούνται οι διαδικασίες της εκπαίδευσης (training) και της αξιολόγησης (validation), συμβολιζόμενες από τα συνεχόμενα μαύρα και τα γαλάζια διακοπτόμενα βέλη αντίστοιχα, εξάγοντας την τελική σχέση της απόστασης.²³

Για τις διαδικασίες της ανάπτυξης (development) και του ελέγχου (testing) ακολουθούνται δύο διαφορετικά **πειραματικά σενάρια**, που αφορούν τα δεδομένα με τα οποία πρόκειται να κατασκευαστούν τα μοντέλα. Κατά το πρώτο, εξετάζεται η επίδοση του μοντέλου που αναπτύσσεται, όταν τα στάδια της εκπαίδευσης (training) και της αξιολόγησης (validation)²⁴ διεξάγονται

²³E. N. Zois, D. Tsourounis and D. Kalivas, "Similarity Distance Learning on SPD Manifold for Writer Independent Offline Signature Verification," in IEEE Transactions on Information Forensics and Security, vol. 19, pp. 1342-1356, 2024, doi: 10.1109/TIFS.2023.3333681.

με υποσύνολα της ίδιας Βάσης Δεδομένων, ενώ ο έλεγχος (testing) γίνεται με τα υπόλοιπα δεδομένα της. Το σενάριο αυτό, ακολουθείται τις περισσότερες φορές στο χώρο της Μηχανικής Μάθησης και εξετάζει την επίδοση του αλγόριθμου εάν κληθεί να αντιμετωπίσει δεδομένα, μορφής παρόμοιας με αυτά που έχει εκπαιδευτεί. Κατά το δεύτερο σενάριο, το μοντέλο αναπτύσσεται με τα δεδομένα μιας Βάσης, ενώ η επίδοσή του ελέγχεται χρησιμοποιώντας την άλλη, στοχεύοντας έτσι να μετρηθεί η ικανότητά του στην περίπτωση που επιχειρηθεί Μεταφορά Μάθησης (Transfer Learning).

5.5.1 Στάδιο Εκπαίδευσης και Αξιολόγησης

Κατά τη διαδικασία της **εκπαίδευσης** (training) ο αλγόριθμος τροφοδοτείται με δεδομένα γνωστών κλάσεων, με σκοπό να αντιληφθεί τις διαφορές μεταξύ τους, ώστε μελλοντικά να καταφέρει να διαχωρίζει νέα στις σωστές κατηγορίες τους. Στη συγκεκριμένη περίπτωση, μέρος της διαδικασίας της εκπαίδευσης, αποτελεί και η **αξιολόγηση** (validation) του μοντέλου που έχει δημιουργηθεί. Κατά την αξιολόγηση, γίνεται μια πρόβλεψη για την προσαρμογή του μοντέλου στη διαδικασία που καλείται να εκτελέσει. Όπως γίνεται και με τη διαδικασία του ελέγχου, η οποία αναλύεται στο επόμενο κεφάλαιο, τα δεδομένα που χρησιμοποιούνται για την αξιολόγηση του αλγορίθμου, δεν έχουν συμμετάσχει στο στάδιο της εκπαίδευσης. Επομένως, όταν κληθούν προς αναγνώριση, τα αποτελέσματα που θα ληφθούν αποτελούν μια αμερόληπτη εκτίμηση της απόδοσης που θα έχει ο αλγόριθμος υπό πραγματικές συνθήκες. Περαιτέρω, επειδή τα δεδομένα αξιολόγησης συνήθως επιλέγονται να είναι αρκετά λιγότερα σε πλήθος από αυτά της εκπαίδευσης και του ελέγχου, η αξιολόγηση του μοντέλου που έχει εκπαιδευτεί διεξάγεται σε σύντομο χρόνο. Λόγω των παραπάνω προτερημάτων μπορούν να γίνουν αντιληπτά από νωρίς προβλήματα που παρουσιάζει το μοντέλο, όπως αυτό της υπερπροσαρμογής (overfitting). Κατά την υπερπροσαρμογή, το μοντέλο δυσκολεύεται να διαχωρίζει νέα αντικείμενα, γιατί έχει προσαρμοστεί σε τέτοιο βαθμό στα δεδομένα με τα οποία έχει εκπαιδευτεί, που έχει χάσει τη γενική εικόνα των πραγμάτων, αναγνωρίζοντας μόνο εκείνα που πληρούν συγκεκριμένες συνθήκες.

Κατά το **πρώτο πειραματικό σενάριο**, ο πληθυσμός του συνόλου των χρηστών (γραφέων, writers, $(W)^{25}$) της κάθε Βάσης, ο οποίος συμβολίζεται ως $n(W)$, χωρίζεται στη μέση σε ακέραια μέρη, κατασκευάζοντας δύο σύνολα. Το **σύνολο ανάπτυξης** (development set) DEV, πλήθους $n(DEV) = \lceil 0.5 \times n(W) \rceil$, προορίζεται για την εκπαίδευση και αξιολόγηση του αλγόριθμου και το **σύνολο ελέγχου** (testing set) TS, πλήθους $n(TS) = \lfloor 0.5 \times n(W) \rfloor$, για τον έλεγχο. Η διαδικασία της ανάπτυξης και του ελέγχου εκτελούνται πέντε φορές για κάθε αλλαγή των παραμέτρων m και p του αλγόριθμου PSSSD, που αφορούν τις διαστάσεις και το πλήθος των προβολών αντίστοιχα, όπως θα γίνει αντιληπτό στο κεφάλαιο παρουσίασης των αποτελεσμάτων. Η διαδικασία αυτή είναι γνωστή ως “k-fold Cross-Validation” και στην προκειμένη περίπτωση χαρακτηρίζεται ως 5 × 2-fold Cross Validation, αφού για κάθε μια από τις δύο Βάσεις Δεδομένων το πλήθος των χρηστών τους χωρίζεται σε δύο υποσύνολα (folds), ενώ η διαδικασία της εκπαίδευσης και του ελέγχου, επαναλαμβάνεται πέντε φορές. Τέλος, ο ρόλος των υποσυνόλων αυτών εναλλάσσεται κατά τη δεύτερη εκτέλεση της ίδιας διαδικασίας, μειώνοντας έτσι την επίδραση της τύχης στην επίδοση του αλγόριθμου, αναδεικνύοντας την πραγματική του ικανότητα. Σημειώνεται ότι, κατά τη διαίρεση του πληθυσμού των χρηστών χωρίζονται στα δύο αυτά σύνολα, το 50% και των αυθεντικών υπογραφών κάθε χρήστη SIG^+ , αλλά και το 50% των επιτηδευμένων πλαστογραφιών SIG_{SK}^- , που επιχειρούν να μιμηθούν την ταυτότητά του. Επίσης, το πλήθος των στοιχείων ενός συνόλου (cardinality), συμβολίζεται ως: $|\cdot|$ ή ως $n(\cdot)$.

²⁴Το στάδιο της αξιολόγησης (validation) αποτελεί ένα στάδιο αντίστοιχο με αυτό του ελέγχου αλλά μικρότερης κλίμακας, το οποίο εφαρμόζεται σε ξεχωριστό σύνολο πειραματικών δεδομένων.

²⁵Ο όρος W , που αναφέρεται σε αυτό το κεφάλαιο, δε συσχετίζεται με τον πίνακα W του αλγόριθμου PSSSD.

Ο διαχωρισμός των δεδομένων της βάσης κατά το πρώτο πειραματικό σενάριο γίνεται ως εξής. Από τους χρήστες που έχουν κατανεμηθεί για το στάδιο της ανάπτυξης, το σύνολο των αυθεντικών SIG^+ και πλαστών υπογραφών SIG_{SK}^- τους, χωρίζεται σε ποσοστό 70% και 30% για τα επιμέρους στάδια της εκπαίδευσης και της αξιολόγησης αντίστοιχα. Το σύνολο των αυθεντικών SIG^+ και πλαστών SIG_{SK}^- υπογραφών κάθε χρήστη της κάθε Βάσης, που προορίζονται για τα στάδια της εκπαίδευσης και αξιολόγησης, συμβολίζονται ως SIG^{TR+} , SIG^{TR-} και SIG^{V+} , SIG^{V-} αντίστοιχα. Επομένως, το πλήθος των αυθεντικών και των πλαστών υπογραφών για τα σύνολα εκπαίδευσης TR και αξιολόγησης V , ισούται με $n(TR) = [0.7 \times (n(SIG^+) + n(SIG_{SK}^-))]$ και $n(V) = [0.3 \times (n(SIG^+) + n(SIG_{SK}^-))]$ αντίστοιχα. Όπως θα αναλυθεί στη συνέχεια, με τις υπογραφές των χρηστών αυτών των κατανομών, δημιουργούνται ζεύγη όμοιων $p^+ = (X_i, X_j)_S$ και ανόμοιων υπογραφών $p^- = (X_i, X_j)_D$, τα οποία θα χρησιμοποιηθούν για τα στάδια της εκπαίδευσης και αξιολόγησης. Τα ζεύγη των όμοιων (αυθεντικών) υπογραφών (similar pairs) συμβολίζονται ως $p^+ = (X_i, X_j)_S$ και αποτελούνται από υπογραφές του ίδιου χρήστη. Από την άλλη, τα ζεύγη των ανόμοιων υπογραφών (dissimilar pairs) χωρίζονται σε δύο κατηγορίες. Στη μια, τα ζεύγη των ανόμοιων υπογραφών αποτελούνται από την αυθεντική υπογραφή SIG^+ , ενός χρήστη και μια τυχαία πλαστογραφία (RF, SIG_{RF}^-), ενώ στη δεύτερη περίπτωση, η αυθεντική υπογραφή, συνδυάζεται με μια επιτηδευμένη πλαστογραφία (SK, SIG_{SK}^-). Τα ζεύγη αυτά, χαρακτηρίζονται ως p_{RF}^- και p_{SK}^- αντίστοιχα. Προτού ξεκινήσουν τα στάδια της εκπαίδευσης, της αξιολόγησης και του ελέγχου, επιλέγεται μια από τις δύο κατηγορίες ανόμοιων ζευγών, για την εκτέλεση όλων των διαδικασιών.

Το πλήθος των ζευγών των όμοιων $|S|$ και ανόμοιων $|D|$ υπογραφών, προκύπτει από τον συνδυασμό των στοιχείων του πληθυσμού των συνόλων εκπαίδευσης και αξιολόγησης για την κάθε περίπτωση, ανά δύο. Από τα σύνολα των ζευγών των όμοιων S και ανόμοιων D υπογραφών, προκύπτουν τα $S^{TR} = p^{TR+}$, $S^V = p^{V+}$ και $D^{TR} = p^{TR-}$, $D^V = p^{V-}$, τα οποία περιέχουν τα ζεύγη των όμοιων (similar) και ανόμοιων (dissimilar) υπογραφών, που προορίζονται για στάδιο της εκπαίδευσης (training) και της αξιολόγησης (validation) αντίστοιχα. Για την παρούσα πειραματική διαδικασία, το πλήθος των όμοιων και ανόμοιων ζευγών, τέθηκε ίσο ($|S| = |D|$) και για τα δύο στάδια της ανάπτυξης, ώστε να υπάρχει ισορροπία μεταξύ των δειγμάτων της κάθε μιας από τις δύο κλάσεις. Για τη Βάση Bangla, λόγω του γεγονότος ότι κάθε χρήστης της συνοδεύεται από 30 επιτηδευμένες πλαστογραφίες, έγινε τυχαία δειγματοληψία (random sampling), έτσι ώστε να ισχύει ότι $|S| = |D|$. Τέλος, παρακάτω παρουσιάζονται τα ποσά των χρηστών, οι κατανομές των δειγμάτων υπογραφών και των ζευγών που σχηματίζονται, για τις Βάσεις CEDAR και Bangla κατά το πρώτο fold τους:

→ Βάση **CEDAR**, $n(W) = 55$ χρήστες, $n(SIG^+) = 24$ αυθεντικές υπογραφές ανά χρήστη, $n(SIG_{SK}^-) = 24$ πλαστές υπογραφές ανά χρήστη:

- Πλήθος χρηστών για το στάδιο της Ανάπτυξης (Development): $n(DEV) = [0.5 \times n(W)] = [27.5] = 28$ χρήστες

– Για τη διαδικασία της Εκπαίδευσης (Training):

- * Πλήθος αυθεντικών υπογραφών ανά χρήστη:

$$n(SIG^{TR+}) = [0.7 \times n(SIG^+)] = [16.8] = 17 \text{ υπογραφές}$$

- * Πλήθος πλαστών υπογραφών ανά χρήστη:

$$n(SIG^{TR-}) = [0.7 \times n(SIG_{SK}^-)] = [16.8] = 17 \text{ υπογραφές}$$

- * Πλήθος ζευγών όμοιων υπογραφών:

$$n(p^{TR+}) = |S^{TR}| = \binom{n(SIG^{TR+})}{2} = \binom{17}{2} = 136 \text{ ζεύγη}$$

- * Πλήθος ζευγών ανόμοιων υπογραφών (p_{RF}^- είτε p_{SK}^-):

$$n(\mathbf{p}^{\text{TR}-}) = |D^{\text{TR}}| = \binom{n(\text{SIG}^{\text{TR}-})}{2} = \binom{17}{2} = 136 \text{ ζεύγη}$$

$$\text{Συνολικά: } n(\text{DEV}) \times |S^{\text{TR}}| = 28 \times 136 = 3808 \text{ ζεύγη υπογραφών}$$

– Για τη διαδικασία της *Αξιολόγησης* (Validation):

* Πλήθος αυθεντικών υπογραφών ανά χρήστη:

$$n(\text{SIG}^{\text{V}+}) = \lfloor 0.3 \times n(\text{SIG}^+) \rfloor = \lfloor 7.2 \rfloor = 7 \text{ υπογραφές}$$

* Πλήθος πλαστών υπογραφών ανά χρήστη:

$$n(\text{SIG}^{\text{V}-}) = \lfloor 0.3 \times n(\text{SIG}_{\text{SK}}^-) \rfloor = \lfloor 7.2 \rfloor = 7 \text{ υπογραφές}$$

* Πλήθος ζευγών όμοιων υπογραφών:

$$n(\mathbf{p}^{\text{V}+}) = |S^{\text{V}}| = \binom{n(\text{SIG}^{\text{V}+})}{2} = \binom{7}{2} = 21 \text{ ζεύγη}$$

* Πλήθος ζευγών ανόμοιων υπογραφών ($\mathbf{p}_{\text{RF}}^{\text{TR}-}$ είτε $\mathbf{p}_{\text{SK}}^{\text{TR}-}$):

$$n(\mathbf{p}^{\text{V}-}) = |D^{\text{V}}| = \binom{n(\text{SIG}^{\text{V}-})}{2} = \binom{7}{2} = 21 \text{ ζεύγη}$$

$$\text{Συνολικά: } n(\text{DEV}) \times |S^{\text{V}}| = 28 \times 21 = 588 \text{ ζεύγη υπογραφών}$$

- Πλήθος χρηστών για το στάδιο του *Ελέγχου* (Testing): $n(\text{TS}) = \lfloor 0.5 \times n(\text{W}) \rfloor = \lfloor 27.5 \rfloor = 27$ χρήστες

→ Βάση **Bangla**, $n(\text{W}) = 100$ χρήστες, $n(\text{SIG}^+) = 24$ αυθεντικές υπογραφές ανά χρήστη, $n(\text{SIG}_{\text{SK}}^-) = 30$ πλαστές υπογραφές ανά χρήστη:

- Πλήθος χρηστών για το στάδιο της *Ανάπτυξης* (Development): $n(\text{DEV}) = \lfloor 0.5 \times n(\text{W}) \rfloor = 50$ χρήστες

– Για τη διαδικασία της *Εκπαίδευσης* (Training):

* Πλήθος αυθεντικών υπογραφών ανά χρήστη:

$$n(\text{SIG}^{\text{TR}+}) = \lceil 0.7 \times n(\text{SIG}^+) \rceil = \lceil 16.8 \rceil = 17 \text{ υπογραφές}$$

* Πλήθος πλαστών υπογραφών ανά χρήστη:

$$n(\text{SIG}^{\text{TR}-}) = \lceil 0.7 \times n(\text{SIG}_{\text{SK}}^-) \rceil = 21 \text{ υπογραφές}$$

* Πλήθος ζευγών όμοιων υπογραφών:

$$n(\mathbf{p}^{\text{TR}+}) = |S^{\text{TR}}| = \binom{n(\text{SIG}^{\text{TR}+})}{2} = \binom{17}{2} = 136 \text{ ζεύγη}$$

* Πλήθος ζευγών ανόμοιων υπογραφών ($\mathbf{p}_{\text{RF}}^{\text{TR}-}$ είτε $\mathbf{p}_{\text{SK}}^{\text{TR}-}$), έπειτα από τυχαία δειγματοληψία:

$$n(\mathbf{p}^{\text{TR}-}) = |D^{\text{TR}}| = \binom{n(\text{SIG}^{\text{TR}-})}{2} = \binom{17}{2} = 136 \text{ ζεύγη}$$

$$\text{Συνολικά: } n(\text{DEV}) \times |S^{\text{TR}}| = 50 \times 136 = 6800 \text{ ζεύγη υπογραφών}$$

– Για τη διαδικασία της *Αξιολόγησης* (Validation):

* Πλήθος αυθεντικών υπογραφών ανά χρήστη:

$$n(\text{SIG}^{\text{V}+}) = \lfloor 0.3 \times n(\text{SIG}^+) \rfloor = \lfloor 7.2 \rfloor = 7 \text{ υπογραφές}$$

* Πλήθος πλαστών υπογραφών ανά χρήστη:

$$n(\text{SIG}^{\text{V}-}) = \lfloor 0.3 \times n(\text{SIG}_{\text{SK}}^-) \rfloor = \lfloor 7.2 \rfloor = 9 \text{ υπογραφές}$$

* Πλήθος ζευγών όμοιων υπογραφών:

$$n(\mathbf{p}^{\text{V}+}) = |S^{\text{V}}| = \binom{n(\text{SIG}^{\text{V}+})}{2} = \binom{7}{2} = 21 \text{ ζεύγη}$$

* Πλήθος ζευγών ανόμοιων υπογραφών ($\mathbf{p}_{\text{RF}}^{\text{TR}-}$ είτε $\mathbf{p}_{\text{SK}}^{\text{TR}-}$), έπειτα από τυχαία δειγματοληψία:

$$n(\mathbf{p}^{\text{V}-}) = |D^{\text{V}}| = \binom{n(\text{SIG}^{\text{V}-})}{2} = \binom{7}{2} = 21 \text{ ζεύγη}$$

Συνολικά: $n(\text{DEV}) \times |S^V| = 50 \times 21 = 1050$ ζεύγη υπογραφών

- Πλήθος χρηστών για το στάδιο του *Ελέγχου* (Testing): $n(\text{TS}) = \lfloor 0.5 \times n(\text{W}) \rfloor = 50$ χρήστες

Κατά το **δεύτερο πειραματικό σενάριο**, όπου η ανάπτυξη γίνεται με τους χρήστες της μίας Βάσης ενώ ο έλεγχος με την άλλη, τα σύνολα χωρίζονται με παρόμοιο τρόπο. Δηλαδή, για την κάθε Βάση, το σύνολο των υπογραφόντων χωρίζεται σε ακέραιο αριθμό κατά 70% και 30%, για τα στάδια της εκπαίδευσης και της αξιολόγησης αντίστοιχα. Από την άλλη, κατά τον έλεγχο της επίδοσης του ταξινομητή που κατασκευάστηκε, συμμετέχουν όλοι οι χρήστες της έτερης Βάσης. Παράδειγμα του διαχωρισμού περιγράφεται αναλυτικά παρακάτω, στην περίπτωση που η ανάπτυξη γίνεται με τη Βάση CEDAR, ενώ ο έλεγχος με την Bangla:

→ *ΑΝΑΠΤΥΞΗ*: Βάση **CEDAR**, $n(\text{W}) = 55$ χρήστες, $n(\text{SIG}^+) = 24$ αυθεντικές υπογραφές ανά χρήστη, $n(\text{SIG}_{SK}^-) = 24$ πλαστές υπογραφές ανά χρήστη

- Πλήθος χρηστών για το στάδιο της *Ανάπτυξης* (Development): $n(\text{DEV}) = n(\text{W}_{\text{CEDAR}}) = 55$ χρήστες

– Για τη διαδικασία της *Εκπαίδευσης* (Training):

- * Πλήθος αυθεντικών υπογραφών ανά χρήστη:
 $n(\text{SIG}^{\text{TR}+}) = \lceil 0.7 \times n(\text{SIG}^+) \rceil = \lceil 16.8 \rceil = 17$ υπογραφές
- * Πλήθος πλαστών υπογραφών ανά χρήστη:
 $n(\text{SIG}^{\text{TR}-}) = \lceil 0.7 \times n(\text{SIG}_{SK}^-) \rceil = \lceil 16.8 \rceil = 17$ υπογραφές
- * Πλήθος ζευγών όμοιων υπογραφών:
 $n(\mathbf{p}^{\text{TR}+}) = |S^{\text{TR}}| = \binom{n(\text{SIG}^{\text{TR}+})}{2} = \binom{17}{2} = 136$ ζεύγη
- * Πλήθος ζευγών ανόμοιων υπογραφών ($\mathbf{p}_{RF}^{\text{TR}-}$ είτε $\mathbf{p}_{SK}^{\text{TR}-}$):
 $n(\mathbf{p}^{\text{TR}-}) = |D^{\text{TR}}| = \binom{n(\text{SIG}^{\text{TR}-})}{2} = \binom{17}{2} = 136$ ζεύγη

Συνολικά: $n(\text{DEV}) \times |S^{\text{TR}}| = 55 \times 136 = 7480$ ζεύγη υπογραφών

– Για τη διαδικασία της *Αξιολόγησης* (Validation):

- * Πλήθος αυθεντικών υπογραφών ανά χρήστη:
 $n(\text{SIG}^{\text{V}+}) = \lfloor 0.3 \times n(\text{SIG}^+) \rfloor = \lfloor 7.2 \rfloor = 7$ υπογραφές
- * Πλήθος πλαστών υπογραφών ανά χρήστη:
 $n(\text{SIG}^{\text{V}-}) = \lfloor 0.3 \times n(\text{SIG}_{SK}^-) \rfloor = \lfloor 7.2 \rfloor = 7$ υπογραφές
- * Πλήθος ζευγών όμοιων υπογραφών:
 $n(\mathbf{p}^{\text{V}+}) = |S^{\text{V}}| = \binom{n(\text{SIG}^{\text{V}+})}{2} = \binom{7}{2} = 21$ ζεύγη
- * Πλήθος ζευγών ανόμοιων υπογραφών ($\mathbf{p}_{RF}^{\text{V}-}$ είτε $\mathbf{p}_{SK}^{\text{V}-}$):
 $n(\mathbf{p}^{\text{V}-}) = |D^{\text{V}}| = \binom{n(\text{SIG}^{\text{V}-})}{2} = \binom{7}{2} = 21$ ζεύγη

Συνολικά: $n(\text{DEV}) \times |S^{\text{V}}| = 55 \times 21 = 1155$ ζεύγη υπογραφών

→ *ΕΛΕΓΧΟΣ*: Βάση **Bangla**, $n(\text{W}) = 100$ χρήστες, $n(\text{SIG}^+) = 24$ αυθεντικές υπογραφές ανά χρήστη, $n(\text{SIG}_{SK}^-) = 30$ πλαστές υπογραφές ανά χρήστη

- Πλήθος χρηστών για το στάδιο του *Ελέγχου* (Testing): $n(\text{TS}) = n(\text{W}_{\text{Bangla}}) = 100$ χρήστες

5.5.2 Στάδιο Ελέγχου

Κατά τη διαδικασία του **ελέγχου** (testing) αξιολογείται το μοντέλο που έχει κατασκευαστεί στο το στάδιο της ανάπτυξης, από την ικανότητά του αναγνωρίζει σωστά, δεδομένα με τα οποία δεν έχει εκπαιδευτεί, όπως γίνεται σε μικρότερη κλίμακα κατά το στάδιο της αξιολόγησης. Από τη διαδικασία του ελέγχου, λόγω του γεγονότος ότι το πλήθος των δεδομένων είναι πολύ μεγαλύτερο από αυτό που αφιερώθηκε για την αξιολόγηση, το μοντέλο θα αποφέρει προβλέψεις που θα ταυτίζονται σε ένα μεγάλο βαθμό με το πως θα ανταποκρίνεται υπό πραγματικές συνθήκες λειτουργίας.

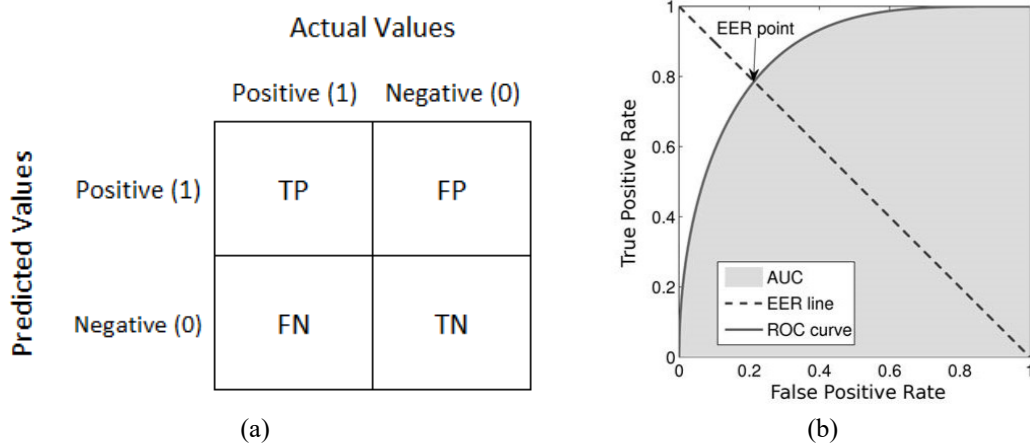
Προτού αναλυθεί η μεθοδολογία που ακολουθείται για τη διεξαγωγή των πειραμάτων, είναι αναγκαία η εισαγωγή στο **Πίνακα Σύγχυσης** (Confusion Matrix) και στη **ROC Ανάλυση**, όπως περιγράφονται στην αναφορά [128]. Η χρήση του Πίνακα Σύγχυσης σε εφαρμογές Ταξινόμησης είναι αρκετά διαδεδομένη, καθώς παρέχει ένα τρόπο καταγραφής της επίδοσης του ταξινομητή που αξιολογείται. Όπως φαίνεται και στην Εικόνα 15a, υπάρχουν δύο κλάσεις, η θετική (positive, 1) και η αρνητική (negative, 0) και τέσσερα πεδία, στα οποία αναγράφεται ο αριθμός των προβλέψεων που έγιναν ορθά ή λανθασμένα για τις εκάστοτε κλάσεις. Στα πεδία Ορθώς Θετικό (True Positive, TP) και Ορθώς Αρνητικό (True Negative, TN), καταγράφονται οι σωστές προβλέψεις για τα στοιχεία που προβλέφθηκαν ορθά ότι ανήκουν στην Κλάση 1 και την Κλάση 2 αντίστοιχα. Παρομοίως, στα πεδία Λανθασμένα Θετικό (False Positive, FP) και Λανθασμένα Αρνητικό (False Negative, FN), καταγράφεται ο αριθμός των φορών που έγιναν λανθασμένες προβλέψεις, για τις κλάσεις στις οποίες ανήκουν τα στοιχεία της Κλάσης 2 και 1 αντίστοιχα. Στις περιπτώσεις που γίνεται δυαδική ταξινόμηση (binary classification), όπως στην παρούσα εφαρμογή, από τις ποσότητες αυτές μπορούν να υπολογιστούν παράγωγες, οι οποίες χρησιμοποιούνται για τη χάραξη της λεγόμενης ROC (Receiver Operating Characteristics) καμπύλης, όπως φαίνεται στην Εικόνα 15b. Η ROC καμπύλη, όπως περιγράφεται στη δημοσίευση του Fawcett [128], παρουσιάζει τη συσχέτιση μεταξύ του ποσοστού της Ορθής Αποδοχής (True Positive Rate, TPR) και του ποσοστού της Εσφαλμένης Αποδοχής (False Positive rate, FPR). Οι ποσότητες αυτές υπολογίζονται με τη χρήση των Σχέσεων 39 και 40 αντίστοιχα.

$$TPR = \frac{TP}{TP + FN} \quad (39)$$

$$FPR = \frac{FP}{FP + TN} \quad (40)$$

Το **όριο απόφασης** (decision boundary), για ένα μοντέλο δυαδικής ταξινόμησης, είναι η τιμή-σκορ εκείνη, η οποία τίθεται με σκοπό να διαχωρίσει τα δεδομένα των δύο κλάσεων. Ανάλογα με την εφαρμογή, μπορεί να είναι πιο σημαντική η αποφυγή της Εσφαλμένης κατηγοριοποίησης δειγμάτων ως αρνητικά (FN), με το ρίσκο να υπάρξουν περισσότερα λανθασμένα θετικά (FP). Για παράδειγμα, στην περίπτωση που γινόταν ταξινόμηση πιθανών κρουσμάτων μιας θανατηφόρας ασθένειας θα ήταν σημαντικότερη η αποφυγή ψευδών αρνητικών διαγνώσεων, παρά ψευδών θετικών, για τον περιορισμό των κρουσμάτων. Η ROC καμπύλη, παρουσιάζει τις τιμές των TPR και FPR , κατά τη μεταβολή του ορίου απόφασης του ταξινομητή. Το βέλτιστο όριο απόφασης, είναι εκείνο που αποφέρει $TPR = 1$ και $FPR = 0$, δηλαδή αυτό που κατατάσσει τα δείγματα προς αναγνώριση πάντα στις σωστές τους κλάσεις.

Από την άλλη, σε εφαρμογές όπως αυτές της βιομετρικής αναγνώρισης, προτεραιότητα έχει η μείωση των πιθανοτήτων να υπάρξουν Λανθασμένες Εγκρίσεις (False Positive, FP) μη εξουσιοδοτημένων χρηστών, που με επιτυχία ξεγέλασαν το σύστημα βιομετρικής αναγνώρισης, ενώ παράλληλα η μείωση του ενδεχομένου της Εσφαλμένης Απόρριψης (False Negative, FN) ενός εξουσιοδοτημένου χρήστη, που υπέβαλε την αυθεντική υπογραφή του. Οι ποσότητες που εκφράζουν τα ποσοστά της Εσφαλμένης Αποδοχής και της Εσφαλμένης Απόρριψης, αποτελούν οι **False Acceptance Rate** (FAR) (Σχέση 41) και **False Rejection Rate** (FRR) (Σχέση 42) αντίστοιχα. Όπως αναφέρθηκε, η μετακίνηση του ορίου απόφασης, θα έχει ως αποτέλεσμα τη με-



Εικόνα 15: (a) Δομή Πίνακα Σύγκρισης σε περίπτωση δυαδικής ταξινόμησης²⁶, (b) Ενδεικτική αναπαράσταση ROC καμπύλης²⁷.

ταβολή του πλήθους των FP και FN κατά την ταξινόμηση. Το σημείο τοποθέτησης εκείνο, το οποίο αποφέρει την ισοβάθμιση των FAR και FRR , αποτελεί το λεγόμενο **Equal Error Rate** (EER), όπως ενδεικτικά παρουσιάζεται στην Εικόνα 15b. Η ποσότητα αυτή, χρησιμοποιείται για την αξιολόγηση της επίδοσης του ταξινομητή που κατασκευάζεται κατά την παρούσα διπλωματική εργασία και αποτελεί ένα συμβιβασμό μεταξύ της ασφάλειας του συστήματος και της ανοχής σε μεταβολές της όψης των υπογραφών των εξουσιοδοτημένων χρηστών.

$$FAR = \frac{FP}{TN + FP} \quad (41)$$

$$FRR = \frac{FN}{FN + TP} \quad (42)$$

Στη συνέχεια, αναλύεται η μέθοδος που υιοθετήθηκε για τη σύγκριση υπογραφών κατά το πειραματικό στάδιο. Ενώ η ανάπτυξη του μοντέλου AHSV γίνεται δημιουργώντας μια ολική αναπαράσταση για κάθε υπογραφή, κατά τη διαδικασία του Ελέγχου υιοθετείται μια διαφορετική μεθοδολογία για τη σύγκριση μεταξύ υπογραφών. Η τεχνική αυτή, όπως περιγράφεται στη δημοσίευση των Zois *et al.*[121] τεμαχίζει την εικόνα, στην οποία αναπαρίσταται η υπογραφή του χρήστη, σε περιοχές οι οποίες περιέχουν τον ίδιο αριθμό pixels (equimass regions). Αυτό επιτυγχάνεται με τη χρήση μιας ειδικά διαμορφωμένης υλοποίησης της χωρικής πυραμίδας (spatial pyramid), η οποία τεμαχίζει την εικόνα σε περιοχές μεταβλητών διαστάσεων, με γνώμονα τον αριθμό των pixel που υπάρχουν σε κάθε περιοχή. Με τη χρήση της τεχνικής αυτής, επισημαίνονται περισσότερο τοπικά στοιχεία της υπογραφής, κατασκευάζοντας έτσι μια πιο λεπτομερή αναπαράσταση.

Η πειραματική διαδικασία διεξάγεται ακολουθώντας τα δύο διαφορετικά πειραματικά σενάρια, για τα οποία μέχρι τώρα, έχει αναφερθεί το πως επηρεάζουν τη διαδικασία της Ανάπτυξης όσο αφορά τις κατανομές των δεδομένων. Σε αυτό το στάδιο, θα αναλυθεί η διαδικασία του Ελέγχου που ακολουθείται, που είναι κοινή για τα δύο αυτά σενάρια. Για το στάδιο του Ελέγχου, ακολουθείται μια διαδικασία, που βασίζεται στη δημοσίευση των Zois *et al.*[127]. Κατά τη μέθοδο αυτή, από το μοντέλο ζητείται να αναγνωρίσει το αν η υπογραφή ενός χρήστη του υποσυνόλου TS , ανήκει σε αυτόν, συγκρίνοντάς τη με δέκα αυθεντικές υπογραφές αναφοράς (reference) του.

²⁶Understanding Confusion Matrix - Towards Data Science

²⁷Tronci, Roberto & Giacinto, Giorgio & Roli, Fabio. (1970). Dynamic Score Combination: A Supervised and Unsupervised Score Combination Method. 163-177. 10.1007/978-3-642-03070-3_13.

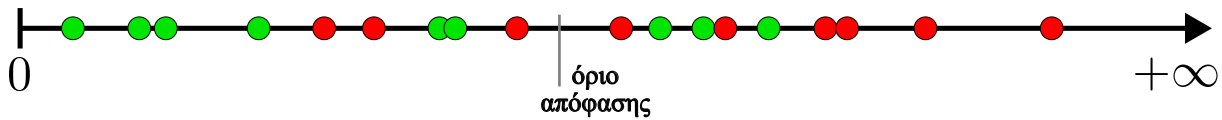
Συγκεκριμένα, για τον κάθε χρήστη του εκάστοτε συνόλου TS και κατά την κάθε επανάληψη του Ελέγχου του αλγόριθμου, επιλέγονται τυχαία δέκα αυθεντικές υπογραφές του οι οποίες ορίζονται ως υπογραφές αναφοράς (reference signatures - REF): $SIG^{REF} \subseteq SIG^{TS}$, με $n(SIG^{REF}) = 10$. Στη συνέχεια, κάθε μια από τις υπόλοιπες αυθεντικές και πλαστές υπογραφές που τον συνοδεύουν, πλήθους $n(SIG^{TS+}) - n(SIG^{REF}) + n(SIG^{TS-})$, χαρακτηρίζονται ως υπογραφές προς αναγνώριση (questioned signatures - Q) SIG^Q και συγκρίνονται μια τη φορά, με τις δέκα αυθεντικές του SIG^{REF} . Σε αντίθεση με τη διαδικασία της Ανάπτυξης, κατά τον Έλεγχο, για τη σύγκριση μεταξύ υπογραφών υπολογίζονται πέντε διαφορετικές αναπαραστάσεις, που εξάγονται όπως περιγράφεται στη συνέχεια. Εφόσον γίνει η προεπεξεργασία της εικόνας της κάθε υπογραφής προς σύγκριση SIG^Q , ακολουθεί ο τεμαχισμός της, σε τέσσερις περιοχές με ίσο αριθμό pixels (equimass regions), διαστάσεων 2×2 . Πλέον, κάθε υπογραφή, αναπαρίσταται από πέντε πίνακες συνδιακύμανσης που προκύπτουν από: α) ολόκληρη την εικόνα, $S_{1,1}$ και β) τις τέσσερις επιμέρους εικόνες $\{S_{2,j}\}_{j=1}^4$, που απέφερε ο τεμαχισμός της.

Για τη σύγκριση μίας υπογραφής προς αναγνώριση SIG^Q με τις υπογραφές αναφοράς SIG^{REF} , χρησιμοποιείται η σχέση της Point-to-Set and Set-to-Set απόστασης D^\ominus , που έχει προκύψει από το στάδιο της ανάπτυξης και οι πέντε αναπαραστάσεις των υπογραφών. Στη συνέχεια συγκρίνεται η (i, j) αναπαράσταση της άγνωστης υπογραφής $S_{i,j}^Q$, με κάθε μια από τις αντίστοιχες (i, j) αναπαραστάσεις, όλων των υπογραφών αναφοράς $\{S_{i,j}^{REF\#k}\}_{k=1}^{10}$, υπολογίζοντας την απόστασή τους. Για κάθε αναπαράσταση, εξάγονται $n(SIG^{REF}) = 10$ αποστάσεις $d_{i,j}$, από τις οποίες επιλέγεται η μικρότερη $d_{i,j}^{min}$, όπως φαίνεται στην ενδεικτική Σχέση 43. Συνολικά, για κάθε υπογραφή επιλέγονται πέντε ελάχιστες αποστάσεις $d_{i,j}^{min}$, από τις οποίες εξάγεται η μέση τους τιμή, όπως παρουσιάζεται ενδεικτικά στη Σχέση 44. Τέλος, η απόσταση εκείνη που προκύπτει από τη μέση τους τιμή, αναπαριστά την τελική απόσταση D_{SIG^Q} της υπό μελέτη υπογραφής SIG^Q .

$$\left. \begin{array}{l} D^\ominus(S_{1,1}^Q, S_{1,1}^{REF\#1}) \\ D^\ominus(S_{1,1}^Q, S_{1,1}^{REF\#2}) \\ \vdots \\ D^\ominus(S_{1,1}^Q, S_{1,1}^{REF\#10}) \end{array} \right\} \xrightarrow{\min(\cdot)} d_{1,1}^{min} \quad (43)$$

$$D_{SIG^Q} = \text{avg}(d_{1,1}^{min}, d_{2,1}^{min}, d_{2,2}^{min}, d_{2,3}^{min}, d_{2,4}^{min}) \quad (44)$$

Έτσι, για κάθε υπογραφή υπό εξέταση SIG^Q προκύπτει μια τελική απόσταση D_{SIG^Q} , που χαρακτηρίζει το πόσο απέχει από τις γνήσιες υπογραφές αναφοράς. Όσο μεγαλύτερη είναι η τιμή της, τόσο πιθανότερο είναι η υπογραφή αυτή να αποτελεί πλαστογραφία. Οι αποστάσεις αυτές, σχεδιάζονται ως σημεία σε ένα Καρτεσιανό επίπεδο όπως παρουσιάζεται στην Εικόνα 16 παρακάτω. Για τον υπολογισμό του EER για τον εκάστοτε χρήστη, μεταβάλλεται το όριο απόφασης, έτσι ώστε οι ποσότητες FAR και FRR , είναι ίσες. Η διαδικασία αυτή επαναλαμβάνεται δέκα φορές για όλους τους i χρήστες που ανήκουν στο σύνολο Ελέγχου TS, συγκεντρώνοντας δέκα διαφορετικά EER για τον κάθε ένα. Από αυτά, προκύπτει το μέσο $\overline{EER}_{W_i \in TS}$ κάθε W_i χρήστη. Τέλος, εξάγεται η μέση τιμή όλων των μέσων EER κάθε χρήστη του συνόλου Ελέγχου $\overline{EER}_W = \text{avg}(\{\overline{EER}_{W_i}\}_{i=1}^{n(TS)})$, η οποία εκφράζει με ένα ποσοτικό τρόπο, τις ακόλουθες πιθανότητες σφάλματος: α) το σύστημα αναγνώρισης να παρέχει λανθασμένα πρόσβαση σε ένα επιτήδριο, που πλαστογραφεί την υπογραφή ενός χρήστη ή β) να απορρίψει έναν εξουσιοδοτημένο χρήστη, του οποίου η υπογραφή διαφέρει περισσότερο από ότι συνήθως.



Εικόνα 16: Παράδειγμα απεικόνισης των αποστάσεων D_{SIG^Q} για ένα σύνολο υπογραφών προς εξέταση SIG^Q ενός χρήστη του συνόλου Ελέγχου. Οι υπογραφές αυτού του συνόλου αποτελούνται από γνήσιες και πλαστές υπογραφές, που συμβολίζονται ως πράσινες και κόκκινες κουκίδες αντίστοιχα. Επιπλέον, για το συγκεκριμένο παράδειγμα παρουσιάζεται το όριο απόφασης, για το οποίο ισχύει ότι $FAR = FRR$.

5.6 Αποτελέσματα

Για την εξαγωγή των αποτελεσμάτων της πειραματικής διαδικασίας αναπτύχθηκε κώδικας στη γλώσσα MATLAB, ο οποίος εκτελεί τις διαδικασίες της προεπεξεργασίας, της εξαγωγής χαρακτηριστικών και της κατασκευής πινάκων συνδιακύμανσης. Επιπλέον, τροποποιήθηκε ο βασικός κώδικας του αλγόριθμου Point-to-Set and Set-to-Set Distance (PSSSD) και πάλι στη γλώσσα MATLAB, για τις διαδικασίες (πρωτόκολλα) της Εκπαίδευσης, της Αξιολόγησης και του Ελέγχου, των μοντέλων που κατασκευάστηκαν με αυτόν. Συνοψίζοντας, η παρούσα διπλωματική εργασία τροποποιεί την αρχική έκδοση του PSSSD, που βρίσκεται εδώ: <https://github.com/zhigao2017/PSSSD>. Στη συνέχεια παρουσιάζονται οι πίνακες που περιέχουν τα αποτελέσματα των πειραματικών διαδικασιών, οι οποίοι κατασκευάζονται ως εξής: Κάθε ξεχωριστός πίνακας, περιέχει τα αποτελέσματα που έχουν προκύψει για την εκτέλεση της διαδικασίας της Εκπαίδευσης, με τις εκάστοτε παραμέτρους m και p , που συμβολίζουν το πλήθος και το μέγεθος των προβολών $\{\mathbf{X}_i^k\}_{k=1}^m \in Sym_p^+$ που κατασκευάστηκαν, για την αναπαράσταση κάθε αρχικού πίνακα συνδιακύμανσης $\mathbf{X}_i \in Sym_n^+$ του αλγόριθμου PSSSD. Επίσης, κάθε στήλη με τίτλο: “**Ανάπτυξη με**” περιγράφει το είδος πλαστογραφιών, τυχαίες (RF) ή επιτηδευμένες (SK), με τις οποίες συζεύχτηκαν οι γνήσιες υπογραφές, κατά τη διαδικασία της Εκπαίδευσης. Κάθε ξεχωριστό κελί περιέχει το μέσο $\overline{EER}_W\%$, που έχει προκύψει από τις δέκα επαναλήψεις της εκτέλεσης του Ελέγχου, κατά τη διαδικασία που αναλύθηκε παραπάνω. Επιπλέον, όπως παρατηρείται για κάθε στήλη με τίτλο: “**Ανάπτυξη με**”, καταγράφονται τέσσερα αποτελέσματα. Το κάθε ένα από αυτά, προκύπτει από την τήρηση του ενός από τα δύο πειραματικά σενάρια που ορίστηκαν στο Κεφάλαιο 5.5.1. Τα στοιχεία της νοητής κύριας διαγώνιου του κάθε πίνακα, είναι τα αποτελέσματα της διαδικασίας, όταν τηρήθηκε το πρώτο πειραματικό σενάριο. Δηλαδή αυτού κατά το οποίο οι διαδικασίες της Εκπαίδευσης, Αξιολόγησης και του Ελέγχου, πραγματοποιούνται αποκλειστικά με τα δεδομένα της ίδιας Βάσης. Τα στοιχεία εκτός της κύριας διαγώνιου, περιέχουν τα αποτελέσματα που λήφθηκαν όταν τηρήθηκε το δεύτερο πειραματικό σενάριο, εκείνο το οποίο υλοποιεί τη Μεταφορά Μάθησης, όπου η Ανάπτυξη έγινε με τους χρήστες της μίας Βάσης, ενώ ο Έλεγχος, με την έτερη. Συγκεκριμένα, για το πάνω δεξιό κελί η Ανάπτυξη έγινε με τη Βάση Bangla και ο Έλεγχος με τη CEDAR, ενώ για το κάτω αριστερό κελί έγινε το αντίστροφο.

$m = 1, p = 6$	Ανάπτυξη με: p_{RF}^-		Ανάπτυξη με: p_{SK}^-	
Έλεγχος	CEDAR	Bangla	CEDAR	Bangla
CEDAR	0.41	0.76	0.33	0.33
Bangla	0.52	0.54	0.65	0.74

$m = 1, p = 7$	Ανάπτυξη με: p_{RF}^-		Ανάπτυξη με: p_{SK}^-	
Έλεγχος	CEDAR	Bangla	CEDAR	Bangla
CEDAR	0.28	0.41	0.14	0.32
Bangla	0.58	0.44	0.55	0.56

$m = 1, p = 8$	Ανάπτυξη με: p_{RF}^-		Ανάπτυξη με: p_{SK}^-	
Έλεγχος	CEDAR	Bangla	CEDAR	Bangla
CEDAR	0.11	0.55	0.17	0.33
Bangla	0.54	0.51	0.6	0.55

$m = 1, p = 9$	Ανάπτυξη με: p_{RF}^-		Ανάπτυξη με: p_{SK}^-	
Έλεγχος	CEDAR	Bangla	CEDAR	Bangla
CEDAR	0.13	0.44	0.13	0.33
Bangla	0.46	0.5	0.63	0.59

$m = 1, p = 10$	Ανάπτυξη με: p_{RF}^-		Ανάπτυξη με: p_{SK}^-	
Έλεγχος	CEDAR	Bangla	CEDAR	Bangla
CEDAR	0.12	0.24	0.18	0.2
Bangla	0.47	0.49	0.54	0.61

$m = 1, p = 11$	Ανάπτυξη με: p_{RF}^-		Ανάπτυξη με: p_{SK}^-	
Έλεγχος	CEDAR	Bangla	CEDAR	Bangla
CEDAR	0.11	0.19	0.16	0.19
Bangla	0.45	0.55	0.45	0.56

$m = 1, p = 12$	Ανάπτυξη με: p_{RF}^-		Ανάπτυξη με: p_{SK}^-	
Έλεγχος	CEDAR	Bangla	CEDAR	Bangla
CEDAR	0.17	0.17	0.17	0.17
Bangla	0.45	0.56	0.45	0.56

Πίνακας 3: Τα $\overline{EER}_W\%$, για τις περιπτώσεις που το πλήθος των προβολών των SPD πινάκων, κατά το στάδιο της Ανάπτυξης, ισούνταν με $m = 1$.

$m = 2, p = 6$	Ανάπτυξη με: p_{RF}^-		Ανάπτυξη με: p_{SK}^-	
Έλεγχος	CEDAR	Bangla	CEDAR	Bangla
CEDAR	0.22	0.53	0.19	0.36
Bangla	0.49	0.71	0.56	0.58

$m = 2, p = 5$	Ανάπτυξη με: p_{RF}^-		Ανάπτυξη με: p_{SK}^-	
Έλεγχος	CEDAR	Bangla	CEDAR	Bangla
CEDAR	0.33	0.58	0.12	0.32
Bangla	0.50	0.60	0.58	0.26

$m = 2, p = 4$	Ανάπτυξη με: p_{RF}^-		Ανάπτυξη με: p_{SK}^-	
Έλεγχος	CEDAR	Bangla	CEDAR	Bangla
CEDAR	0.4	0.67	0.4	0.34
Bangla	0.51	0.76	0.53	0.69

$m = 2, p = 3$	Ανάπτυξη με: p_{RF}^-		Ανάπτυξη με: p_{SK}^-	
Έλεγχος	CEDAR	Bangla	CEDAR	Bangla
CEDAR	0.49	0.59	0.2	0.39
Bangla	0.81	0.71	0.68	0.96

$m = 3, p = 4$	Ανάπτυξη με: p_{RF}^-		Ανάπτυξη με: p_{SK}^-	
Έλεγχος	CEDAR	Bangla	CEDAR	Bangla
CEDAR	0.3	0.47	0.1	0.47
Bangla	0.58	0.77	0.51	0.81

$m = 3, p = 3$	Ανάπτυξη με: p_{RF}^-		Ανάπτυξη με: p_{SK}^-	
Έλεγχος	CEDAR	Bangla	CEDAR	Bangla
CEDAR	0.42	0.53	0.35	0.49
Bangla	0.65	0.73	0.55	0.81

$m = 4, p = 3$	Ανάπτυξη με: p_{RF}^-		Ανάπτυξη με: p_{SK}^-	
Έλεγχος	CEDAR	Bangla	CEDAR	Bangla
CEDAR	0.42	0.46	0.27	0.61
Bangla	0.53	0.63	0.49	0.67

Πίνακας 4: Τα $\overline{EER}_W\%$, για τις περιπτώσεις που το πλήθος των προβολών των SPD πινάκων, κατά το στάδιο της Ανάπτυξης, ήταν $m > 1$.

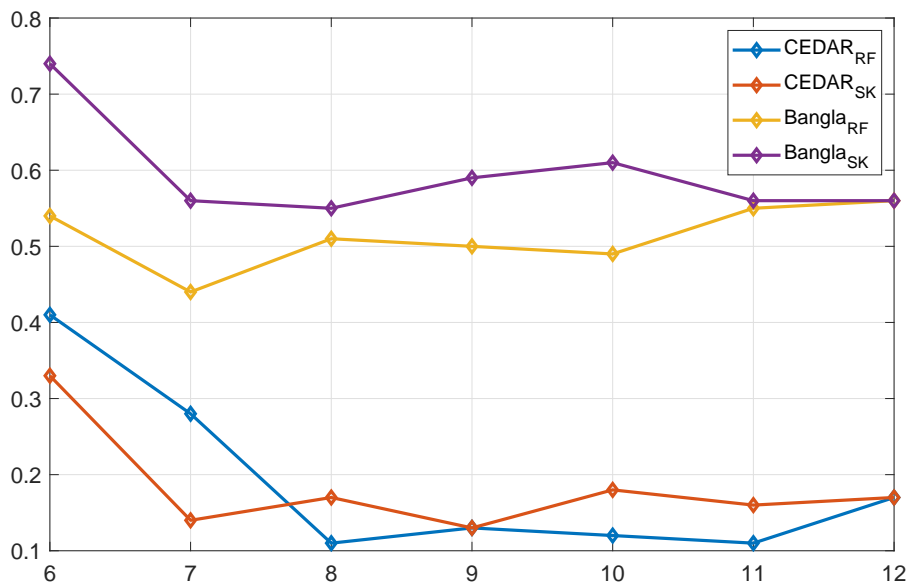
5.7 Σχολιασμός Αποτελεσμάτων

Ερμηνεύοντας τα αποτελέσματα που παρουσιάζονται στους Πίνακες 3 και 4 παραπάνω, επιχειρούμε να εξάγουμε τα ακόλουθα συμπεράσματα:

1. Η συσχέτιση μεταξύ της επίδοσης του αλγόριθμου κατά την ταξινόμηση όμοιων και ανόμοιων υπογραφών, με τη μεταβολή του πλήθους και του μεγέθους των προβολών του.

2. **Η επίδοση του αλγόριθμου**, όταν για την εκπαίδευσή του τα ανόμοια ζεύγη υπογραφών αποτελούν **τυχαίες πλαστογραφίες (RF)**, έναντι **επιτηδευμένων πλαστογραφιών (SK)**, και το αντίστροφο.
3. **Η επίδοση του αλγόριθμου σε σενάριο Μεταφοράς Μάθησης (Transfer Learning)**, όπου η Εκπαίδευσή του γίνεται με τη μία Βάση, ενώ ο Έλεγχός του με την άλλη.

Όπως υποστηρίζεται στη δημοσίευση των Gao *et al.*[57], συνδυάζοντας τη χρήση πολλαπλών m προβολών, αυξημένων διαστάσεων p , αλλά μικρότερων σε σύγκριση με τις αρχικές ($m \times p < n$), ο αλγόριθμος είναι πιο ακριβής, καθώς απορρίπτεται η περιττή πληροφορία που προέρχεται από την αρχική αναπαράσταση των δεδομένων. Παρόλο που δεν είναι εφικτή η μεγάλη μεταβολή του m , λόγω των ήδη μειωμένων διαστάσεων του αρχικού πίνακα συνδιακύμανσης που υλοποιείται σε αυτή τη διπλωματική εργασία, αναδεικνύεται το γεγονός ότι, η μείωση των διαστάσεων της αρχικής αναπαράστασης κατά ένα μικρό βαθμό, αποφέρει μεγαλύτερη ακρίβεια, σε σύγκριση με αυτό που παρουσιάζεται κατά τη μεγάλη ελάττωση. Μάλιστα, μέσω της Εικόνας 17 προσδιορίζεται ότι για $10 < p < 12$, επιτυγχάνεται η μεγαλύτερη μείωση του $\overline{EER}_W\%$, για κάθε μια από τις τέσσερις περιπτώσεις εκπαίδευσης.



Εικόνα 17: Η μεταβολή της τιμής του $\overline{EER}_W\%$, κατά την αύξηση του μεγέθους p των προβολών, για $m = 1$, χρησιμοποιώντας ζεύγη τυχαίων (RF) και επιτηδευμένων (SK) πλαστογραφιών, των δύο Βάσεων, κατά το στάδιο της Ανάπτυξης, ακολουθώντας το πρώτο πειραματικό σενάριο.

Με τη μελέτη του Πίνακα 5 συμπεραίνουμε ότι για τη συγκεκριμένη πειραματική διάταξη, η εκπαίδευση του αλγόριθμου με $m = 1$, δηλαδή με μόνο μια προβολή $X_i^1 \in Sym_p^+$, αποφέρει κατά μέσο όρο καλύτερα αποτελέσματα, σε σύγκριση με την περίπτωση που χρησιμοποιούνται παραπάνω από μια προβολές ($m > 1$). Επιπλέον, από την Εικόνα 17 παρατηρείται ότι, ενώ υπάρχει μείωση του σφάλματος $\overline{EER}_W\%$, όταν τα ανόμοια ζεύγη που χρησιμοποιούνται κατά το στάδιο της Ανάπτυξης αποτελούνται από γνήσιες υπογραφές και επιτηδευμένες πλαστογραφίες (p_{SK}^-), η μείωση αυτή κατά μέσο όρο δεν είναι σε μεγάλο βαθμό, όπως παρουσιάζεται στο

Πίνακα 5. Το γεγονός αυτό είναι σημαντικό, γιατί υπό κανονικές συνθήκες εάν η εκπαίδευση ενός συστήματος αναγνώρισης υπογραφών γίνει εκ νέου, χρησιμοποιώντας τις υπογραφές πραγματικών χρηστών, θα είναι και πάλι ακριβές, παρ' όλο που τα ζεύγη ανόμοιων υπογραφών που θα χρησιμοποιηθούν, θα αποτελούνται από τυχαίες πλαστογραφίες (p_{RF}^-).

		Ανάπτυξη με: p_{RF}^-	Ανάπτυξη με: p_{SK}^-
CEDAR	$m = 1$	0.19	0.18
	$m > 1$	0.37	0.23
Bangla	$m = 1$	0.51	0.6
	$m > 1$	0.7	0.68

Πίνακας 5: Μέση τιμή των $\overline{EER}_W\%$ κάθε πειραματικής διαδικασίας του πρώτου σεναρίου, για σταθερό ($m = 1$) και μεταβαλλόμενο ($m > 1$) πλήθος προβολών.

Αναπτύσσοντας το προηγούμενο συμπέρασμα, από το Πίνακα 6 παρουσιάζεται ότι με τη σωστή επιλογή της Βάσης με την οποία θα γίνει η εκπαίδευση, είναι αρκετά πιθανό ο αλγόριθμος που θα εκπαιδευτεί με εκείνη, να μπορέσει να χρησιμοποιηθεί και για την αναγνώριση άλλου τύπου υπογραφών, όπως σε αυτή την περίπτωση, που οι δύο Βάσεις Δεδομένων περιέχουν υπογραφές διαφορετικού αλφάβητου. Συγκεκριμένα, εάν στην παρούσα περίπτωση χρησιμοποιηθεί η Βάση Bangla για την Ανάπτυξη του μοντέλου AHSV, όπως παρουσιάζεται από το $\overline{EER}_W\%$, η επίδοση του αλγορίθμου στην αναγνώριση υπογραφών της CEDAR, θα είναι αρκετά καλή, και μάλιστα κοντά στις τιμές εκείνες που προκύπτουν όταν τηρείται το πρώτο πειραματικό σενάριο, με την εκπαίδευση υπογραφών της CEDAR.

C: CEDAR, B: Bangla		Ανάπτυξη με: p_{RF}^-	Ανάπτυξη με: p_{SK}^-
C $\xrightarrow[\text{με τη}]{\text{Έλεγχος}}$ B	$m = 1$	0.5	0.55
	$m > 1$	0.58	0.57
B $\xrightarrow[\text{με τη}]{\text{Έλεγχος}}$ C	$m = 1$	0.39	0.27
	$m > 1$	0.55	0.41

Πίνακας 6: Μέση τιμή των $\overline{EER}_W\%$ κάθε πειραματικής διαδικασίας του δεύτερου σεναρίου, για σταθερό ($m = 1$) και μεταβαλλόμενο ($m > 1$) πλήθος προβολών.

6 Συμπεράσματα

Παρατηρώντας την επιτυχία του αλγόριθμου “Point-to-Set and Set-to-Set Distance” [57] σε άλλες εφαρμογές της Υπολογιστικής Όρασης, όπως αυτές της αναγνώρισης αντικειμένων, προσώπων και κινήσεων, επιχειρήθηκε η χρήση του στην επιστημονική πρόκληση της κατασκευής συστήματος ταξινόμησης χειρόγραφων υπογραφών. Έτσι, με τη χρήση δύο διαφορετικών Βάσεων Δεδομένων, η εφαρμογή του αλγόριθμου αποδείχθηκε επιτυχής στη σύνθεση ενός αξιόπιστου συστήματος βιομετρικής αναγνώρισης ατόμων, μέσω της Στατικής (Offline) ανάλυσης των χειρόγραφων υπογραφών τους, χρησιμοποιώντας ένα γενικό Writer Independent ταξινομητή. Περαιτέρω, το σύστημα αυτό μπορεί να χρησιμοποιηθεί για την ταξινόμηση και άλλου είδους υπογραφών, πιθανών και εντελώς άλλων αντικειμένων σε σενάριο Μεταφοράς Μάθησης (Transfer Learning) με την ίδια σχεδόν ακρίβεια, γεγονός που το καθιστά ευπροσάρμοστο και για άλλες εφαρμογές Αναγνώρισης Προτύπων. Έχοντας εξετάσει τις ικανότητες του προτεινόμενου αλγόριθμου σε μια τόσο απαιτητική εφαρμογή, συμπεραίνουμε ότι θα μπορέσει να ανταποκριθεί με παρόμοια επιτυχία και σε άλλες προκλήσεις που αφορούν την Αναγνώριση Προτύπων και την Υπολογιστική Όραση.

Αναφορές

- [1] Wikipedia contributors, *Artificial intelligence* — *Wikipedia, The Free Encyclopedia*, [Online; accessed 11-November-2023], 2023. διεύθυν.: https://en.wikipedia.org/w/index.php?title=Artificial_intelligence&oldid=1184456640.
- [2] Σέργιος Θεοδωρίδης, Κωνσταντίνος Κουτρούμπας, *Αναγνώριση Προτύπων*. Ιατρικές Εκδόσεις Π. Χ. Πασχαλίδης, 2011, ISBN: 978-960-489-145-0.
- [3] C. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Οκτ. 2007, ISBN: 0387310738.
- [4] Βικιπαίδεια, *Δεδομένα* — *Βικιπαίδεια, Η Ελεύθερη Εγκυκλοπαίδεια*, [Online; accessed 9-Νοεμβρίου-2023], 2022. διεύθυν.: <https://el.wikipedia.org/w/index.php?title=%CE%94%CE%B5%CE%B4%CE%BF%CE%BC%CE%AD%CE%BD%CE%B1&oldid=9736941>.
- [5] *Statistical terms and concepts*. διεύθυν.: <https://www.abs.gov.au/statistics/understanding-statistics/statistical-terms-and-concepts>.
- [6] Wikipedia contributors, *Multi-label classification* — *Wikipedia, The Free Encyclopedia*, [Online; accessed 12-November-2023], 2023. διεύθυν.: https://en.wikipedia.org/w/index.php?title=Multi-label_classification&oldid=1178261136.
- [7] GeeksforGeeks, *An introduction to MultiLabel classification*, Ιούλ. 2020. διεύθυν.: <https://www.geeksforgeeks.org/an-introduction-to-multilabel-classification/>.
- [8] F. Villegas, *Categorical Data: Definition, Types, features + Examples*, en-US, Αύγ. 2023. διεύθυν.: <https://www.questionpro.com/blog/categorical-data/>.
- [9] Βικιλεξικό, *γνώρισμα* — *Βικιλεξικό, Η Ελεύθερη Εγκυκλοπαίδεια*, [Online; accessed 15-Νοεμβρίου-2023], 2022. διεύθυν.: <https://el.wiktionary.org/w/index.php?title=%CE%B3%CE%BD%CF%8E%CF%81%CE%B9%CF%83%CE%BC%CE%B1&oldid=5464498>.
- [10] *Feature Vector*, en-us, [Online; accessed 16-November-2023]. διεύθυν.: <https://brilliant.org/wiki/feature-vector/>.
- [11] R. Szeliski, *Computer Vision Algorithms and Applications (Texts in Computer Science)*, eng, 2nd ed. 2022. Cham: Springer International Publishing, 2022, ISBN: 9783030343729.
- [12] S. Theodoridis, *Machine Learning: A Bayesian and Optimization Perspective*. Elsevier Science, 2020, ISBN: 9780128188040. διεύθυν.: <https://books.google.gr/books?id=l-nEDwAAQBAJ>.
- [13] *A guide to the types of machine learning algorithms and their applications*, en-GB. διεύθυν.: https://www.sas.com/en_gb/insights/articles/analytics/machine-learning-algorithms.html.
- [14] R. Duda, P. Hart και D. Stork, *Pattern Classification*. Wiley, 2012, ISBN: 9781118586006. διεύθυν.: <https://books.google.gr/books?id=Br33IRC3PkQC>.
- [15] K. Fukunaga, *Introduction to Statistical Pattern Recognition (Computer science and scientific computing)*. Elsevier Science, 2013, ISBN: 9780080478654. διεύθυν.: <https://books.google.gr/books?id=BIJZTGjTxBgC>.
- [16] A. P. Witkin, “Scale-Space Filtering”, στο *International Joint Conference on Artificial Intelligence*, 1983. διεύθυν.: <https://api.semanticscholar.org/CorpusID:7096897>.

- [17] J. Koenderink, “The structure of images: 1984–2021”, *Biological Cybernetics*, τόμ. 115, Μαρ. 2021. DOI: [10.1007/s00422-021-00870-0](https://doi.org/10.1007/s00422-021-00870-0).
- [18] 7. *Scale Space*. διεύθν.: <https://staff.fnwi.uva.nl/r.vandenboomgaard/IPCV20172018/LectureNotes/IP/ScaleSpace/index.html>.
- [19] E. Adelson, C. Anderson, J. Bergen, P. Burt και J. Ogden, “Pyramid Methods in Image Processing”, *RCA Eng.*, τόμ. 29, Νοέ. 1983.
- [20] D. B. Percival και D. Mondal, “22 - A Wavelet Variance Primer”, στο *Time Series Analysis: Methods and Applications*, σειρά Handbook of Statistics, T. Subba Rao, S. Subba Rao και C. Rao, επιμελητές, τόμ. 30, Elsevier, 2012, σσ. 623–657. DOI: <https://doi.org/10.1016/B978-0-444-53858-1.00022-3>. διεύθν.: <https://www.sciencedirect.com/science/article/pii/B9780444538581000223>.
- [21] K. Chandrasekaran, “2D-Discrete Wavelet Transformation and its applications in Digital Image Processing using MATLAB”, *Medium*, Ιαν. 2022. διεύθν.: <https://medium.com/@koushikc2000/2d-discrete-wavelet-transformation-and-its-applications-in-digital-image-processing-using-matlab-1f5c68672de3>.
- [22] J. Canny, “A Computational Approach to Edge Detection”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, τόμ. PAMI-8, αρθμ. 6, σσ. 679–698, 1986. DOI: [10.1109/TPAMI.1986.4767851](https://doi.org/10.1109/TPAMI.1986.4767851).
- [23] *Structure from Motion Overview*. διεύθν.: <https://www.mathworks.com/help/vision/ug/structure-from-motion.html>.
- [24] S. Ullman, “The interpretation of structure from motion”, *Proceedings of the Royal Society of London. Series B. Biological Sciences*, τόμ. 203, σσ. 405–426, 1979. διεύθν.: <https://api.semanticscholar.org/CorpusID:11995230>.
- [25] *What is photogrammetry - Artec3D*. διεύθν.: <https://www.artec3d.com/learning-center/what-is-photogrammetry>.
- [26] Y. LeCun, Y. Bengio και G. Hinton, “Deep Learning”, *Nature*, τόμ. 521, σσ. 436–44, Μάι. 2015. DOI: [10.1038/nature14539](https://doi.org/10.1038/nature14539).
- [27] A. Krizhevsky, I. Sutskever και G. E. Hinton, “ImageNet Classification with Deep Convolutional Neural Networks”, στο *Advances in Neural Information Processing Systems*, F. Pereira, C. Burges, L. Bottou και K. Weinberger, επιμελητές, τόμ. 25, Curran Associates, Inc., 2012. διεύθν.: https://proceedings.neurips.cc/paper_files/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf.
- [28] *What Is SLAM?* Διεύθν.: <https://www.mathworks.com/discovery/slam.html>.
- [29] M. O. Aqel, M. H. Marhaban, M. I. Saripan και N. B. Ismail, “Review of visual odometry: types, approaches, challenges, and applications”, *SpringerPlus*, τόμ. 5, σσ. 1–26, 2016.
- [30] G. Boesch, *The 100 most popular Computer Vision applications in 2024*, en-US, Νοέ. 2023. διεύθν.: <https://viso.ai/applications/computer-vision-applications/>.
- [31] Ιούλ. 2022. διεύθν.: <https://premioinc.com/blogs/blog/what-is-computer-vision-for-medical-imaging>.
- [32] G. Boesch, *Top 18 applications of Computer Vision in security and Surveillance*, en-US, Ιαν. 2023. διεύθν.: <https://viso.ai/applications/computer-vision-applications-in-surveillance-and-security/>.

- [33] Διεύθυν.: https://www.tesla.com/ownersmanual/model3/en_us/GUID-32E9F9FD-0014-4EB4-8D96-A8BE99DBE1A2.html.
- [34] Διεύθυν.: https://www.tesla.com/ownersmanual/model3/en_us/GUID-0D9B548B-D1A4-494B-8C5F-C4360304D99F.html.
- [35] Διεύθυν.: <https://www.dji.com/gr/guidance/features>.
- [36] C. Cortes και V. N. Vapnik, “Support-vector networks”, *Machine Learning*, τόμ. 20, σσ. 273–297, 2004. διεύθυν.: <https://api.semanticscholar.org/CorpusID:206787478>.
- [37] E. Fix και J. L. Hodges, “Discriminatory analysis. Nonparametric discrimination: Consistency properties”, *International Statistical Review/Revue Internationale de Statistique*, τόμ. 57, αρθμ. 3, σσ. 238–247, 1989.
- [38] T. M. Cover και P. E. Hart, “Nearest neighbor pattern classification”, *IEEE Trans. Inf. Theory*, τόμ. 13, σσ. 21–27, 1967. διεύθυν.: <https://api.semanticscholar.org/CorpusID:5246200>.
- [39] N. Friedman, D. Geiger και M. Goldszmidt, “Bayesian Network Classifiers”, *Machine Learning*, τόμ. 29, σσ. 131–163, Νοέ. 1997. DOI: [10.1023/A:1007465528199](https://doi.org/10.1023/A:1007465528199).
- [40] L. Breiman, J. Friedman, C. Stone και R. Olshen, *Classification and Regression Trees*. Taylor & Francis, 1984, ISBN: 9780412048418. διεύθυν.: <https://books.google.gr/books?id=JwQx-W0mSyQC>.
- [41] J. Tabak, *Geometry: The Language of Space and Form* (Facts on File math library). Facts On File, Incorporated, 2014, ISBN: 9780816068760. διεύθυν.: <https://books.google.gr/books?id=r0HuPiexnYwC>.
- [42] I. Wolfram Research, *L2-Norm – from Wolfram MathWorld*. διεύθυν.: <https://mathworld.wolfram.com/L2-Norm.html>.
- [43] Wikipedia contributors, *Backpropagation — Wikipedia, The Free Encyclopedia*, [Online; accessed 14-December-2023], 2023. διεύθυν.: <https://en.wikipedia.org/w/index.php?title=Backpropagation&oldid=1189477213>.
- [44] L. Bottou και O. Bousquet, “The tradeoffs of large scale learning”, *Advances in neural information processing systems*, τόμ. 20, 2007.
- [45] *What is gradient descent? - IBM*. διεύθυν.: <https://www.ibm.com/topics/gradient-descent>.
- [46] L. Bottou, “Online Algorithms and Stochastic Approximations”, στο *Online Learning and Neural Networks*, D. Saad, επιμελητής, revised, oct 2012, Cambridge, UK: Cambridge University Press, 1998. διεύθυν.: <http://leon.bottou.org/papers/bottou-98x>.
- [47] Wikipedia contributors, *Line (geometry) — Wikipedia, The Free Encyclopedia*, [Online; accessed 3-July-2023], 2023. διεύθυν.: [https://en.wikipedia.org/w/index.php?title=Line_\(geometry\)&oldid=1158843583](https://en.wikipedia.org/w/index.php?title=Line_(geometry)&oldid=1158843583).
- [48] J. Lee, *Introduction to Smooth Manifolds* (Graduate Texts in Mathematics). Springer, 2003, ISBN: 9780387954486. διεύθυν.: <https://books.google.gr/books?id=eqfgZtjQceYC>.
- [49] Wikipedia contributors, *Einstein notation — Wikipedia, The Free Encyclopedia*, [Online; accessed 19-September-2023], 2023. διεύθυν.: https://en.wikipedia.org/w/index.php?title=Einstein_notation&oldid=1158848214.

- [50] Βικιπαίδεια, *Σχέση ισοδυναμίας* — Βικιπαίδεια, Η Ελεύθερη Εγκυκλοπαίδεια, [Online; accessed 20-Οκτωβρίου-2023], 2022. διεύθν.: https://el.wikipedia.org/w/index.php?title=%CE%A3%CF%87%CE%AD%CF%83%CE%B7_%CE%B9%CF%83%CE%BF%CE%B4%CF%85%CE%BD%CE%B1%CE%BC%CE%AF%CE%B1%CF%82&oldid=9382667.
- [51] Wikipedia contributors, *Smoothness* — Wikipedia, The Free Encyclopedia, [Online; accessed 21-September-2023], 2023. διεύθν.: <https://en.wikipedia.org/w/index.php?title=Smoothness&oldid=1176043358>.
- [52] qncubed3. “Manifolds #5: Tangent Space (part 1)”. (16 Ιαν. 2023), διεύθν.: <https://www.youtube.com/watch?v=rWSoPR8j6Gg>.
- [53] qncubed3. “Manifolds #6: Tangent Space (part 2)”. (11 Ιούλ. 2023), διεύθν.: <https://www.youtube.com/watch?v=0V2maKTi0Rw>.
- [54] Wikipedia contributors, *Tangent bundle* — Wikipedia, The Free Encyclopedia, [Online; accessed 6-November-2023], 2023. διεύθν.: https://en.wikipedia.org/w/index.php?title=Tangent_bundle&oldid=1174138843.
- [55] Wikipedia contributors, *Disjoint union* — Wikipedia, The Free Encyclopedia, [Online; accessed 6-November-2023], 2023. διεύθν.: https://en.wikipedia.org/w/index.php?title=Disjoint_union&oldid=1159895839.
- [56] Βικιλεξικό, *disjoint union* — Βικιλεξικό, Η Ελεύθερη Εγκυκλοπαίδεια, 2021. διεύθν.: https://el.wiktionary.org/w/index.php?title=disjoint_union&oldid=5259270.
- [57] Z. Gao, Y. Wu, M. Harandi και Y. Jia, “A Robust Distance Measure for Similarity-Based Classification on the SPD Manifold”, *IEEE Transactions on Neural Networks and Learning Systems*, τόμ. 31, αρθμ. 9, σσ. 3230–3244, 2020. DOI: [10.1109/TNNLS.2019.2939177](https://doi.org/10.1109/TNNLS.2019.2939177).
- [58] Wikipedia contributors, *Stiefel manifold* — Wikipedia, The Free Encyclopedia, [Online; accessed 22-October-2023], 2023. διεύθν.: https://en.wikipedia.org/w/index.php?title=Stiefel_manifold&oldid=1137830694.
- [59] Διεύθν.: https://noodle.med.yale.edu/hdtag/notes/steifel_notes.pdf.
- [60] T. Bendokat, R. Zimmermann και P. A. Absil, *A Grassmann Manifold Handbook: Basic Geometry and Computational Aspects*, 2020. arXiv: [2011.13699](https://arxiv.org/abs/2011.13699) [math.NA].
- [61] Διεύθν.: <https://juliamanifolds.github.io/Manifolds.jl/latest/manifolds/grassmann.html>.
- [62] Wikipedia contributors, *Logarithm of a matrix* — Wikipedia, The Free Encyclopedia, [Online; accessed 5-November-2023], 2023. διεύθν.: https://en.wikipedia.org/w/index.php?title=Logarithm_of_a_matrix&oldid=1178957044.
- [63] Wikipedia contributors, *Natural logarithm* — Wikipedia, The Free Encyclopedia, [Online; accessed 5-November-2023], 2023. διεύθν.: https://en.wikipedia.org/w/index.php?title=Natural_logarithm&oldid=1182868218.
- [64] Wikipedia contributors, *Exponential function* — Wikipedia, The Free Encyclopedia, [Online; accessed 5-November-2023], 2023. διεύθν.: https://en.wikipedia.org/w/index.php?title=Exponential_function&oldid=1181465336.

- [65] Wikipedia contributors, *Matrix exponential* — *Wikipedia, The Free Encyclopedia*, [Online; accessed 5-November-2023], 2023. διεύθν.: https://en.wikipedia.org/w/index.php?title=Matrix_exponential&oldid=1165819418.
- [66] Βικιπαίδεια, *Συσχέτιση και εξάρτηση (Στατιστική)* — *Βικιπαίδεια, Η Ελεύθερη Εγκυκλοπαίδεια*, [Online; accessed 18-Νοεμβρίου-2023], 2023. διεύθν.: [https://el.wikipedia.org/w/index.php?title=%CE%A3%CF%85%CF%83%CF%87%CE%AD%CF%84%CE%B9%CF%83%CE%B7_%CE%BA%CE%B1%CE%B9_%CE%B5%CE%BE%CE%AC%CF%81%CF%84%CE%B7%CF%83%CE%B7_\(%CE%A3%CF%84%CE%B1%CF%84%CE%B9%CF%83%CF%84%CE%B9%CE%BA%CE%AE\)&oldid=10245379](https://el.wikipedia.org/w/index.php?title=%CE%A3%CF%85%CF%83%CF%87%CE%AD%CF%84%CE%B9%CF%83%CE%B7_%CE%BA%CE%B1%CE%B9_%CE%B5%CE%BE%CE%AC%CF%81%CF%84%CE%B7%CF%83%CE%B7_(%CE%A3%CF%84%CE%B1%CF%84%CE%B9%CF%83%CF%84%CE%B9%CE%BA%CE%AE)&oldid=10245379).
- [67] *Συνδιακύμανση - Τι είναι, ορισμός και έννοια*. διεύθν.: <https://el.economy-pedia.com/11036890-covariance>.
- [68] X. Pennec, P. Fillard και N. Ayache, “A Riemannian Framework for Tensor Computing”, *International Journal of Computer Vision*, τόμ. 66, σσ. 41–66, 2005. διεύθν.: <https://api.semanticscholar.org/CorpusID:207251858>.
- [69] V. Arsigny, P. Fillard, X. Pennec και N. Ayache, “Log-Euclidean metrics for fast and simple calculus on diffusion tensors”, *Magnetic Resonance in Medicine*, τόμ. 56, αρθμ. 2, σσ. 411–421, 2006. DOI: <https://doi.org/10.1002/mrm.20965>. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/mrm.20965>. διεύθν.: <https://onlinelibrary.wiley.com/doi/abs/10.1002/mrm.20965>.
- [70] A. Cichocki και S. Amari, “Families of Alpha- Beta- and Gamma- Divergences: Flexible and Robust Measures of Similarities”, *Entropy*, τόμ. 12, σσ. 1532–1568, 2010. διεύθν.: <https://api.semanticscholar.org/CorpusID:15472649>.
- [71] A. Cichocki, S. Cruces και S. Amari, “Log-Determinant Divergences Revisited: Alpha-Beta and Gamma Log-Det Divergences”, *Entropy*, τόμ. 17, σσ. 2988–3034, 2014. διεύθν.: <https://api.semanticscholar.org/CorpusID:12951826>.
- [72] S. Sra, “A new metric on the manifold of kernel matrices with application to matrix geometric means”, στο *Neural Information Processing Systems*, 2012. διεύθν.: <https://api.semanticscholar.org/CorpusID:1234793>.
- [73] M. Moakher και P. Batchelor, “Symmetric Positive-Definite Matrices: From Geometry to Applications and Visualization”, στο *Iav*. 2006, σσ. 285–298, ISBN: 978-3-540-25032-6. DOI: [10.1007/3-540-31272-2_17](https://doi.org/10.1007/3-540-31272-2_17).
- [74] B. Kulis, M. A. Sustik και I. S. Dhillon, “Learning low-rank kernel matrices”, *Proceedings of the 23rd international conference on Machine learning*, 2006. διεύθν.: <https://api.semanticscholar.org/CorpusID:3049471>.
- [75] X. Liu, S. Liu και Z. Ma, “A Framework for Short Video Recognition Based on Motion Estimation and Feature Curves on SPD Manifolds”, *Applied Sciences*, τόμ. 12, αρθμ. 9, 2022, ISSN: 2076-3417. DOI: [10.3390/app12094669](https://doi.org/10.3390/app12094669). διεύθν.: <https://www.mdpi.com/2076-3417/12/9/4669>.
- [76] T. Zhang, W. Zheng, Z. Cui και C. Li, “Deep manifold-to-manifold transforming network for action recognition”, *CoRR*, τόμ. abs/1705.10732, 2017. arXiv: [1705.10732](https://arxiv.org/abs/1705.10732). διεύθν.: <http://arxiv.org/abs/1705.10732>.
- [77] S. Herath, M. T. Harandi και F. Porikli, “Learning an Invariant Hilbert Space for Domain Adaptation”, *CoRR*, τόμ. abs/1611.08350, 2016. arXiv: [1611.08350](https://arxiv.org/abs/1611.08350). διεύθν.: <http://arxiv.org/abs/1611.08350>.

- [78] J. Brownlee, *A gentle introduction to transfer learning for deep learning*, en-US, Σεπτ. 2019. διεύθν.: <https://machinelearningmastery.com/transfer-learning-for-deep-learning/>.
- [79] H. Wang, Q. Wang, M. Gao, P. Li και W. Zuo, “Multi-scale Location-aware Kernel Representation for Object Detection”, *CoRR*, τόμ. abs/1804.00428, 2018. arXiv: [1804.00428](https://arxiv.org/abs/1804.00428). διεύθν.: <http://arxiv.org/abs/1804.00428>.
- [80] J. Zhang, L. Zhou, L. Wang και W. Li, “Functional Brain Network Classification With Compact Representation of SICE Matrices”, *IEEE Trans. Biomed. Eng.*, τόμ. 62, αρθμ. 6, σσ. 1623–1634, 2015. DOI: [10.1109/TBME.2015.2399495](https://doi.org/10.1109/TBME.2015.2399495). διεύθν.: <https://doi.org/10.1109/TBME.2015.2399495>.
- [81] Z. Dong, S. Jia, C. Zhang, M. Pei και Y. Wu, “Deep Manifold Learning of Symmetric Positive Definite Matrices with Application to Face Recognition”, στο *AAAI Conference on Artificial Intelligence*, 2017. διεύθν.: <https://api.semanticscholar.org/CorpusID:13411844>.
- [82] Z. Huang και L. V. Gool, “A Riemannian Network for SPD Matrix Learning”, *CoRR*, τόμ. abs/1608.04233, 2016. arXiv: [1608.04233](https://arxiv.org/abs/1608.04233). διεύθν.: <http://arxiv.org/abs/1608.04233>.
- [83] P. Li, Q. Wang, W. Zuo και L. Zhang, “Log-Euclidean Kernels for Sparse Representation and Dictionary Learning”, στο *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, Δεκ. 2013.
- [84] R. Wang, H. Guo, L. S. Davis και Q. Dai, “Covariance discriminative learning: A natural and efficient approach to image set classification”, στο *2012 IEEE Conference on Computer Vision and Pattern Recognition*, 2012, σσ. 2496–2503. DOI: [10.1109/CVPR.2012.6247965](https://doi.org/10.1109/CVPR.2012.6247965).
- [85] R. Vemulapalli, J. K. Pillai και R. Chellappa, “Kernel Learning for Extrinsic Classification of Manifold Features”, στο *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Ιούν. 2013.
- [86] J. Zhang, L. Wang, L. Zhou και W. Li, “Learning Discriminative Stein Kernel for SPD Matrices and Its Applications”, *IEEE transactions on neural networks and learning systems*, τόμ. 27, Μάι. 2015. DOI: [10.1109/TNNLS.2015.2435154](https://doi.org/10.1109/TNNLS.2015.2435154).
- [87] M. T. Harandi, C. Sanderson, R. I. Hartley και B. C. Lovell, “Sparse Coding and Dictionary Learning for Symmetric Positive Definite Matrices: A Kernel Approach”, *CoRR*, τόμ. abs/1304.4344, 2013. arXiv: [1304.4344](https://arxiv.org/abs/1304.4344). διεύθν.: <http://arxiv.org/abs/1304.4344>.
- [88] M. T. Harandi και M. Salzmann, “Riemannian coding and dictionary learning: Kernels to the rescue”, *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, σσ. 3926–3935, 2015. διεύθν.: <https://api.semanticscholar.org/CorpusID:13915435>.
- [89] A. Cherian, P. Stanitsas, M. Harandi, V. Morellas και N. Papanikolopoulos, “Learning Discriminative Alpha-Beta-divergence for Positive Definite Matrices (Extended Version)”, *CoRR*, τόμ. abs/1708.01741, 2017. arXiv: [1708.01741](https://arxiv.org/abs/1708.01741). διεύθν.: <http://arxiv.org/abs/1708.01741>.
- [90] T. Sanger και P. N. Baljekar, “The perceptron: a probabilistic model for information storage and organization in the brain.”, *Psychological review*, τόμ. 65 6, σσ. 386–408, 1958. διεύθν.: <https://api.semanticscholar.org/CorpusID:12781225>.

- [91] D. Beeta, S. Cruces, J. Olias και A. Cichocki, “Optimization of Alpha-Beta Log-Det Divergences and their Application in the Spatial Filtering of Two Class Motor Imagery Movements”, *Entropy*, τόμ. 19, σ. 89, Φεβ. 2017. DOI: [10.3390/e19030089](https://doi.org/10.3390/e19030089).
- [92] R. Vemulapalli και D. W. Jacobs, “Riemannian Metric Learning for Symmetric Positive Definite Matrices”, *CoRR*, τόμ. abs/1501.02393, 2015. arXiv: [1501.02393](https://arxiv.org/abs/1501.02393). διεύθν.: <http://arxiv.org/abs/1501.02393>.
- [93] Z. Huang, R. Wang, S. Shan, X. Li και X. Chen, “Log-Euclidean Metric Learning on Symmetric Positive Definite Manifold with Application to Image Set Classification”, στο *International Conference on Machine Learning*, 2015. διεύθν.: <https://api.semanticscholar.org/CorpusID:1040234>.
- [94] L. Zhou, L. Wang, J. Zhang, Y. Shi και Y. Gao, “Revisiting metric learning for SPD matrix based visual representation”, στο *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, σσ. 3241–3249.
- [95] H. Q. Minh, M. San-Biagio και V. Murino, “Log-Hilbert-Schmidt metric between positive definite operators on Hilbert spaces”, στο *Neural Information Processing Systems*, 2014. διεύθν.: <https://api.semanticscholar.org/CorpusID:2448883>.
- [96] M. T. Harandi, M. Salzmann και R. I. Hartley, “Dimensionality Reduction on SPD Manifolds: The Emergence of Geometry-Aware Methods”, *CoRR*, τόμ. abs/1605.06182, 2016. arXiv: [1605.06182](https://arxiv.org/abs/1605.06182). διεύθν.: <http://arxiv.org/abs/1605.06182>.
- [97] J. Lu, J. Hu και Y.-P. Tan, “Discriminative deep metric learning for face and kinship verification”, *IEEE Transactions on Image Processing*, τόμ. 26, αρθμ. 9, σσ. 4269–4282, 2017.
- [98] H. Yan, “Collaborative discriminative multi-metric learning for facial expression recognition in video”, *Pattern Recognition*, τόμ. 75, σσ. 33–40, 2018.
- [99] Y. Zhang, H. Zhang, N. M. Nasrabadi και T. S. Huang, “Multi-metric learning for multi-sensor fusion based classification”, *Information Fusion*, τόμ. 14, αρθμ. 4, σσ. 431–440, 2013.
- [100] P.-A. Absil, R. Mahony και R. Sepulchre, *Optimization Algorithms on Matrix Manifolds*. USA: Princeton University Press, 2007, ISBN: 0691132984.
- [101] M. Harandi, M. Salzmann και R. Hartley, “Joint Dimensionality Reduction and Metric Learning: A Geometric Take”, στο *Proceedings of the 34th International Conference on Machine Learning*, D. Precup και Y. W. Teh, επιμελητές, σειρά Proceedings of Machine Learning Research, τόμ. 70, PMLR, Αύγ. 2017, σσ. 1404–1413. διεύθν.: <https://proceedings.mlr.press/v70/harandi17a.html>.
- [102] S. Ying, Z. Wen, J. Shi, Y. Peng, J. Peng και H. Qiao, “Manifold Preserving: An Intrinsic Approach for Semisupervised Distance Metric Learning”, *IEEE Transactions on Neural Networks and Learning Systems*, τόμ. 29, αρθμ. 7, σσ. 2731–2742, 2018. DOI: [10.1109/TNNLS.2017.2691005](https://doi.org/10.1109/TNNLS.2017.2691005).
- [103] M. Faraki, M. Harandi και F. Porikli, “Large-Scale Metric Learning: A Voyage From Shallow to Deep”, *IEEE Transactions on Neural Networks and Learning Systems*, τόμ. PP, σσ. 1–8, Νοέ. 2017. DOI: [10.1109/TNNLS.2017.2761773](https://doi.org/10.1109/TNNLS.2017.2761773).

- [104] J. V. Davis, B. Kulis, P. Jain, S. Sra και I. S. Dhillon, “Information-Theoretic Metric Learning”, στο *Proceedings of the 24th International Conference on Machine Learning*, σειρά ICML '07, Association for Computing Machinery, 2007, σσ. 209–216, ISBN: 9781595937933. DOI: [10 . 1145 / 1273496 . 1273523](https://doi.org/10.1145/1273496.1273523). διεύθν.: <https://doi.org/10.1145/1273496.1273523>.
- [105] M. Harandi και B. Fernando, “Generalized BackPropagation, Étude De Cas: Orthogonality”, *CoRR*, τόμ. abs/1611.05927, 2016. arXiv: [1611 . 05927](https://arxiv.org/abs/1611.05927). διεύθν.: <http://arxiv.org/abs/1611.05927>.
- [106] S. K. Roy, Z. Mhammedi και M. Harandi, “Geometry Aware Constrained Optimization Techniques for Deep Learning”, στο *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018, σσ. 4460–4469. DOI: [10.1109/CVPR.2018.00469](https://doi.org/10.1109/CVPR.2018.00469).
- [107] S. Bonnabel, “Stochastic Gradient Descent on Riemannian Manifolds”, *IEEE Transactions on Automatic Control*, τόμ. 58, αρθμ. 9, σσ. 2217–2229, Σεπτ. 2013, ISSN: 1558-2523. DOI: [10 . 1109 / tac . 2013 . 2254619](https://doi.org/10.1109/tac.2013.2254619). διεύθν.: <http://dx.doi.org/10.1109/TAC.2013.2254619>.
- [108] S. Mitra και M. Gofman, *Biometrics in a Data Driven World: Trends, Technologies, and Challenges*. Chapman & Hall/CRC, 2016, ISBN: 1498737641.
- [109] A. K. Jain και D. Maltoni, *Handbook of Fingerprint Recognition*. Berlin, Heidelberg: Springer-Verlag, 2003, ISBN: 0387954317.
- [110] M. Diaz, M. A. Ferrer, D. Impedovo, M. I. Malik, G. Pirlo και R. Plamondon, “A Perspective Analysis of Handwritten Signature Technology”, *ACM Comput. Surv.*, τόμ. 51, αρθμ. 6, Ιαν. 2019, ISSN: 0360-0300. DOI: [10.1145/3274658](https://doi.org/10.1145/3274658). διεύθν.: <https://doi.org/10.1145/3274658>.
- [111] R. Bajaj και S. Chaudhury, “Signature verification using multiple neural classifiers”, *Pattern Recognition*, τόμ. 30, αρθμ. 1, σσ. 1–7, 1997, ISSN: 0031-3203. DOI: [https://doi.org/10.1016/S0031-3203\(96\)00059-3](https://doi.org/10.1016/S0031-3203(96)00059-3). διεύθν.: <https://www.sciencedirect.com/science/article/pii/S0031320396000593>.
- [112] H. Baltzakis και N. Papamarkos, “A new signature verification technique based on a two-stage neural network classifier”, *Engineering Applications of Artificial Intelligence*, τόμ. 14, αρθμ. 1, σσ. 95–103, 2001, ISSN: 0952-1976. DOI: [https://doi.org/10.1016/S0952-1976\(00\)00064-6](https://doi.org/10.1016/S0952-1976(00)00064-6). διεύθν.: <https://www.sciencedirect.com/science/article/pii/S0952197600000646>.
- [113] K. Huang και H. Yan, “Off-line signature verification based on geometric feature extraction and neural network classification”, *Pattern Recognition*, τόμ. 30, αρθμ. 1, σσ. 9–17, 1997, ISSN: 0031-3203. DOI: [https://doi.org/10.1016/S0031-3203\(96\)00063-5](https://doi.org/10.1016/S0031-3203(96)00063-5). διεύθν.: <https://www.sciencedirect.com/science/article/pii/S0031320396000635>.
- [114] Ö. Yedekçioğlu, M. Akhan και Y. Lim, *Off-line Signature Verification with Thickened Templates* (Κοç University Working Paper). Κοç University, 1995. διεύθν.: <https://books.google.gr/books?id=LCAatgEACAAJ>.

- [115] G. S. Eskander, R. Sabourin και E. Granger, “Hybrid writer-independent–writer-dependent offline signature verification system”, *IET Biometrics*, τόμ. 2, αρθμ. 4, σσ. 169–181, 2013. DOI: <https://doi.org/10.1049/iet-bmt.2013.0024>. eprint: <https://ietresearch.onlinelibrary.wiley.com/doi/pdf/10.1049/iet-bmt.2013.0024>. διεύθν.: <https://ietresearch.onlinelibrary.wiley.com/doi/abs/10.1049/iet-bmt.2013.0024>.
- [116] S.-H. Cha και S. N. Srihari, “Writer Identification: Statistical Analysis and Dichotomizer”, στο *Advances in Pattern Recognition*, F. J. Ferri, J. M. Iñesta, A. Amin και P. Pudil, επιμελητές, Berlin, Heidelberg: Springer Berlin Heidelberg, 2000, σσ. 123–132, ISBN: 978-3-540-44522-7.
- [117] V. L. Souza, A. L. Oliveira, R. M. Cruz και R. Sabourin, “A white-box analysis on the writer-independent dichotomy transformation applied to offline handwritten signature verification”, *Expert Systems with Applications*, τόμ. 154, σ. 113 397, 2020, ISSN: 0957-4174. DOI: <https://doi.org/10.1016/j.eswa.2020.113397>. διεύθν.: <https://www.sciencedirect.com/science/article/pii/S0957417420302219>.
- [118] M. K. Kalera, S. N. Srihari και A. Xu, “Offline Signature Verification And Identification Using Distance Statistics”, *Int. J. Pattern Recognit. Artif. Intell.*, τόμ. 18, σσ. 1339–1360, 2004. διεύθν.: <https://api.semanticscholar.org/CorpusID:177167>.
- [119] A. Alaei, S. Pal, U. Pal και M. Blumenstein, “An Efficient Signature Verification Method Based on an Interval Symbolic Representation and a Fuzzy Similarity Measure”, *IEEE Transactions on Information Forensics and Security*, τόμ. 12, σσ. 2360–2372, 2017. διεύθν.: <https://api.semanticscholar.org/CorpusID:28367281>.
- [120] N. Otsu, “A Threshold Selection Method from Gray-Level Histograms”, *IEEE Transactions on Systems, Man, and Cybernetics*, τόμ. 9, αρθμ. 1, σσ. 62–66, 1979. DOI: [10.1109/TSMC.1979.4310076](https://doi.org/10.1109/TSMC.1979.4310076).
- [121] E. Zois, D. Tsourounis, I. Theodorakopoulos, A. Kesidis και G. Economou, “A Comprehensive Study of Sparse Representation Techniques for Offline Signature Verification”, *IEEE Transactions on Biometrics, Behavior, and Identity Science*, τόμ. PP, σσ. 1–1, Φεβ. 2019. DOI: [10.1109/TBIOM.2019.2897802](https://doi.org/10.1109/TBIOM.2019.2897802).
- [122] D. G. Lowe, “Object recognition from local scale-invariant features”, στο *Proceedings of the Seventh IEEE International Conference on Computer Vision*, τόμ. 2, 1999, 1150–1157 vol.2. DOI: [10.1109/ICCV.1999.790410](https://doi.org/10.1109/ICCV.1999.790410).
- [123] D. G. Lowe, “Distinctive Image Features from Scale-Invariant Keypoints”, *International Journal of Computer Vision*, τόμ. 60, σσ. 91–110, 2004. διεύθν.: <https://api.semanticscholar.org/CorpusID:174065>.
- [124] H. Bay, T. Tuytelaars και L. Van Gool, “SURF: Speeded Up Robust Features”, στο *Computer Vision – ECCV 2006*, A. Leonardis, H. Bischof και A. Pinz, επιμελητές, Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, σσ. 404–417, ISBN: 978-3-540-33833-8.
- [125] P. F. Alcantarilla, A. Bartoli και A. J. Davison, “KAZE Features”, στο *Computer Vision – ECCV 2012*, A. Fitzgibbon, S. Lazebnik, P. Perona, Y. Sato και C. Schmid, επιμελητές, Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, σσ. 214–227, ISBN: 978-3-642-33783-3.
- [126] O. Tuzel, F. M. Porikli και P. Meer, “Region Covariance: A Fast Descriptor for Detection and Classification”, στο *European Conference on Computer Vision*, 2006. διεύθν.: <https://api.semanticscholar.org/CorpusID:6638556>.

- [127] E. N. Zois, D. Tsourounis και D. Kalivas, “Similarity Distance Learning on SPD Manifold for Writer Independent Offline Signature Verification”, *IEEE Transactions on Information Forensics and Security*, τόμ. 19, σσ. 1342–1356, 2024. DOI: [10.1109/TIFS.2023.3333681](https://doi.org/10.1109/TIFS.2023.3333681).
- [128] T. Fawcett, “An introduction to ROC analysis”, *Pattern Recognition Letters*, τόμ. 27, αρθμ. 8, σσ. 861–874, 2006, ROC Analysis in Pattern Recognition, ISSN: 0167-8655. DOI: <https://doi.org/10.1016/j.patrec.2005.10.010>. διεύθν.: <https://www.sciencedirect.com/science/article/pii/S016786550500303X>.