



ΠΑΝΕΠΙΣΤΗΜΙΟ ΔΥΤΙΚΗΣ ΑΤΤΙΚΗΣ
ΣΧΟΛΗ ΜΗΧΑΝΙΚΩΝ
ΤΜΗΜΑ ΗΛΕΚΤΡΟΛΟΓΩΝ & ΗΛΕΚΤΡΟΝΙΚΩΝ ΜΗΧΑΝΙΚΩΝ

Διπλωματική Εργασία

Από τον Ευκλείδη στον Riemann: Ένα άλμα στην υπολογιστική όραση σε γεωμετρικούς χώρους



Φοιτητής: Βασιλάκης Νικόλαος
ΑΜ: eee18387056

Επιβλέποντες Καθηγητές

Ζώης Ηλίας, Ph.D.
Αναπληρωτής Καθηγητής

Βενέτης Ιωάννης, Ph.D.
Επίκουρος Καθηγητής, Τμήμα
Πληροφορικής, Πανεπιστήμιο Πειραιά

ΑΘΗΝΑ-ΑΙΓΑΛΕΩ, ΣΕΠΤΕΜΒΡΙΟΣ 2024



UNIVERSITY OF WEST ATTICA
FACULTY OF ENGINEERING
DEPARTMENT OF ELECTRICAL & ELECTRONICS ENGINEERING

Diploma Thesis

A leap of faith from Euclidean to Riemannian Manifolds for computer vision applications



Student: Vasilakis Nikolaos
Registration Number: eee18387056

Supervisors

Elias Zois, Ph.D.
Associate Professor

Venetis Ioannis, Ph.D.
Assistant Professor, Department of
Informatics University of Piraeus

ATHENS-EGALEO, SEPTEMBER 2024

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

Ζώης Ηλίας, Αναπληρωτής καθηγητής	Βενέτης Ιωάννης, Επίκουρος Καθηγητής- Πανεπιστήμιο Πειραιώς, Τμήμα Πληροφορικής	Χωριανόπουλος Χρήστος, Επίκουρος Καθηγητής
(Υπογραφή)	(Υπογραφή)	(Υπογραφή)

Copyright © Με επιφύλαξη παντός δικαιώματος. All rights reserved.

ΠΑΝΕΠΙΣΤΗΜΙΟ ΔΥΤΙΚΗΣ ΑΤΤΙΚΗΣ και Βασιλάκης Νικόλαος, Σεπτέμβριος 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/την 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/η Βασιλάκης Νικόλαος του Μιχαήλ με αριθμό μητρώου 18387056 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.

Ο Δηλών
(Ονοματεπώνυμο φοιτητή/ήτριας)

(Υπογραφή φοιτητή/ήτριας)

Στην μητέρα μου Όλγα

Στα έργα του, ο Ναπόλεον Χιλ έγραψε πως η ποιότητα του έργου μας στη ζωή καθορίζεται από τους ανθρώπους που επιλέγουμε να έχουμε κοντά μας. Στο κύκνειο άσμα των φοιτητικών μου χρόνων, θα ήταν εσφαλμένο να υποστηρίξω ότι κατάφερα ο,τι κατάφερα μόνος μου. Αισθάνομαι υπέρμετρα τυχερός και θα ήθελα να ευχαριστήσω:

- Τον επιβλέποντα καθηγητή μου κ. Ηλία Ζώη. Παρατηρώντας το πάθος και τον ζήλο του για αυτό που κάνει να τον εξυψώνει σε επίπεδο μαεστρίας, ο κ. Ζώης αποτέλεσε πραγματικό παράδειγμα και φώτισε τον δρόμο μου προς το μονοπάτι που θέλω να ακολουθήσω. Το συγγραφικό του έργο σαν επιστήμονας, η προσέγγισή του πέραν της ακαδημαϊκής του ιδιότητας αλλά πρωτίστως το ποιόν του σαν άνθρωπος με βοήθησαν πολύ περισσότερο από ότι μπορεί κανείς να φανταστεί. Τον ευχαριστώ για την υπομονή και την εμπιστοσύνη που μου έδειξε. Θεωρώ ότι είμαι πραγματικά τυχερός που τον γνώρισα και με δίδαξε.
- Τους καθηγητές μου, Δημήτρα, Κώστα και Γιάννη. Ο καθένας με την πάροδο των χρόνων τοποθετούσε ένα μικρό λίθο στο σχηματισμό της πολύπτυχης προσωπικότητας μου και των μαθησιακών μου ιδιοτήτων. Ήταν οι δικοί μου φάροι γνώσης και αρχών. Τους ευχαριστώ.
- Τον φίλο μου Αλέξιο. Η σχέση μας ξεκίνησε μέσω του μαθήματος του CVPR και μετουσιώθηκε σε κάτι που ξεπέρασε τα τείχη της σχολής. Για όλες τις συμβουλές του, προσωπικές και επαγγελματικές, το χρόνο του και την ικανότητα του να με ακούει καθώς του υποδείκνυα ότι το MATLAB ξεπερνά την ερπετόμορφη γλώσσα προγραμματισμού του, τον ευχαριστώ.
- Τον φίλο μου Αγάπιο. Η υπομονή και η αστείρευτη ικανότητα του να ακούει οποιοδήποτε παράπονο, γκρίνια ή πρόβλημά μου παραπάνω από όσο θα ήθελε είναι αξιοζήλευτο ενώ οι προτάσεις του σε πολλά από τα προηγούμενα με έχουν αποτρέψει από δύσκολες καταστάσεις. Για όλα αυτά και άλλα πολλά, τον ευχαριστώ.
- Τους φίλους μου, Ανδρέα, Κώστα, Νίκο, Δημήτρη και Τάσο για τις όμορφες στιγμές που περάσαμε μαζί ως παρέα στην σχολή αλλά και εκτός από αυτήν. Για την υπομονή, την ανεκτικότητα και την αποδοχή τους, τούς ευχαριστώ.
- Το εργαστήριο TelSiP στην αίθουσα ZB203, το δεύτερό μου σπίτι. Οι ώρες που πέρασα στην αίθουσα του 203 δύσκολα μπορούν να μετρηθούν. Οι άνθρωποι που γνώρισα και οι αναμνήσεις που σμιλεύτηκαν εκεί θα με συνοδεύουν για πάντα. Τούς ευχαριστώ όλους.

*“I overcame myself, the sufferer;
I carried my own ashes to the mountains;
I invented a brighter flame for myself.”
- Friedrich Nietzsche*

*“No man ever steps in the same river twice,
for it's not the same river and he's not the same man”
-Heraclitus*

*«Our virtues and our failings are inseparable, like force and matter.
When they separate, man is no more.»
-Nikola Tesla*

Περίληψη

Η παγκόσμια επιστημονική κοινότητα, μέχρι πρότινος, παραμένοντας πιστή στις κοινότητες προσεγγίσεις των μεθόδων που χρησιμοποιούνται στην Μηχανική Μάθηση στηρίζεται στην υπόθεση ότι η δόμηση και κατά συνέπεια η επεξεργασία των δεδομένων θα πρέπει να υπόκειται σε πράξεις οι οποίες ορίζονται σε διανυσματικούς χώρους με πρωταρχική βάση την Ευκλείδεια γεωμετρία. Πορίσματα από πληθώρα ερευνών έχουν δείξει ότι ο επιστημονικός τομέας της Υπολογιστική Όραση επωφελείται από αναπαραστάσεις δεδομένων που είναι *συμπαγείς*, *διακριτές*, και *εύρωστες*. Εδώ λοιπόν εγείρεται το ακόλουθο ερώτημα: Μπορούμε να υποστηρίξουμε με απόλυτη βεβαιότητα ότι τα δεδομένα, όπως αυτά συλλέγονται και επεξεργάζονται, ακολουθούν πιστά την Ευκλείδεια Γεωμετρία; Αν όχι, υπάρχει κάποιου είδους υποκείμενης γεωμετρικής δομής η οποία δρα συμπληρωματικά στην ήδη απλοϊκή φύση του Ευκλείδειου Γεωμετρικού χώρου ενισχύοντας τα μοντέλα Μηχανικής Μάθησης με επιπρόσθετη περιγραφικότητα οδηγώντας έτσι σε υψηλότερα επίπεδα ακρίβειας προβλέψεων; Μια πιθανή απάντηση είναι η χρήση Συμμετρικών Θετικά Ορισμένων πινάκων οι οποίοι αντιμετωπίζονται πιο φυσικά μέσω των εννοιών της διαφορικής γεωμετρίας που εμφανίζεται σε μη-Ευκλείδειους χώρους, όπως αυτός του Riemann.

Στόχος της παρούσας διπλωματικής είναι να γίνει μια μελέτη σε αλγόριθμους μη επιβλεπόμενης μάθησης τόσο σε Ευκλείδειους όσο και σε Riemannian χώρους με την χρήση προγραμματιστικών εργαλείων όπως το MATLAB για εφαρμογές Υπολογιστικής Όρασης. Συγκεκριμένα, στον Ευκλείδειο χώρο μελετώνται οι αλγόριθμοι: α) *Bag of Features*, β) *Spatial Pyramid Matching*, γ) *Fisher Kernels* και δ) *Vectors of Locally Aggregated Descriptors* ενώ στον Riemannian χώρο μελετώνται τεχνικές αλγόριθμων που εκμεταλλεύονται την εγγενώς μη γραμμική φύση των πολλαπλοτήτων στις οποίες απαντώνται οι πίνακες συνδιακύμανσης. Οι τέσσερις αριθμημένοι αλγόριθμοι του Ευκλείδειου Χώρου δοκιμάζονται σε μια βάση εικόνων που περιέχουν σκηνές για την μελέτη της προοδευτικής αύξησης της ακρίβειας της τεχνικής της διανυσματικής αναπαράστασης των εικόνων ενώ οι αλγόριθμοι στον Riemannian Χώρο επεξεργάζονται εικόνες χειρόγραφων υπογραφών από μια δημοφιλή βάση με σκοπό την βιομετρική ταυτοποίησή τους. Όπου κατέστη δυνατό, εφαρμόζεται επιτάχυνση του λογισμικού για την μείωση του χρόνου λήψης αποτελεσμάτων στην υλοποίηση των αλγόριθμων χρησιμοποιώντας παράλληλη επεξεργασία και χρήση καρτών γραφικών.

Λέξεις – κλειδιά

Υπολογιστική Όραση, Ψηφιακή Επεξεργασία Εικόνας, Ευκλείδειοι χώροι, Riemannian χώροι, Συμμετρικό Θετικά Ορισμένο Πολύπτυχο, Παράλληλη Επεξεργασία, Επεξεργασία με Κάρτες Γραφικών, MATLAB, Διανυσματική Αναπαράσταση Εικόνων.

Abstract

The global scientific community, until recently, remaining faithful to the common approaches of the methods used in Machine Learning, has assumed that the construction and consequently the processing of data should be subject to operations defined in vector spaces with Euclidean geometry as the primary basis. Findings from a plethora of research have shown that the scientific field of Computational Vision benefits from data representations that are *compact*, *discrete*, and *robust*, the following question therefore arises here: Can we assert with absolute certainty that the data, as collected and processed, faithfully follow Euclidean geometry? If not, is there some kind of underlying geometric structure that acts in a complementary way to the already simplistic nature of the Euclidean Geometry space by enhancing Machine Learning models with additional descriptiveness leading to higher levels of prediction accuracy? One possible answer is the use of Symmetric Positively Defined matrices which are treated more naturally through the concepts of differential geometry occurring in non-Euclidean spaces such as Riemannian spaces.

The objective of this diploma thesis is to provide a thorough review on unsupervised learning algorithms in both Euclidean as well as Riemannian spaces with the use of programming tools such as the MATLAB for computer vision applications. In the Euclidean space the following algorithms are under examination: a) Bag of Visual Words (BoVW), b) Spatial Pyramid Matching (SPM), c) Fisher Kernels (FK) or Fisher Vectors (FV), and d) Vectors of Locally Aggregated Descriptors (VLAD) whereas in Riemannian space techniques for algorithms that exploit the intrinsic non-linear nature of the manifolds that comprise the covariance matrices are studied. The four numbered algorithms in Euclidean Space are tested on an image database containing scenes to study the progressive increase in the accuracy of the vector image representation technique, while the algorithms in Riemannian Space process images of handwritten manuscripts from a popular signature database for biometric identification. Wherever feasible, software acceleration is applied to reduce the time to obtain results in the implementation of the algorithms using parallel processing and GPU computing.

Keywords

Computer Vision, Digital Image Processing, Euclidean spaces, Riemannian spaces, SPD Manifold, Parallel Processing, GPU Computing, MATLAB, Biometrics, Automated Signature Verification, Image Vector Representation.

Περιεχόμενα

Κατάλογος Πινάκων.....	12
Κατάλογος Εικόνων	13
Κατάλογος Ορισμών	15
Κατάλογος Αλγορίθμων.....	16
ΕΙΣΑΓΩΓΗ.....	17
Αντικείμενο της διπλωματικής εργασίας.....	18
Σκοπός και στόχοι	19
Μεθοδολογία.....	20
Καινοτομία	21
Δομή	22
1 ΚΕΦΑΛΑΙΟ 1^ο : Ψηφιακή Επεξεργασία Εικόνας.....	23
1.1 Βασικές αρχές της Ψηφιακής Επεξεργασίας Εικόνας	23
1.1.1 Θεμελιώδεις λίθοι στην Ψηφιακή Επεξεργασία Εικόνας.....	24
1.1.2 Αναπαράσταση της Ψηφιακής Εικόνας.....	31
1.2 Βασικές σχέσεις εικονοστοιχείων	32
1.2.1 Η έννοια του «Γείτονα» στα εικονοστοιχεία	32
1.2.2 Η έννοια της Γειτνίασης στα εικονοστοιχεία.....	35
1.2.3 Η έννοια της Συνδεσιμότητας στα εικονοστοιχεία	36
1.3 Βασικές πράξεις στην Επεξεργασία Εικόνας.....	37
1.3.1 Πράξεις στοιχείο-προς-στοιχείο	37
1.3.2 Γραμμικές και μη-Γραμμικές πράξεις	38
1.3.3 Χωρικές Πράξεις: Γεωμετρικοί Μετασχηματισμοί	39
1.4 Επεξεργασία Ιστογράμματος.....	42
1.4.1 Τι είναι Ιστόγραμμα	42
1.5 Εξαγωγή Χαρακτηριστικών.....	45
2 ΚΕΦΑΛΑΙΟ 2^ο : Αλγόριθμοι Μάθησης.....	48
2.1 Βασικές κατηγορίες Αλγορίθμων Μάθησης	49
2.2 Μηχανές Υποστήριξης Διανυσμάτων (Support Vector Machines)	51
2.2.1 Διαχωριστικό Υπερεπίπεδο	53
2.2.2 Η έννοια του Περιθωρίου	55
2.2.3 Από το Πρωτεύον στο Δυαδικό πρόβλημα :Κυρτή δυαδικότητα μέσω πολλαπλασιαστών Lagrange	57
2.2.4 Ο ταξινομητής βέλτιστου Περιθωρίου	58
2.2.5 Πυρήνες (Kernels)	60
2.3 Αλγόριθμοι Μη Επιβλεπόμενης Μάθησης.....	63
2.3.1 Αλγόριθμοι Συσταδοποίησης: Γκαουσιανά Μοντέλα Μίξης (GMM).....	65
2.3.2 Γκαουσιανές Κατανομές (Gaussian Distributions).....	65
2.3.3 Γκαουσιανά Μοντέλα Μίξης	67
2.3.4 Εκτίμηση Μέγιστης Πιθανοφάνειας (Maximum Likelihood Estimation)	71
2.3.5 Εκμάθηση παραμέτρων μέσω της Εκτίμησης Μέγιστης Πιθανοφάνειας.....	72
2.3.6 Βάρη Ευθυνών (Responsibilities)	73
2.3.7 Ενημέρωση των παραμέτρων.....	73
2.3.8 Ο αλγόριθμος Μεγιστοποίησης Προσδοκιών (Expectation Maximization).....	74
2.3.9 Ο αλγόριθμος Βηματικής Μεγιστοποίησης Προσδοκιών (stepwise-Expectation Maximization)....	78
3 ΚΕΦΑΛΑΙΟ 3^ο : Μαθηματικό Υπόβαθρο.....	81
3.1 Riemannian Πολλαπλότητες (Riemannian Manifolds)	82
3.1.1 Η έννοια της μετρικής και του μετρικού Χώρου	82
3.1.2 Ορισμός της Riemannian Πολλαπλότητας	83

3.1.3	Γεωδαισιακές.....	84
3.1.4	Εκθετική Απεικόνιση.....	87
3.1.5	Οι Πολλαπλότητες των Συμμετρικών Θετικά Ορισμένων Πινάκων	88
4	ΚΕΦΑΛΑΙΟ 4^ο : Αλγόριθμοι στον Ευκλείδειο και Riemannian Χώρο.....	94
4.1	Αλγόριθμοι στον Ευκλείδειο Χώρο	95
4.1.1	Bag of Features	96
4.1.2	Spatial Pyramid Matching	102
4.1.3	Fisher Kernels.....	110
4.1.4	Vectors of Locally Aggregated Descriptors (VLAD)	114
4.2	Αλγόριθμοι στην Riemannian Πολλαπλότητα	120
4.2.1	Βιομετρικά Συστήματα	121
4.2.2	Σφάλματα Μονάδων Αντιστοίχισης (<i>Matching Module Errors</i>).....	122
4.2.3	Είδη πλαστογραφίας στις υπογραφές.....	125
4.2.4	Προσεγγίσεις Ταξινόμησης Υπογραφών	126
4.2.5	Αξιοποιώντας τον Εφαπτομενικό Χώρο	129
4.2.6	Αξιοποιώντας τη Riemannian Πολλαπλότητα και τον Εφαπτομενικό Χώρο	135
5	ΚΕΦΑΛΑΙΟ 5^ο : Συμπεράσματα.....	140
6	Προτάσεις προς επέκταση	141
	Παράρτημα Α: Ο χώρος του Riemann - Πολλαπλότητες.....	143
	Χώρος και αριθμοί	143
	Ομοιομορφισμός και Διαφορομορφισμός (<i>Homeomorphism & Diffeomorphism</i>)	144
	Λείες Πολλαπλότητες (<i>Smooth Manifolds</i>).....	146
	Χάρτες Συντεταγμένων (<i>Coordination Charts</i>).....	147
	Μεταβατικές Απεικονίσεις (<i>Transition Maps</i>)	148
	Άτλαντες και Ομαλοί Άτλαντες	149
	Η έννοια του Μέγιστου Λείου Άτλαντα	149
	Λείες Συναρτήσεις και Λείες Απεικονίσεις (<i>Smooth Functions and Smooth Maps</i>).....	150
	Παράρτημα Β: Βασικοί όροι της Διαφορικής Γεωμετρίας	151
	Βασικοί όροι της Διαφορικής Γεωμετρίας	151
	Γεωμετρικά Εφαπτομενικά διανύσματα και Εφαπτομενικοί Χώροι	152
	Η Εφαπτομενική Δέσμη (<i>The Tangent Bundle</i>).....	154
	Βιβλιογραφία – Αναφορές - Διαδικτυακές Πηγές	156

Κατάλογος Πινάκων

Πίνακας 1:	Συγκεντρωτικός πίνακας απεικόνισης των affine μετασχηματισμών.....	41
Πίνακας 2:	Ενδεικτικές τιμές των παραμέτρων των τεσσάρων πληθυσμών.....	70
Πίνακας 3:	Αποτίπωση των τιμών της ακρίβειας για τις δέκα διακριτές επαναλήψεις.....	102
Πίνακας 4:	Αποτίπωση του συνολικού πίνακα για τις μετρήσεις που πραγματοποιήθηκαν σε 10 διαφορετικά πειράματα. Με έντονα μαύρη γραφή υποδηλώνεται η καλύτερη τιμή ακρίβειας με τις παραμέτρους προς βελτιστοποίηση και τον αντίστοιχο αριθμό κέντρων.....	107
Πίνακας 5:	Αποτίπωση των χρόνων για την λήψη των αποτελεσμάτων από τις τρεις συναρτήσεις.....	113
Πίνακας 6:	Πίνακας αποτίπωσης των μετρήσεων που συλλέχθηκαν από 10 διαφορετικά πειράματα για τα VLAD. Με έντονη γραφή σημειώνεται η υψηλότερη μέση ακρίβεια που εντοπίστηκε.....	116
Πίνακας 7:	Αποτίπωση του EER ανά επανάληψη κι ανά fold καθώς και του συνολικού μέσου EER για το πρώτο Riemannian σενάριο.....	135
Πίνακας 8:	Αποτίπωση του EER ανά επανάληψη κι ανά fold καθώς και του συνολικού μέσου EER για το πρώτο Riemannian σενάριο.....	140
Πίνακας 9:	Τελικός πίνακας αποτίπωσης της αύξησης της ακρίβειας ανά τους αλγορίθμους.....	141

Κατάλογος Εικόνων

Εικόνα 1: Από αριστερά προς τα δεξιά: A) Αρχική εικόνα προς επεξεργασία. B) Grayscale αναπαράσταση της αρχικής εικόνας. Γ) Υπόδειξη ενός αυθαίρετου ευθύγραμμου τμήματος κατά μήκος του οποίου εξάγονται οι τιμές έντασης φωτεινότητας. -----	26
Εικόνα 2: Από πάνω προς τα κάτω A) Απεικόνιση των τιμών έντασης που εντοπίστηκαν στο ευθύγραμμο τμήμα που ορίζεται στην Εικόνα 1A) B) Υπέρθυση 30 διακεκριμένων τιμών/δειγμάτων (με κόκκινο χρώμα) πάνω στις τιμές εντάσεων Γ) Απομόνωση και απεικόνιση των 30 διακεκριμένων τιμών/δειγμάτων. -----	26
Εικόνα 3: Από πάνω προς τα κάτω A) Απεικόνιση των δειγματοληπτούμενων τιμών από το ευθύγραμμο τμήμα της Εικόνας 1Γ). B) Αποτύπωση των κβαντισμένων δειγματοληπτούμενων τιμών. Γ) Αποτύπωση της διαδικασίας της ανάθεσης των δειγματοληπτούμενων τιμών σε διακεκριμένες τιμές. -----	28
Εικόνα 4: Από πάνω προς τα κάτω κι από αριστερά προς τα δεξιά A) Αποτύπωση της αρχικής εικόνας με το παράθυρο επεξεργασίας πληροφοριών των εικονοστοιχείων. B) Αρχική σμίκρυνση του παραθύρου επεξεργασίας πληροφοριών. Γ) Περαιτέρω σμίκρυνση του παραθύρου επεξεργασίας πληροφοριών. Δ) Τελική σμίκρυνση του παραθύρου επεξεργασίας πληροφοριών με την αποτύπωση των τιμών της RGB τριπλέτας. -----	29
Εικόνα 5: Από αριστερά προς τα δεξιά A) Απεικόνιση των grayscale τιμών έντασης της φωτεινότητας. B) Απεικόνιση με mesh plot των τιμών έντασης. -----	30
Εικόνα 6: Απεικόνιση της πυραμιδικής δομής των διαδικασιών της Ψηφιακής Επεξεργασίας Εικόνας. -----	30
Εικόνα 7: Από αριστερά προς τα δεξιά A) Απεικόνιση της συμβατικής μορφής αρίθμησης ενός πίνακα. B) Απεικόνιση της δεικτοδότησης με βάση τις στήλες ενός πίνακα. -----	32
Εικόνα 8: Απεικόνιση της έννοιας των τεσσάρων γειτόνων. Με κόκκινο υποδεικνύεται το εικονοστοιχείο προς εξέταση ενώ με τα υπόλοιπα χρώματα υποδηλώνουν τους τέσσερις γείτονες. -----	33
Εικόνα 9: Απεικόνιση της έννοιας των οκτώ γειτόνων. Με λευκό χρώμα απεικονίζονται οι συμπληρωματικοί τέσσερις γείτονες. -----	34
Εικόνα 10: Από αριστερά προς τα δεξιά A) Αποτύπωση του εύρους των δυνατών τιμών έντασης φωτεινότητας μιας grayscale εικόνας. B) Αποτύπωση του υποσυνόλου των τιμών έντασης που ορίζει την γειννίαση στις Εικόνα 8 και 9. --	35
Εικόνα 11: Από αριστερά προς τα δεξιά Απεικόνιση της γειννίασης-κατά-4 και της γειννίασης-κατά-8. -----	36
Εικόνα 12: Από επάνω προς τα κάτω A) Απεικόνιση της grayscale εικόνας αναφοράς με την μορφή πίνακα στο MATLAB. B) Απεικόνιση του διανύσματος των 3000 στηλών. -----	43
Εικόνα 13: Απεικόνιση μέρους των τιμών έντασης του διανύσματος στήλης. -----	43
Εικόνα 14: Απεικόνιση της εύρεσης τριών διακριτών τιμών έντασης στο πεπερασμένο πλήθος του διανύσματος στήλης. -	43
Εικόνα 15: Από αριστερά προς τα δεξιά A) Η αρχική εικόνα προς επεξεργασία. B) Το ιστόγραμμα της αρχικής εικόνας. 44	
Εικόνα 16: Από αριστερά προς τα δεξιά A) Απεικόνιση λεπτομέρειας της αρχικής εικόνας του ματιού της Εικόνας 4A B) Περιστροφή της αρχικής εικόνας κατά 180 μοίρες. -----	46
Εικόνα 17: Απεικόνιση της ροής των διαδικασιών που λαμβάνουν μέρος για την υλοποίηση του SIFT. -----	48
Εικόνα 18: Απεικόνιση δυο ευδιάκριτα διαχωρίσιμων πληθυσμών. -----	52
Εικόνα 19: Απεικόνιση του βέλτιστου διαχωριστικού επιπέδου, των διανυσμάτων υποστήριξης καθώς και των άνω και κάτω περιθωρίων. -----	53
Εικόνα 20: Από αριστερά προς τα δεξιά A) Απεικόνιση ενός γραμμικού υπερεπίπεδου. B) Απεικόνιση ενός διανύσματος κάθετο στο γραμμικό υπερεπίπεδο. -----	54
Εικόνα 21: Αποτύπωση της απόστασης του σημείου x_1 από το υπερεπίπεδο. -----	55
Εικόνα 22: Το SVM με έναν πυρήνα που δίνεται από την σχέση $\phi a, b = a, b, a^2 + b^2$ δίνοντας ότι $kx, y = x \cdot y + x^2y^2$. Τα δεδομένα εκπαίδευσης αντιστοιχίζονται σε έναν τρισδιάστατο χώρο όπου το υπερεπίπεδο μπορεί να εντοπιστεί εύκολα. -----	61
Εικόνα 23: Απεικόνιση της γκαουσιανής κατανομής πυκνότητας πιθανότητας μιας μονοσήμαντης τυχαίας μεταβλητής. -	66
Εικόνα 24: Δισδιάστατη αναπαράσταση της γκαουσιανής κατανομής πυκνότητας πιθανότητας πολυσήμαντων τυχαίων (εδώ δύο) μεταβλητών. -----	66
Εικόνα 25: Τρισδιάστατη αναπαράσταση της γκαουσιανής κατανομής πυκνότητας πιθανότητας της Εικόνας 24. -----	67
Εικόνα 26: Απεικόνιση των τεσσάρων πληθυσμών προς εξέταση. -----	68
Εικόνα 27: Δισδιάστατη αναπαράσταση της πυκνότητας των τεσσάρων πληθυσμών. -----	68
Εικόνα 28: Τρισδιάστατη αναπαράσταση της πυκνότητας των τεσσάρων πληθυσμών. -----	69
Εικόνα 29: Αποτύπωση του συνδυασμού των τεσσάρων μονοσήμαντων γκαουσιανών κατανομών σε ένα συνολικό μείγμα κατανομών. -----	70

Εικόνα 30: Αποτύπωση των συστάδων όπως αυτές προκύπτουν από την χρήση της συνάρτησης <i>ksdensity</i> . -----	76
Εικόνα 31: Αποτύπωση της διδιάστατης κατανομής των συστάδων με βάση τα αποτελέσματα του <i>GMM</i> . -----	76
Εικόνα 32: Αποτύπωση της τριδιάστατης κατανομής των συστάδων με βάση τα αποτελέσματα του <i>GMM</i> . -----	76
Εικόνα 33: Αποτύπωση της ανάθεσης των δεδομένων στα κέντρα των πληθυσμών καθώς και η ανάλυση με χρήση της μετρικής <i>Silhouette</i> τόσο για τον <i>k-means</i> όσο και για τα <i>GMM</i> . -----	77
Εικόνα 34: Απεικόνιση της έννοιας του μήκους σε μια Πολλαπλότητα. -----	85
Εικόνα 35: Αποτύπωση της παραλληλίας του διανύσματος μετακίνησης προς τον εαυτό του. -----	86
Εικόνα 36: Απεικόνιση της Εκθετικής Απεικόνισης. -----	88
Εικόνα 37: Απεικόνιση της ημι-κωνικής δομής του <i>SPD</i> Πολύπτυχου ενός πίνακα συνδιακύμανσης 2×2 . -----	90
Εικόνα 38: Απεικόνιση της διαδικασίας της Εκθετικής και Λογαριθμικής Απεικόνισης από ένα σημείο του Εφαπτομενικού Χώρου πίσω στο Πολύπτυχο και το αντίστροφο. -----	92
Εικόνα 39: Απεικόνιση του εντοπισμού ακμών με την χρήση του αλγορίθμου <i>Canny</i> . -----	99
Εικόνα 40: Απεικόνιση του διαγράμματος για την υλοποίηση του <i>BoVW</i> σχηματικού. -----	99
Εικόνα 41: Γραφική αναπαράσταση ενός τυπικού Πίνακα Σύγχυσης. -----	100
Εικόνα 42: Αποτύπωση της διαδικασίας μείωσης του σφάλματος της αντικειμενικής συνάρτησης συναρτήσει του πλήθους επαληθεύσεων της συνάρτησης. -----	101
Εικόνα 43: Από αριστερά προς τα δεξιά Α) Αποτύπωση της κατάτμησης της εικόνας σε 0 επίπεδα (Αρχική εικόνα) Β) Αποτύπωση της κατάτμησης της εικόνας σε 1 επίπεδο. Γ) Αποτύπωση της κατάτμησης της εικόνας σε 2 επίπεδα. -----	103
Εικόνα 44: Από αριστερά προς τα δεξιά Α) Αποτύπωση της εύρεσης τριών χαρακτηριστικών και της στάθμισής τους στο επίπεδο 0. Β) Αποτύπωση της εύρεσης τριών χαρακτηριστικών και της στάθμισής τους στο επίπεδο 1. Α) Αποτύπωση της εύρεσης τριών χαρακτηριστικών και της στάθμισής τους στο επίπεδο 2. -----	105
Εικόνα 45: Σχηματική αναπαράσταση του <i>one-vs-all</i> σχηματικού -----	106
Εικόνα 46: Από αριστερά προς τα δεξιά Α) Αποτύπωση της πορείας εκτίμησης των τιμών της αντικειμενικής συνάρτησης συναρτήσει της υπερπαραμέτρου προς βελτιστοποίηση. Β) Αποτύπωση της εξέλιξης της ελάχιστης τιμής της αντικειμενικής συνάρτησης συναρτήσει των αξιολογήσεών της. -----	109
Εικόνα 47: Από αριστερά προς τα δεξιά Α) Αποτύπωση του επιπέδου που ορίζεται για την εύρεση των τιμών της αντικειμενικής συνάρτησης προς βελτιστοποίηση εκπεφρασμένο σε λογαριθμική κλίμακα. Β) Αποτύπωση της εξέλιξης της ελάχιστης τιμής της αντικειμενικής συνάρτησης συναρτήσει των αξιολογήσεών της. -----	109
Εικόνα 48: Απεικόνιση της αποτυχίας της <i>fitgmdist</i> λόγω <i>ill-conditioned</i> πινάκων συνδιακύμανσης. -----	113
Εικόνα 49: Απεικόνιση του <i>Voronoi</i> διαγράμματος με την υπέρθεση των τοπικών περιγραφέων ως έγχρωμες κουκίδες και των κέντρων με μαύρους σταυρούς. -----	116
Εικόνα 50: Αποτύπωση των υπολοίπων ανάμεσα στα κέντρα και τους τοπικούς περιγραφείς αυτών. -----	117
Εικόνα 51: Αποτύπωση των <i>Voronoi</i> κελιών μόνο με τα υπόλοιπα και τα κέντρα. -----	117
Εικόνα 52: Απομόνωση δυο μόνο <i>Voronoi</i> κελιών και αποτύπωση απομακρυσμένη οπτική των τοπικών περιγραφέων με τα κέντρα τους. -----	118
Εικόνα 53: Μεγεθυμένη οπτική του αριστερά <i>Voronoi</i> κελιού με τους τοπικούς περιγραφείς και τα κέντρο του κελιού. Με διακεκομμένες γραμμές αποτυπώνονται τα υπόλοιπα ανάμεσα στους τοπικούς περιγραφείς και το εκάστοτε κέντρο. ---	118
Εικόνα 54: Μεγεθυμένη οπτική του δεξιά <i>Voronoi</i> κελιού με τους τοπικούς περιγραφείς και τα κέντρο του κελιού. Με διακεκομμένες γραμμές αποτυπώνονται τα υπόλοιπα ανάμεσα στους τοπικούς περιγραφείς και το εκάστοτε κέντρο. ---	119
Εικόνα 55: Αποτύπωση του διαγράμματος ροής που ακολουθεί ένα σύστημα επαλήθευσης. -----	121
Εικόνα 56: Αποτύπωση του διαγράμματος ροής που ακολουθεί ένα σύστημα ταυτοποίησης. -----	122
Εικόνα 57: Απεικόνιση της κατανομής των αυθεντικών και των πλαστογραφημένων κατανομών, του σημείου στο οποίο τοποθετείται το κατώφλι απόφασης καθώς και των σφαλμάτων <i>FNMR</i> και <i>FMR</i> . -----	124
Εικόνα 58: Αποτύπωση της <i>ROC</i> καμπύλης και του αντίστοιχου <i>EER</i> σημείου που προσδιορίζεται ως το σημείο ισότητας του <i>FAR</i> με το <i>FRR</i> . -----	124
Εικόνα 59: Παραδείγματα υπογραφών από την βάση δεδομένων <i>CEDAR</i> . Στο άνωθεν τμήμα εντοπίζεται ένα σύνολο από αυθεντικές υπογραφές ενώ στο κάτωθεν τμήμα εντοπίζεται ένα σύνολο από πλαστογραφημένες υπογραφές. -----	126
Εικόνα 60: Απεικόνιση του Χώρου Ανομοιότητας που επάγεται από την χρήση του μετασχηματισμού διχοτόμησης. ---	128
Εικόνα 61: Αποτύπωση του σχηματισμού του περιγραφέα περιοχών ως πίνακα συνδιακύμανσης. Για ένα συγκεκριμένο εικονοστοιχείο που υποδεικνύεται με «x», δημιουργείται μια στοιβάδα από τιμές που προκύπτουν μέσω των 9 φίλτρων που εφαρμόζονται στην αρχική εικόνα. -----	130
Εικόνα 62: Αποτύπωση της αναγωγής των σημείων της Πολλαπλότητας στον κοινό Εφαπτομενικό Χώρο που επάγεται από τον μοναδιαίο πίνακα. Με μαύρο, σημειώνεται το σημείο που επάγεται από τον μοναδιαίο πίνακα σχηματίζοντας τον	

κοινό πόλο. Με κόκκινο χρώμα στα αριστερά του κοινού πόλου εντοπίζεται το σημείο της Πολλαπλότητας X και το διάνυσμα που δημιουργείται μέσω της λογαριθμικής αντιστοίχισής του. Με πράσινο χρώμα στα δεξιά του κοινού πόλου εντοπίζεται το σημείο της Πολλαπλότητας Y και το διάνυσμα που δημιουργείται μέσω της λογαριθμικής αντιστοίχισής του. ----- 131

Εικόνα 63: Αποτύπωση της δεικτοδότησης των δεδομένων για την εφαρμογή του μετασχηματισμού διχοτόμησης στο σύνολο της εκπαίδευσης για τις γνήσιες υπογραφές. ----- 132

Εικόνα 64: Αποτύπωση του διανύσματος \mathbf{xy} , της γεωδαισιακής που το ακολουθεί και της διαφοράς των σημείων y και x . ----- 136

Εικόνα 65: Απεικόνιση του διανύσματος y που υποδεικνύει προς την κατεύθυνση του σημείου Y ενώ εντοπίζεται στον Εφαπτομενικό Χώρο του X . Παράλληλα απεικονίζεται και η γεωδαισιακή που συνδέει μη-γραμμικά τα δυο σημεία του Πολύπτυχου. Με διακεκομμένη γραμμή σημειώνεται ο μετασχηματισμός της βάσης του Y στο σύστημα συντεταγμένων του Εφαπτομενικού Χώρου του X . ----- 138

Εικόνα 66: Απεικόνιση της έννοιας της αμφιμονοσήμαντης συνάρτησης (bijection). ----- 145

Εικόνα 67: Απεικόνιση του χώρου των διαφορομορφισμών. ----- 146

Εικόνα 68: Απεικόνιση ενός Χάρτη Συντεταγμένων. ----- 148

Εικόνα 69: Αποτύπωση των μεταβατικών απεικονίσεων ανάμεσα σε δυο Χάρτες συντεταγμένων ----- 148

Εικόνα 70: Απεικόνιση του ορισμού μιας λείας συνάρτησης. ----- 150

Εικόνα 71: Αποτύπωση του ορισμού των Λείων Απεικονίσεων. ----- 151

Εικόνα 72: Αποτύπωση του γεωμετρικού εφαπτομενικού διανύσματος. ----- 152

Εικόνα 73: Απεικόνιση του μοναδιαίου ακτινικού διανύσματος, u (μαύρες διακεκομμένες γραμμές), το οποίο είναι ορθοκανονικό σε ένα τυχαίο διάνυσμα, v (λευκό χρώμα), που έχει ως αρχή το σημείο a και αυθαίρετη κατεύθυνση. ---- 153

Κατάλογος Ορισμών

Ορισμός 1: Θεώρημα Nyquist (ή Θεώρημα Δειγματοληψίας) [5],[6] 25

Ορισμός 2: 4-Γείτονες..... 33

Ορισμός 3: 8-Γείτονες..... 34

Ορισμός 4: Γειτνίαση-κατά-4 36

Ορισμός 5: Γειτνίαση-κατά-8..... 36

Ορισμός 6: Πορεία..... 37

Ορισμός 7: Σύνδεση εικονοστοιχείων 37

Ορισμός 8: Γραμμικοί τελεστές 38

Ορισμός 9: Ιστόγραμμα..... 43

Ορισμός 10: Υπερεπίπεδο..... 54

Ορισμός 11: Θεώρημα ενημέρωσης των μέσων τιμών του GMM 74

Ορισμός 12: Θεώρημα ενημέρωσης των πινάκων συνδιακύμανσης του GMM 74

Ορισμός 13: Θεώρημα ενημέρωσης των βαρών μίξης 74

Ορισμός 14: Μετρικός Χώρος (Metric Space) 82

Ορισμός 15: Μετρική 83

Ορισμός 16: Riemannian Μετρική 83

Ορισμός 17: Riemannian Πολλαπλότητα 84

Ορισμός 18: Μήκος σε μια Πολλαπλότητα 84

Ορισμός 19: Γεωδαισιακή..... 85

Ορισμός 20: Θεώρημα μοναδικότητας Γεωδαισιακών..... 86

Ορισμός 21: Διάστημα μέγιστης ύπαρξης (maximal existence interval)..... 87

Ορισμός 22: Εκθετική Απεικόνιση 87

Ορισμός 23: Συμμετρικός Θετικά Ορισμένος Πίνακας..... 89

Ορισμός 24: Το SPD Πολύπτυχο..... 89

Ορισμός 25: Η SPD Πολλαπλότητα ως διαφορίσιμο Riemannian Πολύπτυχο 90

Ορισμός 26: Γεωδαισιακή στο SPD Πολύπτυχο..... 91

Ορισμός 27: Εκθετική Απεικόνιση στο SPD Πολύπτυχο 92

Ορισμός 28: Λογαριθμική Απεικόνιση..... 92

Ορισμός 29: Διανυσματική Αναπαράσταση στον Εφαπτομενικό Χώρο με ορθοκανονικές συντεταγμένες	93
Ορισμός 30: Τυχαίες Πλαστογραφίες (Random Forgeries).....	125
Ορισμός 31: Απλές Πλαστογραφίες (Simple Forgeries).....	125
Ορισμός 32: Εξειδικευμένες Πλαστογραφίες (Skilled Forgeries).....	126
Ορισμός 33: Ορθοκανονικές συντεταγμένες.....	136
Ορισμός 34: Πολλαπλότητα ή Πολύπτυχο (Manifold).....	143
Ορισμός 35: Ομοιομορφισμός.....	144
Ορισμός 36: Διαφορομορφισμός.....	145
Ορισμός 37: Τοπολογική Πολλαπλότητα n -διαστάσεων.....	146
Ορισμός 38: Λεία Πολλαπλότητα (Smooth Manifold).....	147
Ορισμός 39: Χάρτες Συντεταγμένων	147
Ορισμός 40: Μεταβατικές Απεικονίσεις (Transition Maps).....	148
Ορισμός 41: Λεία συμβατές Μεταβατικές Απεικονίσεις.....	149
Ορισμός 42: Άτλαντας.....	149
Ορισμός 43: Λείοι Άτλαντες.....	149
Ορισμός 44: Ολοκληρωμένος Άτλαντας (Maximal/Complete Atlas).....	149
Ορισμός 45: Λεία δομή Πολλαπλότητας (revisited).....	150
Ορισμός 46: Λεία Συνάρτηση.....	150
Ορισμός 47: Λείες Απεικονίσεις (Smooth Maps).....	151
Ορισμός 48: Εφαπτομενικό Διάνυσμα.....	152
Ορισμός 49: Παραγωγή σημείου σε Λείο Πολύπτυχο (Derivation at point a of a Smooth Manifold).....	153
Ορισμός 50: Εφαπτομενικός Χώρος ενός Πολύπτυχου (Tangent Space of a Manifold).....	154
Ορισμός 51: Εφαπτομενική Δέσμη (Tangent Bundle).....	154
Ορισμός 52: Η Εφαπτομενική Δέσμη ως Λείο Πολύπτυχο	155

Κατάλογος Αλγορίθμων

Αλγόριθμος 1: Συσταδοποίηση κατά k -Means.....	64
Αλγόριθμος 2: Αποτύπωση του EM αλγορίθμου μόνο για διαγώνιους πίνακες συνδιακύμανσης.....	75
Αλγόριθμος 3: Αποτύπωση του sEM όπως έχει διαρθρωθεί βασιζόμενος στον ψευδοκώδικα που διατίθεται στην δημοσίευση [49].....	80
Αλγόριθμος 4: Ψευδοκώδικας για τον αλγόριθμο BoVW.....	101
Αλγόριθμος 5: Ψευδοκώδικας για την εκτέλεση του πειράματος.....	108
Αλγόριθμος 6: Ψευδοκώδικας για την εκτέλεση του πειράματος.....	119
Αλγόριθμος 7: Βασικός αλγόριθμος για το πρώτο Riemannian σενάριο.....	133
Αλγόριθμος 8: Αλγόριθμος για την συνάρτηση Train_Val_Split.m.....	134
Αλγόριθμος 9: Αλγόριθμος για την συνάρτηση OmegaPlusFormation.m.....	134
Αλγόριθμος 10: Αλγόριθμος για την συνάρτηση OmegaMinusFormation.m.....	134
Αλγόριθμος 11: Αλγόριθμος για την συνάρτηση RefsQuestionNV.m.....	135
Αλγόριθμος 12: Αλγόριθμος για το δεύτερο Riemannian σενάριο.....	139
Αλγόριθμος 13: Αλγόριθμος για την συνάρτηση OmegaPlusFormation.m.....	139
Αλγόριθμος 14: Αλγόριθμος για την συνάρτηση OmegaMinusFormation.m.....	139
Αλγόριθμος 15: Αλγόριθμος για την συνάρτηση RefsQuestionNV.m.....	140

ΕΙΣΑΓΩΓΗ

Τα τελευταία χρόνια ο κόσμος έχει υπάρξει μάρτυρας μιας τεράστιας τεχνολογικής εξέλιξης στον τομέα της επιστήμης των υπολογιστών. Έννοιες όπως Τεχνητή Νοημοσύνη, Μηχανική Μάθηση και Υπολογιστική Όραση κερδίζουν καθημερινά έδαφος στον παγκόσμιο στίβο ενδιαφέροντος της επιστημονικής κοινότητας, διευρύνοντας το πλαίσιο εφαρμογής τους σε οποιοδήποτε τομέα επωφελείται από την συλλογή και επεξεργασία δεδομένων για την παροχή υπηρεσιών που αποσκοπούν στην ενίσχυση του βιοτικού επιπέδου και την επέκταση της ανθρώπινης γνώσης. Από την χρήση δικτύων βαθιάς μάθησης με σκοπό την πρωτεϊνοσύνθεση μέχρι και τον σχηματισμό εφαρμογών που μπορούν να περιγράψουν φωνητικά το περιβάλλον που τους υποδεικνύει ο χρήστης μέσω μιας κάμερας, ο 21^{ος} αιώνας βρίθεται από παραδείγματα που παγιώνουν την δεσπόζουσα θέση της Υπολογιστικής Νοημοσύνης στην ανθρώπινη καθημερινότητα σε όλο το σύνολό της με τρόπους οι οποίοι είτε γίνονται έντονα αισθητοί είτε περνούν απαρατήρητοι από το ευρύ κοινό. Ένας από τους πιο υπολογιστικά απαιτητικούς αλλά συνάμα καινοτομικά καρποφόρους τομείς είναι αυτός της Υπολογιστικής Όρασης όπου, όπως υποδηλώνουν και οι λέξεις, η ανθρώπινη όραση αποκρυπτογραφείται και μετουσιώνεται σε ένα είδος πληροφορίας που μπορεί να αξιοποιηθεί από μια υπολογιστική μηχανή, να εξαχθεί νόημα και να ληφθούν αποφάσεις για την επιτέλεση αυτόματων διαδικασιών. Με μια πλούσια ιστορία που εντοπίζεται να ξεκινά στο, όχι και τόσο μακρινό, 1950, η σύλληψη της ιδέας της Υπολογιστικής Όρασης κεντροποιείται στην προσπάθεια αντίληψης της λειτουργίας του φυσικού οπτικού συστήματος και ενισχύεται με επιπρόσθετες μεθόδους επεξεργασίας με απώτερο σκοπό την απόδοση νοήματος του πραγματικού κόσμου στις μηχανές. Βασικός θεμελιωτής της γεφύρωσης ανάμεσα στην ακατέργαστη οπτική πληροφορία και την μετατροπή αυτής σε μια αξιοποιήσιμη μορφή για τις υπολογιστικές μηχανές αποτέλεσε ο κλάδος της επιστήμης των υπολογιστών γνωστός ως Ψηφιακή Επεξεργασία Εικόνας, αξιοποιώντας κάποια από τα βασικά αξιώματα των μαθηματικών σε συνδυασμό με την απεικονιστική δυνατότητα και την ευρωστία σε θέματα μνήμης και υπολογισμών που προσφέρουν οι υπολογιστικές διατάξεις. Η πραγματική άνθηση επήλθε από το αμάλγαμα ανάμεσα στην προαναφερθείσα και την ανέγερση τεχνικών Μηχανικής Μάθησης, όπου στατιστικά μοντέλα εκπαιδεύονται από το πλαίσιο των διαθέσιμων δεδομένων επιδιώκοντας την σύνθεση ενός συστήματος ικανού να γενικεύει σε άγνωστα δεδομένα. Διαφαίνεται πως ο δυναμικά εξελισσόμενος κλάδος της Τεχνητής Νοημοσύνης, με έμφαση στην Υπολογιστική Όραση, διαδραματίζει ολοένα και πιο σημαντικό ρόλο στην κατανόηση του οπτικού κόσμου. Η ικανότητα των μηχανών να ερμηνεύουν και να κατανοούν τον κόσμο γύρω τους ανοίγει νέες δυνατότητες για έρευνα και καινοτομία σε πληθώρα πεδίων, από την αναγνώριση προσώπων έως την ιατρική απεικόνιση και την αυτόνομη οδήγηση. Συνεχείς εξελίξεις στους αλγόριθμους μηχανικής μάθησης, σε συνδυασμό με την ολοένα αυξανόμενη διαθεσιμότητα δεδομένων, άρουν τους περιορισμούς της ανθρώπινης φυσιολογίας αποκαλύπτοντας μοτίβα και

γνώσεις που θα ήταν αδύνατον να εντοπιστούν από το ανθρώπινο μάτι ωθώντας την επιστημονική κοινότητα και κατ' επέκτασιν την κοινωνία, προς νέα όρια. Με αξιοσημείωτες προκλήσεις όπως η διαχείριση μεγάλων όγκων δεδομένων, η εξασφάλιση ηθικής χρήσης αλλά και το κόστος της καινοτομίας, η Υπολογιστική Όραση καλεί σε μια βαθύτερη κατανόηση της σχέσης μεταξύ ανθρώπου και μηχανής, ανοίγοντας τον δρόμο για μια νέα εποχή συνεργασίας και αλληλεπίδρασης.

Αντικείμενο της διπλωματικής εργασίας

Τι κοινά στοιχεία μπορεί να παρουσιάζουν εικόνες που αναπαριστούν ένα πορτρέτο ενός προσώπου, ένα αυθαίρετο παρασκήνιο, ένα βιολί, ένα φυσικό τοπίο ή μια υπογραφή; Ποια είναι άραγε η ειδοποιός διαφορά των χαρακτηριστικών που ξεχωρίζουν το ένα αντικείμενο ενδιαφέροντος από το άλλο; Οι απαντήσεις σε ερωτήματα τέτοιας φύσης αποτελούν τετριμμένες για έναν άνθρωπο, όμως το ίδιο δεν μπορεί να λεχθεί για μια υπολογιστική μηχανή που έχει αναλάβει να φέρει εις πέρας ένα παρόμοιο έργο καθώς η τελευταία θα πρέπει να υποβληθεί σε μια σειρά διαδικασιών με απώτερο σκοπό την λήψη μιας απόφασης που θα υποστηρίζεται εννοιολογικά από την εξαχθείσα γνώση πίσω από τα παρεχόμενα δεδομένα. Κι ενώ η ανθρώπινη νοημοσύνη είναι ικανή να επιτελέσει με υψηλή ακρίβεια και επαναληψιμότητα διαδικασίες όπως η αναγνώριση ή η ταυτοποίηση τυπικών κατηγοριών, για παράδειγμα η διαφορά ανάμεσα σε μια γάτα κι ένα σκύλο, αλλά και πιο σύνθετων, όπως ο διαχωρισμός ανάμεσα σε δυο δίδυμα αδέρφια (δεδομένου ότι έχει δοθεί κάποια επιπρόσθετη πληροφορία όπως το όνομα του κάθε προσώπου), η ενασχόληση με τομείς που στηρίζονται και επωφελούνται από την υψηλή διακριτικότητα ανάμεσα σε γνήσιες και πλαστές πληροφορίες, λόγου χάρη ο τομέας της ασφάλειας ή της εγκληματολογίας, απαιτούν σχολαστικότητα και εκτεταμένη προσοχή στην λεπτομέρεια: χαρακτηριστικά τα οποία ο άνθρωπος συχνά επηρεαζόμενος από ενδοπροσωπικούς αλλά και εξωγενείς παράγοντες μπορεί να υπονομεύσει οδηγούμενος σε λάθη που ενδεχομένως να αποβούν μοιραία. Η χρήση αυτοματοποιημένων διαδικασιών που βασίζονται σε εμπειριστατώμενο μαθηματικό υπόβαθρο και ακριβείς οδηγίες, γνωστές και ως *αλγόριθμοι*, μπορούν και εφοπλίζουν τις υπολογιστικές μηχανές με τα απαραίτητα εφόδια που άρουν τις ανησυχίες και τους προβληματισμούς που οφείλονται σε ανθρώπινα σφάλματα τέτοιας φύσης. Το αντικείμενο της παρούσας διπλωματικής εργασίας ξεκινά με μια ενδελεχή μελέτη αλγορίθμων Υπολογιστικής Όρασης που βασίζονται σε δεδομένα που εντοπίζονται στον Ευκλείδειο Χώρο καθώς και μεθοδολογίες και τεχνικές που εκμεταλλεύονται κωδικοποιήσεις για την μείωση του υπολογιστικού κόστους με την παράλληλη προσπάθεια διατήρησης της περιγραφικότητας και οδηγείται με φυσικό τρόπο στην συνειδητοποίηση ότι τα βέλτιστα δυνατά αποτελέσματα επιτυγχάνονται αξιοποιώντας τον εγγενή χώρο τον οποίο συνιστούν τα δεδομένα: τον Riemannian Χώρο.

Σκοπός και στόχοι

Η παρούσα εργασία αποτελεί ένα συνονθύλευμα από επιμέρους μη συσχετιζόμενους στόχους, οι οποίοι αθροιστικά οδηγούν στην υιοθέτηση τεχνικών που ενισχύουν την ακρίβεια των μοντέλων Υπολογιστικής Όρασης για δυο διακεκριμένες εφαρμογές.

Ξεκινώντας από τον Ευκλείδειο Χώρο και μελετώντας εφαρμογές ταξινόμησης εικόνων:

1. Ο πρώτος στόχος είναι η μελέτη και αξιολόγηση της διαφοράς ανάμεσα στην ιστογραμματική αναπαράσταση μέσω Οπτικών Λέξεων μιας συλλογής εικόνων που βασίζονται στην ολική περιγραφή αυτών και την αντίστοιχη ιστογραμματική αναπαράσταση που προκύπτει από την συνένωση ιστογραμμάτων τοπικών χαρακτηριστικών, όπως αυτά εντοπίζονται από την σταδιακή τμηματοποίηση των εικόνων σε μικρότερα κομμάτια, οδηγώντας στον σχηματισμό μιας εικονικής χωρικής πυραμίδας.
2. Παραμένοντας στον ίδιο γεωμετρικό Χώρο, ο δεύτερος στόχος είναι η επέκταση των Οπτικών Λεξικών και της περιγραφικότητας που αυτά παράγουν με την χρήση πυρήνων Fisher, όπου η πιθανοτική φύση παραγωγής δεδομένων των Γκαουσιανών Μοντέλων Μίξης σε συνεργασία με την αξιοποίηση των παραγώγων πρώτης τάξης των παραμέτρων των πολυμεταβλητών Γκαουσιανών κατανομών, οδηγεί στην σύνθεση διανυσματικών αναπαραστάσεων των εικόνων με αξιόπαινα αποτελέσματα ακρίβειας.
3. Τρίτος στόχος είναι η διερεύνηση μιας τεχνικής που, εμπνεόμενη από τους πυρήνες Fisher, διατηρεί την διανυσματική αναπαράσταση των εικόνων όπως η τελευταία προκύπτει από το άθροισμα του υπολοίπου της διαφοράς ανάμεσα στην εκάστοτε εικόνα και τα κοντινότερα κέντρα στα οποία η τελευταία έχει βρεθεί ότι εΐθισται να ανατεθεί. Η προαναφερθείσα μέθοδος είναι κομβικής σημασίας καθώς όχι μόνο καταφέρνει να αυξήσει την ακρίβεια ταξινόμησης αλλά συνεισφέρει δραματικά στην μείωση της απαιτούμενης μνήμης

Περνώντας στον Riemannian χώρο, η εφαρμογή που τίθεται υπό εξέταση είναι η χρήση της πολλαπλότητας των Συμμετρικών Θετικά Ορισμένων Πινάκων για την βιομετρική ταυτοποίηση χειρόγραφων υπογραφών. Συγκεκριμένα:

4. Ερευνάται η αποτελεσματικότητα της χρήσης αποκλειστικά των προβαλλόμενων διανυσμάτων των σημείων της πολλαπλότητας στον Εφαπτομενικό Χώρο εφαρμόζοντας τον Μετασχηματισμό Διχοτόμησης (*Dichotomy Transform*) ως μέτρο της ανομοιότητας ανάμεσα σε γνήσιες και πλαστές υπογραφές γραφών.
5. Ερευνάται η αποτελεσματικότητα της εφαρμογής της Λογαριθμικής Αντιστοίχισης, η φυσική έννοια της οποίας ορίζει την αφαίρεση ανάμεσα σε δυο σημεία της Πολλαπλότητας, χρησιμοποιώντας ως σταθερό σημείο στην Πολλαπλότητα των Συμμετρικών Θετικά Ορισμένων Πινάκων πάντοτε τις γνήσιες υπογραφές του εκάστοτε γραφέα και προβάλλοντας στον Εφαπτομενικό Χώρο τις πλαστές υπογραφές του ίδιου γραφέα.

Μεθοδολογία

Η μεθοδολογία της παρούσας εργασίας μιμείται τα στάδια που είθισται να παρουσιάζει μια ερευνητική διαδικασία. Με γνώμονα το τελευταίο, το πρώτο στάδιο περιλαμβάνει την μελέτη της σχετικής βιβλιογραφίας, ξεκινώντας με την κατανόηση της λειτουργίας του ανθρώπινου οπτικού συστήματος, όπως και των μέσων που χρησιμοποιούνται για την μετατροπή των εικόνων σε επεξεργάσιμη πληροφορία, και μετουσιώνεται στην δημιουργία του απαραίτητου θεωρητικού υποβάθρου για τις κοινές πρακτικές και μεθοδολογίες που επιτάσσει η Ψηφιακή Επεξεργασία Εικόνων, οι τυπικοί στατιστικοί αλγόριθμοι Μηχανικής Μάθησης για την εξαγωγή νοήματος πίσω από τα δεδομένα αλλά και τα μαθηματικά που πλαισιώνουν τον μη-γραμμικό Riemannian Χώρο. Για να μπορέσουν οι ιδέες να λάβουν μορφή και να αποδοθεί η θεωρία στην πράξη, επιτάσσεται το πρόγραμμα εκείνο που φέρει μεγαλύτερης ευελιξίας σε προσωπική χρήση του συγγραφέα, με το προκείμενο να αποτελεί το υπολογιστικό περιβάλλον του MATLAB. Αν και μεγάλη μερίδα της προγραμματιστικής κοινότητας εξοστρακίζει τη χρήση του προϊόντος της MathWorks για διαδικασίες Μηχανικής Μάθησης, Βαθιάς Μάθησης ακόμη και Υπολογιστικής Όρασης, η απλοποιημένη δομή της διεπαφής με τον χρήστη σε συνδυασμό με τις επιπρόσθετες μορφές διαχείρισης δεδομένων και ελέγχου συναρτήσεων που παρέχει σε σχέση με άλλες γλώσσες, ενισχυόμενο από το γεγονός ότι από την φύση του έχει σχεδιαστεί ώστε να διαχειρίζεται όλες τις μεταβλητές με την μορφή πινάκων αξιοποιώντας στο έπακρον στοιχεία της Γραμμικής Άλγεβρας και υποστηριζόμενο από μια ενεργή διαδικτυακή κοινότητα με πλούσιο documentation, το καθιστούν ένα αξιόπιστο και αποτελεσματικό εργαλείο με αδιαμφισβήτητη διεθνή αναγνωσιμότητα στον τομέα της έρευνας και της καινοτομίας.

Καινοτομία

Οι καινοτομίες που φέρει η παρούσα διπλωματική εργασία περιστρέφονται γύρω από δυο κεντρικούς άξονες. Ο πρώτος άξονας αφορά την αξιοποίηση της Παράλληλης Επεξεργασίας και των Καρτών Γραφικών για την μείωση του χρόνου λήψης των αποτελεσμάτων ενώ ταυτόχρονα επιτάσσονται και τεχνικές βελτιστοποίησης των υπερπαραμέτρων των ταξινομητών με σκοπό την περαιτέρω ενίσχυση της αποδοτικότητας των μεθοδολογιών που ερευνώνται. Ενδεικτικά παραδείγματα αποτελούν τα παρακάτω:

- Στο κομμάτι της ταξινόμησης των σκηνών με την χρήση της μεθόδου των πυρήνων Fisher, η καινοτομία έγκειται στην επιτάχυνση της παροχής αποτελεσμάτων από τα Γκαουσιανά Μοντέλα Μίξης. Το υψηλό υπολογιστικό κόστος σε συνδυασμό με την περιορισμένη, προς εκτέλεση μόνο στην CPU, έκδοση που παρέχει η MathWorks για τον επαναληπτικό αλγόριθμο Μεγιστοποίησης Προσδοκιών (*Expectation Maximization*) επιφέρει σημαντικό τροχοπέδη στην περάτωση του τελικού στόχου. Εντοπίζοντας τα προηγούμενα προβλήματα υλοποιείται συνάρτηση που αξιοποιεί τους υπολογιστικούς πυρήνες της GPU για την επιτάχυνση του γενικευμένου EM (*GMM_NV*). Επιπρόσθετα σχηματίζεται ο βηματικός-EM (*stepwise-EM*), μια online παραλλαγή του τυπικού αλγορίθμου, ο οποίος συνεισφέρει στην ταχύτερη αλλά και αποδοτικότερη σύγκλιση σε καλύτερες λύσεις.

Ο δεύτερος άξονας καινοτομιών αφορά αποκλειστικά την θεωρητική αξιοποίηση του Riemannian Χώρου. Εν γνώσει του συγγραφέα, δεν υπάρχει κάποια συνεισφορά στην παγκόσμια βιβλιογραφία που να αποτυπώνει και να εφαρμόζει τις παρακάτω ιδέες:

- *Αξιοποίηση αποκλειστικά των προβαλλόμενων διανυσμάτων των σημείων της πολλαπλότητας στον Εφαπτομενικό Χώρο εφαρμόζοντας τον Μετασχηματισμό Διχοτόμησης (Dichotomy Transform) ως μέτρο της ανομοιότητας ανάμεσα σε γνήσιες και πλαστογραφημένες υπογραφές γραφών.*
- *Έρευνα της αποτελεσματικότητας της εφαρμογής της Λογαριθμικής Απεικόνισης, η φυσική έννοια της οποίας ορίζει την αφαίρεση ανάμεσα σε δυο σημεία της Πολλαπλότητας, χρησιμοποιώντας ως σταθερό σημείο στην Πολλαπλότητα των Συμμετρικών Θετικά Ορισμένων Πινάκων πάντοτε τις γνήσιες υπογραφές του εκάστοτε γραφέα και προβάλλοντας στον Εφαπτομενικό Χώρο τις πλαστές υπογραφές του ίδιου γραφέα.*

Δομή

Η παρούσα διπλωματική εργασία συνίσταται από πέντε διακριτά κεφάλαια. Στο πρώτο κεφάλαιο παρατίθενται οι τρόποι με τους οποίους μια εικόνα μετατρέπεται από μια ακατέργαστη σε μια επεξεργάσιμη μορφή πληροφορίας που δύναται να γίνει αντιληπτή από τα υπολογιστικά συστήματα, αναδεικνύονται τυπικές συμβάσεις στην δεικτοδότηση και διαμορφώνεται ένα στιβαρό θεωρητικό υπόβαθρο των τεχνικών επεξεργασίας εικόνας που αποτελούν τα θεμέλια όλων των σύγχρονων αρχιτεκτονικών Υπολογιστικής Όρασης. Το δεύτερο κεφάλαιο, ακολουθώντας την επαγωγική μέθοδο, ξεκινά με την παράθεση των βασικών κατηγοριών αλγορίθμων μάθησης και κεντροποιείται στα μοντέλα μάθησης εκείνα που αποτέλεσαν τα βασικά εργαλεία περάτωσης του συνόλου της εργασίας. Ειδικά, διαγράφονται ο αλγόριθμος των Μηχανών Υποστήριξης Διανυσμάτων (*Support Vector Machines*), ο αλγόριθμος μη επιβλεπόμενης μάθησης των Γκαουσιανών Μοντέλων Μίξης (*Gaussian Mixture Models*) καθώς και ο αλγόριθμος του βηματικού-EM (*stepwise-EM*). Το τρίτο κεφάλαιο αποτελεί τον κορμό της εργασίας, καθώς ο αναγνώστης μπορεί να καταφύγει σε αυτό με σκοπό να «εφοπλιστεί» με τα απαραίτητα μαθηματικά εργαλεία που ενδεχομένως να χρειαστεί για να μπορέσει να αντιληφθεί πλήρως τις έννοιες που διακατέχουν τον Ευκλείδειο Χώρο, να οδηγηθεί στην αντίληψη της σημαντικότητας των αξιωμάτων της Γραμμικής Άλγεβρας και τον κομβικό ρόλο που αυτά διαδραματίζουν στην κατανόηση του χώρου των Συμμετρικών Θετικά Ορισμένων Πινάκων. Το τέταρτο κεφάλαιο αποτυπώνει το σύνολο των αλγορίθμων που υλοποιήθηκαν τόσο στον Ευκλείδειο όσο και στον Riemannian Χώρο. Κλείνοντας, στο πέμπτο, και τελευταίο κεφάλαιο, αποτυπώνονται τα συγκεντρωτικά πορίσματα της μελέτης που εκπονήθηκε μαζί με επιπρόσθετες μεθόδους για επέκταση αυτής της εργασίας ενώ στα παραρτήματα παρατίθενται ένα εκτεταμένο θεωρητικό υπόβαθρο για τις έννοιες που εντοπίζονται στο τρίτο κεφάλαιο.

1 ΚΕΦΑΛΑΙΟ 1^ο : Ψηφιακή Επεξεργασία Εικόνας

Το κεφάλαιο αυτό αποσκοπεί στην πλαισίωση του θεωρητικού υποβάθρου για τις κοινότητες τεχνικές που εφαρμόζονται στον τομέα της Ψηφιακής Επεξεργασίας Εικόνας. Αφετηρία της ανάλυση αποτελεί η αποτύπωση της πάγιας και υψίστης σημασίας θέση της όρασης στην καθημερινότητα του ανθρώπου, συνεχίζει με την παράθεση του ορισμού της Ψηφιακής Επεξεργασίας Εικόνας και βαθμιαία σκιαγραφεί τους πυλώνες του τομέα με έννοιες όπως *εικονοστοιχεία*, *γεινίαση εικονοστοιχείων*, *ιστογράμματα* και ολοκληρώνει με την αποτύπωση της έννοιας των χαρακτηριστικών.

1.1 Βασικές αρχές της Ψηφιακής Επεξεργασίας Εικόνας

Η ανθρώπινη καθημερινότητα, σε όλες τις εκφάνσεις αλλά και τα στάδιά της, αποτελεί ένα σύνολο περίπλοκων και πολυδιάστατων μεταβλητών. Κι ενώ η συσσώρευση συγκεκριμένων μεταβλητών υπό ένα ενοποιημένο περιεχόμενο ορίζει αυτό που είναι γνωστό ως *πρόβλημα*, το προηγούμενο τοποθετείται εκτός περιεχομένου όταν δεν υπάρχει κάποιου είδους μετρικής που να δίνει νόημα και να μπορεί να το ερμηνεύσει με τρόπους τους οποίους το ανθρώπινο είδος να μπορεί να αντιληφθεί και να συλλάβει ώστε να αντιμετωπίσει αποτελεσματικά. Η ίδια η αντίληψη ορίζεται ως η διαδικασία συλλογής, οργάνωσης και ερμηνείας των πληροφοριών που συγκεντρώνονται από το περιβάλλον, μέσω των διακριτών αισθητηρίων οργάνων του ανθρώπινου σώματος, με την μορφή βιολογικών σημάτων. Οι βασικότερες των αισθήσεων, όπως απαριθμήθηκαν και αναγνωρίστηκαν από τον Αριστοτέλη, είναι η αφή, η ακοή, η όσφρηση, η γεύση και η όραση, με την τελευταία να θεωρείται ενδεχομένως η πιο σημαντική και η πιο περίπλοκη. Μάλιστα, στο ερευνητικό άρθρο με τίτλο: *«Multisensory task demands temporally extend the causal requirement for visual cortex in perception»* [1] υπογραμμίζεται ότι ο χρόνος που απαιτείται για τον εγκέφαλο ώστε να μπορέσει να λάβει μια απόφαση που βασίζεται σε ένα οπτικό ερεθίσμα, εξαρτάται από επιπρόσθετα απτικά και ακουστικά ερεθίσματα, φωτίζοντας άμεσα ότι *η πολυπλοκότητα της όρασης εκτείνεται πέρα από την επεξεργασία των οπτικών πληροφοριών καθώς ενσωματώνει λεπτομέρειες κι από άλλες αισθήσεις*.

Η όραση συνεισφέρει επιπρόσθετα στην εξελικτική πορεία του ανθρώπου σε τομείς όπως η ασφάλεια και η επιβίωση, ο προσανατολισμός και η αλληλεπίδραση αλλά ίσως το πιο σημαντικό είναι η επιρροή που έχει στην επικοινωνία και την μάθηση. Έρευνες έχουν αποδείξει ότι τα οπτικά ερεθίσματα μπορούν να ενισχύσουν την ανάκτηση πληροφοριών και την αφομοίωση αυτών χάρις στην εγγενή φύση του εγκεφάλου να λειτουργεί ως ένας μεγάλος επεξεργαστής εικόνων μειώνοντας έτσι δραστικά τον χρόνο μάθησης [2],[3],[4]. Συγκεκριμένα, σε πείραμα που διεξήχθη από τους McDaniel M.A και Einstein G.O [2], δόθηκε η οδηγία σε μαθητές να απομνημονεύσουν πολλές κατηγορίες από τρεις λέξεις, με ενδεικτικά παραδείγματα: σκυλί, ποδήλατο και δρόμος. Οι μαθητές που προσπάθησαν να απομνημονεύσουν τις λέξεις επαναλαμβάνοντάς τες ξανά και ξανά είχαν

χαμηλότερη επίδοση από αυτούς που προσπάθησαν να δημιουργήσουν μια οπτική συσχέτιση με τις τρεις αυτές λέξεις, όπως για παράδειγμα με τον σχηματισμό ενός σεναρίου κατά το οποίο έκαναν ποδήλατο ή έβγαζαν βόλτα τον σκύλο τους. Τα προηγούμενα πορίσματα αποτελούν φυσικό επόμενο αν αναλογιστεί κανείς ότι οι λέξεις είναι αφηρημένες έννοιες που δύσκολα μπορούν να διατηρηθούν από τον εγκέφαλο, ενώ οι εικόνες αποτελούν μια συμπαγή και παράλληλα περιγραφική αναπαράσταση πληροφορίας.

Κι ενώ αναδεικνύεται ότι η όραση αποτελεί έναν θαυμαστό μηχανισμό με τον οποίο η φύση έχει προικίσει τον άνθρωπο, οι απαιτήσεις της σύγχρονης εποχής στην επιτέλεση συγκεκριμένων, περίπλοκων και, αρκετά συχνά, χρονοβόρων καθηκόντων καθιστά την φυσιολογία του ανθρώπου παρωχημένη. Οι προαναφερθέντες αποτέλεσαν σημαντικοί λόγοι για να μηχανευτούν προσεγγίσεις και μέθοδοι σύμφωνα με τις οποίες θα επιτάσσονταν υπολογιστικά συστήματα που θα μπορούσαν πρωτίστως να μετατρέψουν την αναλογική μορφή της εικόνας σε ψηφιακή και στην συνέχεια να επεξεργαστούν την οπτική πληροφορία που περιέχεται σε αυτήν εφαρμόζοντας συγκεκριμένες διαδικασίες ώστε να εξάγουν ένα είδος χρήσιμης πληροφορίας ή κρυφά μοτίβα που το ανθρώπινο μάτι δεν θα μπορούσε να «συλλάβει». Αυτό ακριβώς πραγματεύεται ο τομέας της Ψηφιακής Επεξεργασίας Εικόνας (ΨΕΕ).

1.1.1 Θεμελιώδεις λίθοι στην Ψηφιακή Επεξεργασία Εικόνας

Για να γίνει αποτελεσματικά αντιληπτή η έννοια της Ψηφιακής Επεξεργασίας Εικόνας, θα πρέπει κανείς να εντρυφήσει στον τρόπο με τον οποίο οι υπολογιστικές μηχανές αντιλαμβάνονται την οπτική πληροφορία που τους παρέχεται. Υπάρχει μια αχανής έκταση βιβλιογραφικών πηγών που αναδεικνύουν ότι ***η εικόνα μπορεί να γίνει αντιληπτή ως μια δισδιάσταση συνάρτηση, $f(x, y)$, οι μεταβλητές της οποίας αποτελούν τις χωρικές συντεταγμένες της εικόνας, και οι τιμές που λαμβάνονται σε αυτές τις συντεταγμένες συντελούν τις διακεκριμένες τιμές έντασης της φωτεινότητας*** [5], [6], [7], [8], [9], [10]. Όμως πως πραγματικά οδηγείται κανείς σε αυτού του είδους την αναπαράσταση; Πως παράγεται αυτή η οπτική πληροφορία; Ποια είναι τα φυσικά μεγέθη και ποια τα μαθηματικά εργαλεία που επιτάσσονται για να μπορέσει να δημιουργηθεί η ψηφιακή μορφή της εικόνας;

Οι απαντήσεις στα προηγούμενα ερωτήματα ξεδιπλώνονται σταδιακά ξεκινώντας από την λήψη ενός στιγμιότυπου. Σε μια προσπάθεια μίμησης της λειτουργίας των ανθρώπινων αισθητηρίων οργάνων της όρασης, μια μεγάλη ποικιλία ψηφιακών μέσων έχουν κατασκευαστεί ανά καιρούς με το πιο κοινότυπο και ευρέως διαθέσιμο να είναι μια φωτογραφική μηχανή. Από επαγγελματικούς φακούς που μπορούν να αποτυπώσουν υψηλά επίπεδα λεπτομέρειας, μέχρι και τους πιο εμπορικούς που εντοπίζονται στις φορητές κινητές συσκευές, οι συστοιχίες των αισθητήριων οργάνων των καμερών «συλλαμβάνουν» την οπτική πληροφορία μιας σκηνής με την μορφή κυματομορφής συνεχόμενης

τάσης [11], [12], [13]. Η μετατροπή αυτού του συνεχούς μεγέθους σε διακριτό επιτελείται αξιοποιώντας δυο βασικές διαδικασίες που αποτελούν τον θεμέλιο λίθο της Ψηφιακής Επεξεργασίας Σήματος.

Η πρώτη είναι η Δειγματοληψία μέσω της οποίας ψηφιοποιούνται, δηλαδή αποκτούν διακριτές και πεπερασμένες τιμές, οι χωρικές συντεταγμένες της εικόνας. Έχοντας ως αναφορά grayscale εικόνες, η διαδικασία ξεκινά με την επιλογή ενός ευθύγραμμου τμήματος που εκτείνεται στο μήκος της εικόνας και συνεχίζει με την επιλογή ενός πεπερασμένου αριθμού σημείων τα οποία ισαπέχουν το ένα από το άλλο, αποδίδοντας με αυτόν τον τρόπο τόσο την έννοια της συχνότητας δειγματοληψίας στο σήμα όσο και την έννοια της ομοιομορφίας, οριοθετούμενη από το θεώρημα Nyquist:

Ορισμός 1: Θεώρημα Nyquist (ή Θεώρημα Δειγματοληψίας) [5],[6]

Ένα σήμα πεπερασμένου εύρους ζώνης με ισχύ μέχρι την συχνότητα f_{max} , μπορεί να ανακατασκευαστεί με μηδενικό σφάλμα από δείγματά του, αρκεί αυτά να έχουν ληφθεί ομοιόμορφα, με συχνότητα δειγματοληψίας:

$$f_s \geq 2f_{max}$$

Η διαδικασία ολοκληρώνεται εφαρμόζοντας μια επαναληπτική μέθοδο η οποία αξιοποιεί τα δυο προηγούμενα βήματα διατρέχοντας όλο το ύψος της υπο εξέταση εικόνας¹. Ως ένδειξη καλής πρακτικής, μπορεί να μελετηθεί το παρακάτω παράδειγμα. Έστω η *Εικόνα 1A*) που υποδεικνύει την αρχική εικόνα προς επεξεργασία. Το πρώτο βήμα είναι να μετατραπεί η έγχρωμη εικόνα σε grayscale, ενώ στην συνέχεια επιλέγεται ένα αυθαίρετο ευθύγραμμο τμήμα κατά μήκος του οποίου θα εξαχθούν οι τιμές έντασης της φωτεινότητας.

¹ Στο σημείο αυτό είθισται να αναφερθεί πως μια από τις μορφές αναπαράστασης των εικόνων είναι με τρισδιάστατες δομές πινάκων, όπου στην πρώτη διάσταση εντοπίζεται το κανάλι που περιλαμβάνει την χρωματική πληροφορία του κόκκινου, στην δεύτερη διάσταση εντοπίζεται το κανάλι που περιλαμβάνει την χρωματική πληροφορία του πράσινου και στην τρίτη διάσταση εντοπίζεται το κανάλι που περιλαμβάνει την χρωματική πληροφορία του μπλε. Έτσι, το μήκος και το ύψος αποτελούν τις δισδιάστατες αναπαραστάσεις της εικόνας ενώ το πλάτος δίνει την έννοια του βάθους στην τρισδιάστατη δομή.



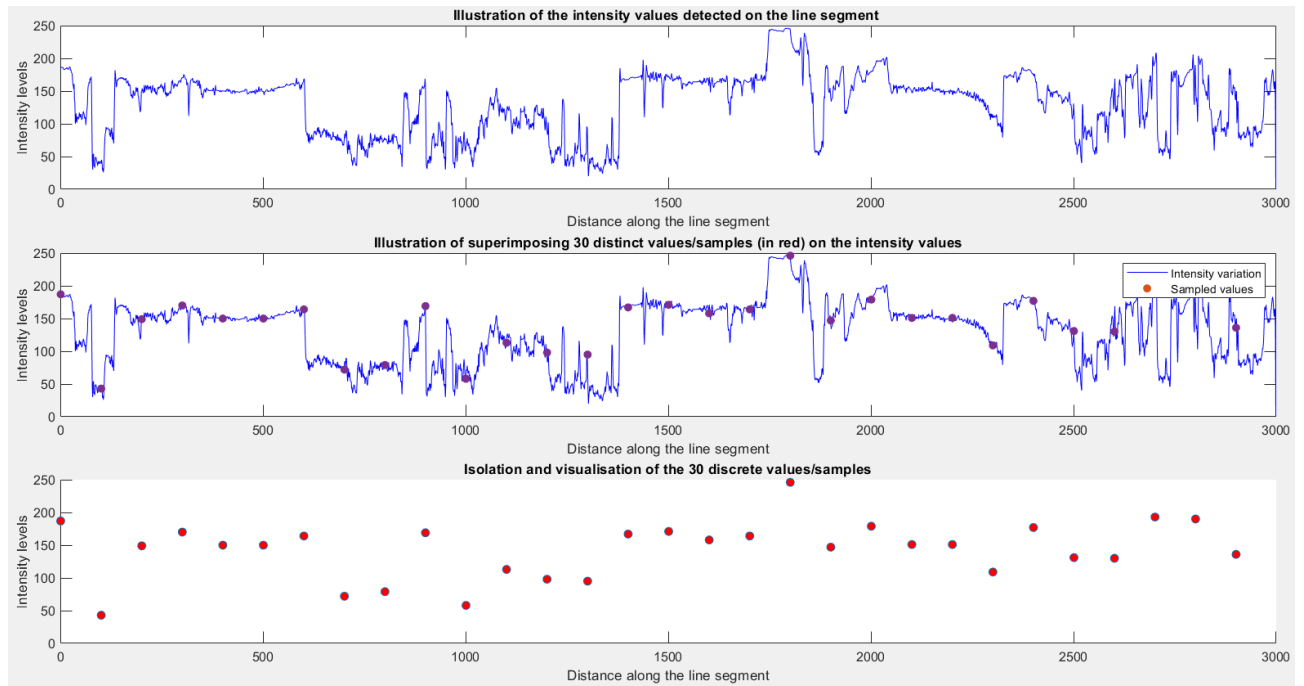
Εικόνα 1: Από αριστερά προς τα δεξιά:

A) Αρχική εικόνα προς επεξεργασία.

B) Grayscale αναπαράσταση της αρχικής εικόνας.

Γ) Υπόδειξη ενός αυθαίρετου ευθύγραμμου τμήματος κατά μήκος του οποίου εξάγονται οι τιμές έντασης φωτεινότητας.

Αξιοποιώντας το προγραμματιστικό περιβάλλον επεξεργασίας εικόνων του MATLAB, στην Εικόνα 2A) μπορεί κανείς να εντοπίσει την διακύμανση των συνεχόμενων τιμών φωτεινότητας που εμφανίζονται στο υπό εξέταση ευθύγραμμο τμήμα. Επιλέγοντας ένα αυθαίρετο σύνολο 30 διακεκριμένων τιμών που απέχουν ίσα και σταθερά βήματα, με κόκκινες τελείες στην Εικόνα 2B) εντοπίζεται η υπέρθεση των προαναφερθέντων τιμών πάνω στο αναλογικό σήμα του ευθύγραμμου τμήματος. Με γνώμονα την αρτιότερη απεικόνιση, στην Εικόνα 2Γ) παρατίθενται οι δειγματοληπτούμενες τιμές έντασης φωτεινότητας απομονωμένες από το αναλογικό σήμα.



Εικόνα 2: Από πάνω προς τα κάτω

A) Απεικόνιση των τιμών έντασης που εντοπίστηκαν στο ευθύγραμμο τμήμα που ορίζεται στην Εικόνα 1A)

B) Υπέρθεση 30 διακεκριμένων τιμών/δειγμάτων (με κόκκινο χρώμα) πάνω στις τιμές έντασεων

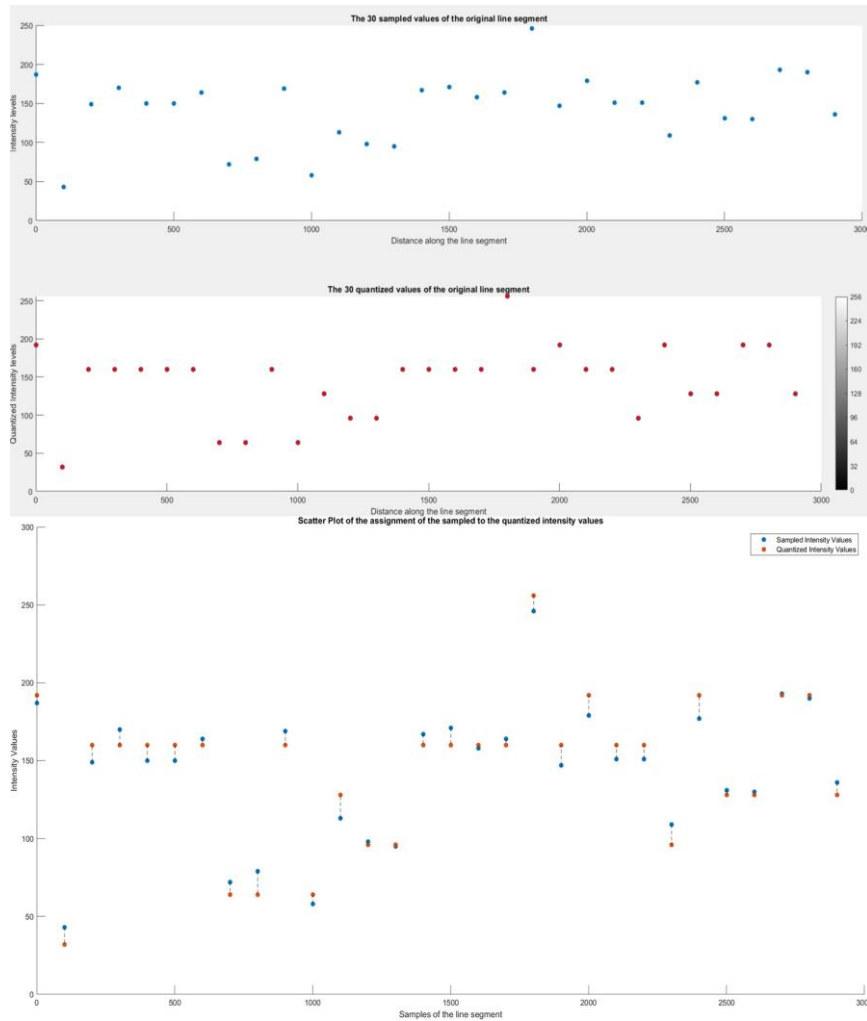
Γ) Απομόνωση και απεικόνιση των 30 διακεκριμένων τιμών/δειγμάτων.

Μέχρι στιγμής, επιτεύχθηκε η απόδοση διακριτών τιμών στις χωρικές συντεταγμένες. Όμως μια εικόνα συνίσταται κι από την αναλογική μορφή των τιμών έντασης της φωτεινότητας που είθισται να λάβει, κάτι που εντοπίζεται εμμέσως πλην σαφώς από τον άξονα y των γραφημάτων της Εικόνας 2A) και Εικόνας 2B), καθιστώντας επιτακτική την απόδοση διακριτών τιμών και σε αυτό το φυσικό

μέγεθος. Η διαδικασία που αναλαμβάνει να φέρει εις πέρας αυτό το ζήτημα ονομάζεται *κβαντισμός*, και συντελείται από δυο βήματα:

1. Το πρώτο βήμα κεντρίζεται στην επιλογή των επιπέδων φωτεινότητας στα οποία θα κβαντιστεί το αναλογικό σήμα, αξιοποιώντας την γνώση ότι τα επίπεδα φωτεινότητας σε μια grayscale εικόνα λαμβάνουν διακεκριμένες τιμές στο πεπερασμένο εύρος $[0,255]$.
2. Έχοντας διαθέσιμα τα δειγματοληπτούμενα δεδομένα, το επόμενο βήμα είναι να αντιστοιχηθούν στα επιλεγμένα επίπεδα φωτεινότητας του βήματος 1. Ένα μέτρο που μπορεί να αξιοποιηθεί είναι η χρήση της Ευκλείδειας απόστασης, σύμφωνα με την οποία υπολογίζεται η κοντινότερη απόσταση της υπό μελέτης δειγματοληπτούμενης τιμής φωτεινότητας από τις διαθέσιμες διακεκριμένες τιμές φωτεινότητας του πεπερασμένου εύρους $[0,255]$

Έχοντας ως βάση το προηγούμενο παράδειγμα και χρησιμοποιώντας οκτώ (8) διακριτά επίπεδα φωτεινότητας, η διαδικασία του κβαντισμού μπορεί να αποδοθεί σχηματικά μέσω των αποκομμάτων της *Εικόνας 3*. Στην *Εικόνα 3A)* απεικονίζονται ξανά οι δειγματοληπτούμενες τιμές του υπό εξέταση ευθύγραμμου τμήματος της *Εικόνας 1Γ)* ενώ στην *Εικόνα 3B)* παρατίθενται οι κβαντισμένες τιμές αυτών. Η χρωματική παλέτα, που εντοπίζεται στα δεξιά του γραφήματος, λειτουργεί ως ένας προσεγγιστικός οδηγός στην ανάθεση των δειγματοληπτούμενων τιμών. Τέλος στην *Εικόνα 3Γ)* αποτυπώνεται με παραστατικό τρόπο η διαδικασία της κβάντισης, όπου με μπλε αναδεικνύονται οι δειγματοληπτούμενες τιμές, με κόκκινο οι κβαντισμένες τιμές στις οποίες ανατίθενται ενώ οι διακεκομμένες γραμμές κάνουν αισθητή την έννοια της Ευκλείδειας απόστασης ως μέτρο περάτωσης της διαδικασίας της κβάντισης.



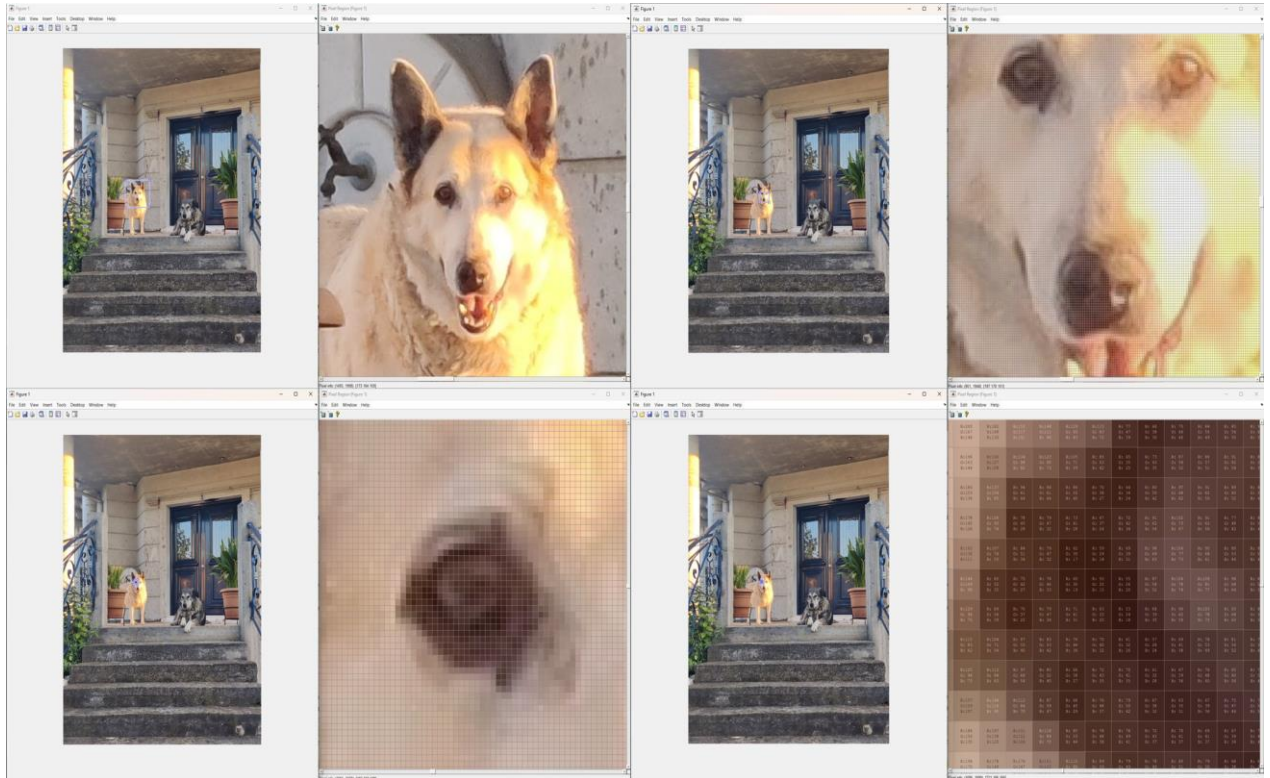
Εικόνα 3: Από πάνω προς τα κάτω

A) Απεικόνιση των δειγματοληπτούμενων τιμών από το ευθύγραμμο τμήμα της Εικόνας 1Γ).

B) Αποτύπωση των κβαντισμένων δειγματοληπτούμενων τιμών.

Γ) Αποτύπωση της διαδικασίας της ανάθεσης των δειγματοληπτούμενων τιμών σε διακεκριμένες τιμές.

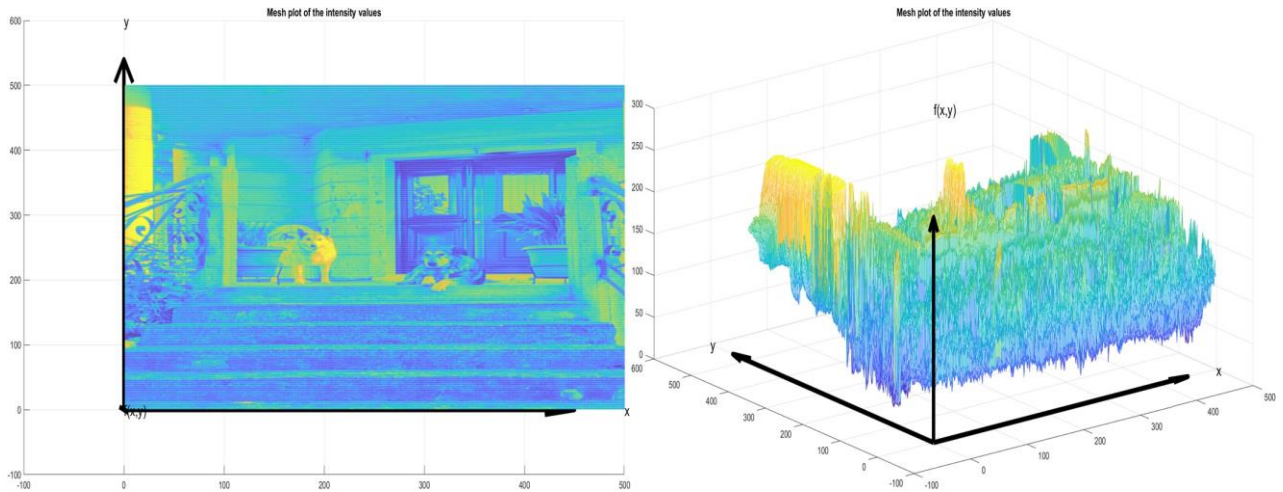
Ο συγκερασμός της παραγωγής διακριτών και ακέραιων τιμών της χωρικής πληροφορίας που προκύπτει από τη Δειγματοληψία με την αντίστοιχη απόδοση διακριτών ακέραιων τιμών επιπέδων φωτεινότητας σε όλα τα πιθανά ζευγάρια συντεταγμένων που προκύπτουν από τον Κβαντισμό, «φέρνουν στην επιφάνεια» ένα νέο ποσοτικό μέγεθος, το οποίο είναι κοινώς γνωστό ως εικονοστοιχείο (ή *pixel*). Χωρίς να αποκλίνει από την πραγματικότητα, κανείς μπορεί να φανταστεί τα στοιχεία αυτά ως τετράγωνα προκαθορισμένου μήκους και πλάτους που επικαλυπτόμενα πάνω στην εικόνα δημιουργούν ένα πλέγμα τιμών, το οποίο αθροιστικά αναπαράγει την εκάστοτε οπτική πληροφορία, όπως μπορεί να φανεί κι από την *Εικόνα 4*.



Εικόνα 4: Από πάνω προς τα κάτω κι από αριστερά προς τα δεξιά

- A) Αποτύπωση της αρχικής εικόνας με το παράθυρο επεξεργασίας πληροφοριών των εικονοστοιχείων.
- B) Αρχική σμίκρυνση του παραθύρου επεξεργασίας πληροφοριών.
- Γ) Περαιτέρω σμίκρυνση του παραθύρου επεξεργασίας πληροφοριών.
- Δ) Τελική σμίκρυνση του παραθύρου επεξεργασίας πληροφοριών με την αποτύπωση των τιμών της RGB τριπλέτας.

Με το εικονοστοιχείο να θεωρείται η θεμελιώδης μονάδα μέτρησης, δεικτοδότησης και επεξεργασίας της φωτεινότητας σε συγκεκριμένες τοποθεσίες της εικόνας, η αρχική παραδοχή πως η εικόνα μπορεί να ιδωθεί ως μια δισδιάστατη συνάρτηση ανάθεσης ακέραιων τιμών έντασης στις εκάστοτε χωρικές συντεταγμένες είναι πλέον αδιαμφισβήτητη οδηγώντας μάλιστα και στην κλιμάκωση της ψηφιοποίησης μέχρι και τα τρία επίπεδα της εικόνας (δηλαδή τις RGB εικόνες) που εισάγουν την έννοια της χρωματικής πληροφορίας. Προκειμένου να γίνουν περαιτέρω αντιληπτές οι παραπάνω έννοιες, ο αναγνώστης μπορεί να επεξεργαστεί την *Εικόνα 5*. Στην *Εικόνα 5A*) εντοπίζεται η δισδιάστατη αναπαράσταση της grayscale *Εικόνας 1A*) με εμφανείς τους άξονες x και y ενώ στην *Εικόνα 5B*) εντοπίζεται η τρισδιάστατη απεικόνιση πλέγματος των τιμών φωτεινότητας ως συνάρτηση των χωρικών συντεταγμένων.



Εικόνα 5: Από αριστερά προς τα δεξιά

- A) Απεικόνιση των grayscale τιμών έντασης της φωτεινότητας.
 B) Απεικόνιση με mesh plot των τιμών έντασης.

Τα παραπάνω αποτελούν τους πυλώνες γύρω από τους οποίους έχει δομηθεί η επιστήμη της Ψηφιακής Επεξεργασίας Εικόνας. Κι ενώ δεν υπάρχει κάποια από κοινού συμφωνία για το τι όντως τη συνιστά, για να μπορέσει να εφαρμοστεί μια αποτελεσματική και εννοιολογική διαφορά ανάμεσα στην προαναφερθείσα από την Υπολογιστική Όραση, κανείς θα μπορούσε να φανταστεί μια πυραμιδική οικοδόμηση τριών ξεχωριστών διαδικασιών. Στην βάση της πυραμίδας τοποθετούνται διαδικασίες χαμηλού επιπέδου, οι οποίες χαρακτηρίζονται από το γεγονός ότι τόσο το σήμα εισόδου όσο και το σήμα εξόδου είναι εικόνες. Στον κορμό της πυραμίδας εντοπίζονται διαδικασίες μεσαίου επιπέδου, σύμφωνα με τις οποίες τα σήματα εισόδου είναι γενικά εικόνες, αλλά οι έξοδοι είναι σήματα που μεταφέρουν πληροφορία σχετικά με κάποιες ιδιότητες που εξάγονται από αυτές, όπως είναι για παράδειγμα ακμές. Τέλος στην κορυφή της πυραμίδας εντοπίζονται διαδικασίες υψηλού επιπέδου, όπου αποτελεί και το επίπεδο της απόφασης, το επίπεδο της αναγνώρισης και γενικά το επίπεδο στο οποίο επιτελούνται οι νοητικές λειτουργίες που σχετίζονται με την εξαγωγή νοήματος από το οπτικό περιεχόμενο της σκηνής [5].



Εικόνα 6: Απεικόνιση της πυραμιδικής δομής των διαδικασιών της Ψηφιακής Επεξεργασίας Εικόνας².

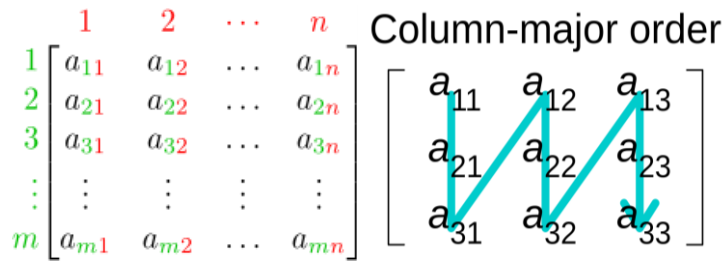
² Η εικόνα έχει ληφθεί από τον [σύνδεσμο](#) και έχει επεξεργαστεί για να αποτυπώσει το περιεχόμενο προς εξέταση.

1.1.2 Αναπαράσταση της Ψηφιακής Εικόνας

Η σημασία των τιμών των εικονοστοιχείων, όπως αυτά ορίστηκαν στην παράγραφο 1.1.1, εκτείνονται σε μεγαλύτερο βάθος από την απλή παράθεση των χωρικών συντεταγμένων της ψηφιακής εικόνας. Η πλεγματική δομή που συντίθεται από την υπέρθεση των τιμών έντασης φωτεινότητας στην χωρική πληροφορία σχηματίζει μια επιπρόσθετη μορφή αναπαράστασης³ την αναπαράσταση με μορφή πίνακα. Για να τεθεί ένα πιο γόνιμο έδαφος με στόχο την αρτιότερη κατανόηση του τελευταίου σχολίου, κανείς θα μπορούσε να καταφύγει στις βασικές αρχές της Γραμμικής Άλγεβρας όπου μπορεί να εντοπίσει ότι ένας πίνακας αποτελεί μια συστοιχία δεδομένων με συχνά τετραγωνική μορφή, κατανεμημένα σε γραμμές και στήλες [16]. Το αξίωμα αυτό εμφανίζει άμεση ταύτιση με την απόδοση της διακεκριμένης τοποθεσίας των εικονοστοιχείων από την διαδικασία της ψηφιοποίησης μιας αναλογικής εικόνας, επιβεβαιώνοντας πως στον τομέα της υπολογιστικής μηχανικής, μια εικόνα μπορεί και είθισται να αναπαρίσταται ως ένας πίνακας αυθαίρετου μεγέθους διακριτών τιμών φωτεινότητας.

Τα μόνα κομμάτια που απομένουν για να συμπληρωθεί το παζλ, είναι η υιοθέτηση ενός τρόπου αρίθμησης των εικονοστοιχείων, η επιλογή ενός σημείου έναρξης καθώς και μια μέθοδος δεικτοδότησης αυτών³. Η παγκόσμια βιβλιογραφία αποτυπώνει πως η αρίθμηση των εικονοστοιχείων είναι μια διαδικασία που εξαρτάται σε μεγάλο βαθμό από το πρόγραμμα επεξεργασίας που επιλέγει ο κάθε χρήστης [5], με χαρακτηριστικό παράδειγμα την διαφορά στην έναρξη δεικτών του προγραμματιστικού περιβάλλοντος της Python όπου το 0 θεωρείται η έναρξη της αρίθμησης ενώ το αντίστοιχο περιβάλλον του MATLAB υιοθετεί το 1 [17], [18]. Κάτι όμως το οποίο είναι από κοινού αποδεκτό, είναι ότι η έναρξη της αρίθμησης είθισται και θα πρέπει να γίνεται από το αριστερότερο και πάνω σημείο ενός πίνακα προσομοιώνοντας την σύμβαση που πολλές απεικονιστικές οθόνες χρησιμοποιούσαν για να σαρώσουν μια εικόνα [19]. Κλείνοντας, και με γνώμονα το τελευταίο σχόλιο περί της σύμβασης των απεικονιστικών οθονών, γνωστοποιείται ότι η παρούσα διπλωματική εργασία, σε όλο το σύνολό της, εναντιώνεται στον κοινότυπο τρόπο δεικτοδότησης και σάρωσης των στοιχείων των εικόνων ή και των δεδομένων γενικότερα. Συγκεκριμένα, προκειμένου να μπορέσει να χαλιναγωγηθεί στο έπακρον η υπολογιστική δυνατότητα που έχει να προσφέρει το MATLAB, η μέθοδος δεικτοδότησης επιλέγεται να είναι βασισμένη στις στήλες (*column-wise linear matrix indexing*) προσφέροντας άμεση συμβατότητα με τον προεπιλεγμένο τρόπο δεικτοδότησης δεδομένων του προγραμματιστικού περιβάλλοντος του αναφερόμενου [20].

³ Παρ' όλο που η αρίθμηση και η δεικτοδότηση αποτελούν δυο τελείως διαφορετικές διαδικασίες, στο πλαίσιο της παρούσας ανάλυσης θα μπορούσε να γίνει μια άτυπη συμφωνία απόδοσης του ίδιου νοήματος και στις δυο λέξεις ώστε να δοθεί ένα αρχικό σύστημα που ωφελεί και τις δυο προαναφερθείσες σημειογραφίες.



Εικόνα 7: Από αριστερά προς τα δεξιά

A) Απεικόνιση της συμβατικής μορφής αρίθμησης ενός πίνακα.⁴

B) Απεικόνιση της δεικτοδότησης με βάση τις στήλες ενός πίνακα.⁵

1.2 Βασικές σχέσεις εικονοστοιχείων

Η αναπαράσταση της εικόνας με την μορφή πίνακα αποτελεί ένα κρίσιμο σημείο παρατήρησης καθώς παρέχει την δυνατότητα επεξεργασίας των εικονοστοιχείων, στο σύνολό τους, με την εφαρμογή σε αυτά μαθηματικών πράξεων που εξάγουν χρήσιμες πληροφορίες. Για να επιτευχθεί αποτελεσματική εξαγωγή όμως είναι επιτακτικό να είναι γνωστή, ή τουλάχιστον να μπορεί να βρεθεί, η ενδεχόμενη εγγενής σχέση που συνδέει τα pixel μεταξύ τους. Το παρών κομμάτι επικεντρώνεται στην παράθεση της έννοιας του «γείτονα» συνεχίζει με τον ορισμό της *γεινίασης* και κλιμακώνεται στην έννοια της *συνδεσιμότητας*.

1.2.1 Η έννοια του «Γείτονα» στα εικονοστοιχεία

Φτάνοντας στο παρών σημείο, η τετραγωνική μορφή των εικονοστοιχείων αποτελεί μια σφαιρικός εμπειρισταωμένη έννοια. Η μοναδική και συμπαγής δομή απόδοσης πληροφοριών της έντασης της φωτεινότητας που παρέχουν τα pixel κρύβει επιπρόσθετα προτερήματα που οφείλουν περαιτέρω μελέτης, ένα εκ των οποίων είναι η επεξεργασία της «επαφής» που εμφανίζει με τα γειτονικά του pixel. Έχοντας ως βήμα την *Εικόνα 8*, εντοπίζεται το pixel της grayscale *Εικόνας 1* στις συντεταγμένες $(x,y) = (1899,1098)$ και ένταση φωτεινότητας 63. Κάνοντας χρήση των βασικών ομοκεντρικών κατευθύνσεων (*egocentric coordinates*)⁶ [21], διαφαίνεται ότι το υπό εξέταση εικονοστοιχείο εμφανίζει τέσσερις «γείτονες», δηλαδή τέσσερα διαφορετικά στοιχεία με τα οποία επικοινωνεί σε κάθε μια από τις πλευρές του, σε αυτές τις κατευθύνσεις. Γενικεύοντας και εισάγοντας την μαθηματική σημειογραφία, μπορεί να υποστηριχτεί ότι:

⁴ Η αριστερή εικόνα μπορεί να εντοπιστεί στον ακόλουθο [σύνδεσμο](#).

⁵ Η δεξιά εικόνα είναι επεξεργασμένη. Η πρωτότυπη μπορεί να εντοπιστεί στον ακόλουθο [σύνδεσμο](#).

⁶ Χωρίς παράβαση της γενικότητας, οι ομοκεντρικές κατευθύνσεις περιλαμβάνουν την τριδιάστατη κίνηση στο χώρο. Στο παρών πλαίσιο, ο όρος αναφέρεται στις τυπικές κινήσεις που μιμούνται στο δισδιάστατο επίπεδο τις κινήσεις προς τα επάνω, προς τα κάτω προς τα δεξιά και προς τα αριστερά ($\uparrow \downarrow \rightarrow \leftarrow$).

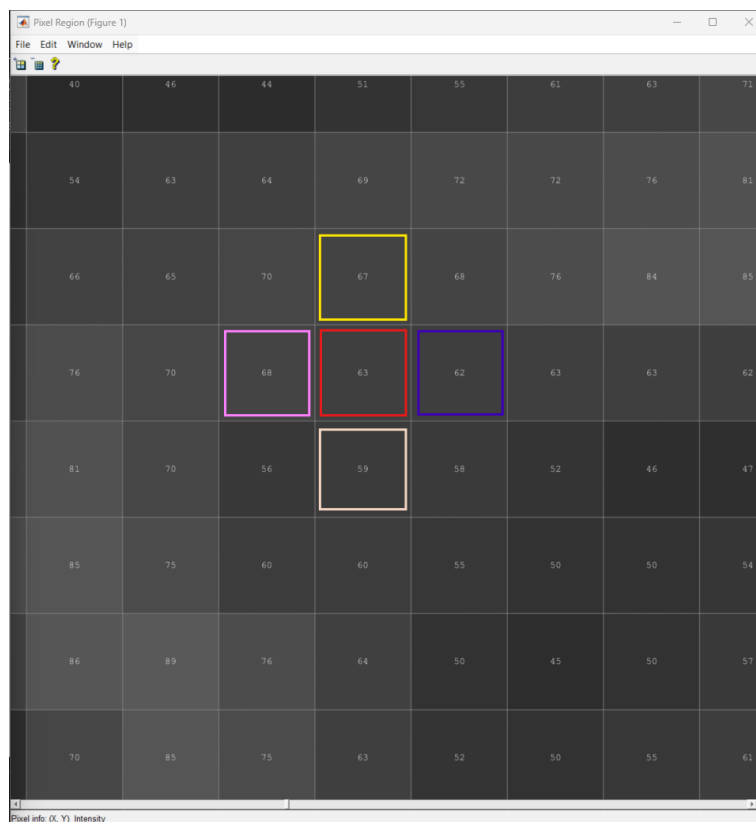
Ορισμός 2: 4-Γείτονες

Ένα εικονοστοιχείο, έστω A , με χωρικές συντεταγμένες (x,y) εμφανίζει δυο οριζόντιους και δυο κάθετους γείτονες, η τοποθεσία των οποίων δίνεται από την παρακάτω σειρά ομοκεντρικών συντεταγμένων [5]:

$$(x, y + 1), (x, y - 1), (x + 1, y), (x - 1, y)$$

Το σύνολο αυτών των τεσσάρων εικονοστοιχείων ονομάζεται *4-γείτονες* του A και η σημειογραφία που το συνοδεύει είναι:

$$N_4(A)$$



Εικόνα 8: Απεικόνιση της έννοιας των τεσσάρων γειτόνων.
 Με κόκκινο υποδεικνύεται το εικονοστοιχείο προς εξέταση
 ενώ τα υπόλοιπα χρώματα υποδηλώνουν τους τέσσερις γείτονες.

Η έννοια του γείτονα μπορεί να επεκταθεί σε περισσότερα από τέσσερα γειτονικά εικονοστοιχεία. Εξετάζοντας ξανά την *Εικόνα 8*, εντοπίζεται ότι το υπό εξέταση pixel «επικοινωνεί» με τέσσερα ακόμη στοιχεία στις κάθετες γωνίες του σε διαγώνιες τοποθεσίες συναρτήσεως του σημείου αναφοράς, όπως διαφαίνονται με άσπρο χρώμα στην *Εικόνα 9*. Τα προηγούμενα προσφέρουν προσοδοφόρο έδαφος ώστε ο *Ορισμός 2* να ανανεωθεί ως:

Ορισμός 3: 8-Γείτονες

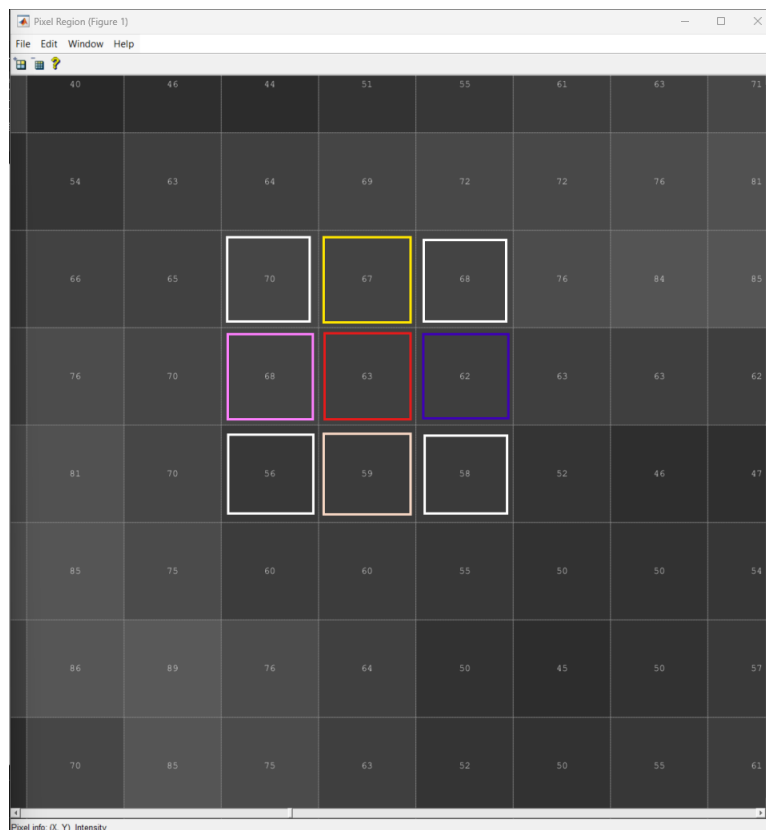
Ένα εικονοστοιχείο, έστω A , με χωρικές συντεταγμένες (x,y) εμφανίζει δυο οριζόντιους δυο κάθετους και δυο διαγώνιους γείτονες, με την τοποθεσία των τελευταίων να δίνεται από την παρακάτω σειρά ομοκεντρικών συντεταγμένων [5]:

$$(x + 1, y), (x - 1, y), (x, y + 1), (x, y - 1),$$

$$(x + 1, y + 1), (x + 1, y - 1), (x - 1, y + 1), (x - 1, y - 1)$$

Το σύνολο αυτών των οκτώ εικονοστοιχείων ονομάζεται δ -γείτονες του A και η σημειογραφία που το συνοδεύει είναι:

$$N_8(A)$$



Εικόνα 9: Απεικόνιση της έννοιας των οκτώ γειτόνων.
Με λευκό χρώμα απεικονίζονται οι συμπληρωματικοί τέσσερις γείτονες.

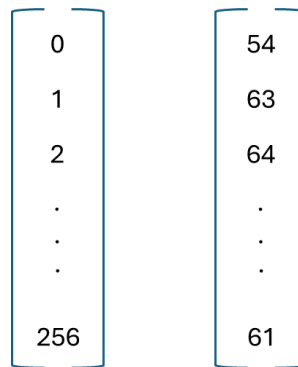
Οι 4 και 8-γείτονες αντίστοιχα αποτελούν τους θεμέλιους λίθους σύνδεσης των εικονοστοιχείων μεταξύ τους, αλλά μπορούν να επεκταθούν και με την έννοια της m -γειτνίασης, η οποία μπορεί να ιδωθεί ως μια τροποποίηση των 8-γειτόνων ώστε να αρθούν ασάφειες που μπορεί να εμφανιστούν ανάμεσα σε διαφορετικές περιοχές εικονοστοιχείων στα οποία ζητείται να βρεθεί μια συσχέτιση, αυξάνοντας έτσι τις πληροφορίες που μπορούν να εξαχθούν για το υπό εξέταση σημείο. Κλείνοντας,

αξίζει να ειπωθεί ότι το σύνολο των τοποθεσιών των γειτόνων ενός σημείου A , ονομάζεται γειτονιά του A . Η γειτονιά αυτή είναι κλειστή όταν περιλαμβάνει το A και ανοιχτή όταν δεν το περιλαμβάνει [5].

1.2.2 Η έννοια της Γειννίαςης στα εικονοστοιχεία

Θέτοντας υπό περαιτέρω επεξεργασία το κομμάτι της γειτονιάς των εικονοστοιχείων με ταυτόχρονη αλλαγή σε μακροσκοπική οπτική, παρατηρείται ότι μια περιοχή εικονοστοιχείων αποτελείται από ένα σύνολο διακεκριμένων ακέραιων τιμών έντασης φωτεινότητας. Στόχος είναι να μπορέσει να εντοπιστεί κάποιου είδους επιπρόσθετης συνδετικότητας ανάμεσα στα pixel τα οποία συντελούν την προκειμένη περιοχή αλλά κατ' επέκτασιν και όλη την εικόνα. Το τελευταίο μπορεί να επιτευχθεί από την αξιοποίηση του ορισμού της *γειννίαςης*.

Στρέφοντας το ενδιαφέρον στην *Εικόνα 8*, εντοπίζεται ένα πλήθος τιμών φωτεινότητας σε όλο το μήκος και το πλάτος της απεικόνισης⁷, με τιμές που εμφανίζονται τυπικά στο γενικότερο εύρος τιμών μιας grayscale εικόνας, δηλαδή [0,255]. Όπως όμως διαπιστώνεται και μέσω της χρήσης των λέξεων, δεν εμφανίζονται και οι 256 τιμές έντασης της φωτεινότητας υποχρεωτικά σε μια εξεταζόμενη περιοχή αλλά μόνο ένα συγκεκριμένο υποσύνολο τιμών για την αποτύπωση της *γειννίαςης* των εξεταζόμενων εικονοστοιχείων. Έτσι λοιπόν η γειννίαςη επηρεάζεται τόσο από την εξεταζόμενη περιοχή όσο κι από το υποσύνολο⁸ των τιμών φωτεινότητας που εμφανίζονται σε αυτή [5], [22], [23], [24].



Εικόνα 10: Από αριστερά προς τα δεξιά
Α) Αποτύπωση του εύρους των δυνατών τιμών έντασης φωτεινότητας μιας grayscale εικόνας.
Β) Αποτύπωση του υποσυνόλου των τιμών έντασης που ορίζει την γειννίαςη στις Εικόνα 8 και 9.

Μέχρι το παρών σημείο έχει οριστεί το σύνολο των τιμών που αντιπροσωπεύουν και ορίζουν την γειννίαςη. Αυτό που απομένει είναι να εμπεριστατωθεί κι ένα τρόπος με τον οποίο να μπορεί να εκφραστεί η σχέση ανάμεσα στο εξεταζόμενο εικονοστοιχείο με τα *γειτονικά* του. Η χρήση της λέξης

⁷ Εφιστάται προσοχή στο γεγονός ότι δεν σχολιάζεται η χρωματική περιοχή των τεσσάρων γειτόνων μόνο όπως έγινε σε ανώτερο επίπεδο, αλλά η επεξεργασία εκτείνεται σε όλη την απεικόνιση.

⁸ Ως ένδειξη καλής πρακτικής, πρέπει να αναφερθεί ότι αυτό το υποσύνολο μπορεί να περιλαμβάνει έναν αυθαίρετο αριθμό τιμών έντασης, τόσο από την άποψη της αριθμητικής παρουσίας όσο κι από την άποψη των ίδιων των τιμών που λαμβάνουν οι εντάσεις της φωτεινότητας.

«γειτονικά» δεν έγινε τυχαία καθώς η γειτονικότητα είναι αυτή που καταφέρνει να αποδώσει την τελική μορφή της έννοιας της *γεινίασης*. Όπως εξετάστηκε στην *υποενότητα 1.2.1* για την γειτονιά έτσι και εδώ, η γεινίαση μπορεί να εμφανίζει δυο βασικές κατηγορίες:

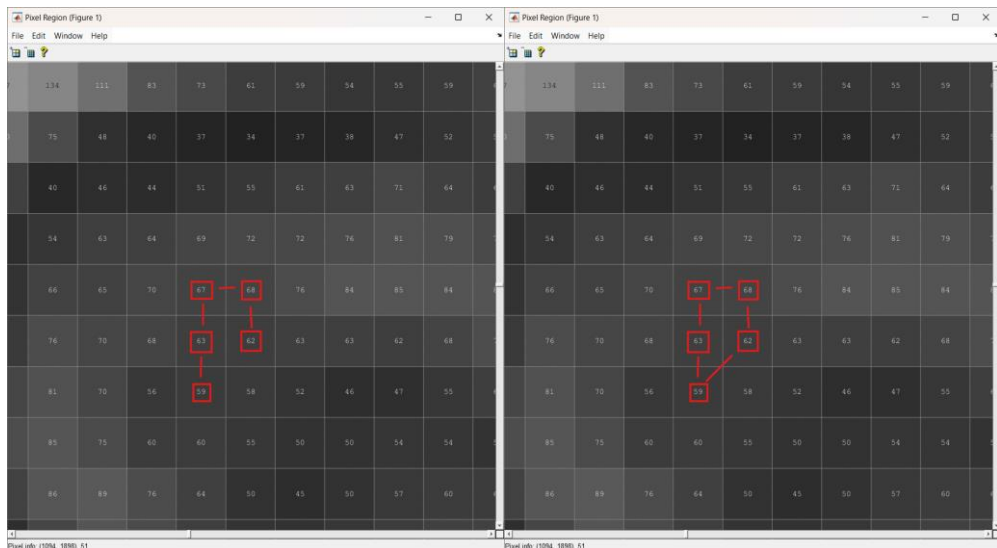
Ορισμός 4: Γεινίαση-κατά-4

Δυο εικονοστοιχεία A και B με τιμές που εκμαιεύονται από ένα σύνολο τιμών V , εμφανίζουν γεινίαση-κατά-4 αν το B ανήκει στο σύνολο $N_4(A)$ [5]

Ορισμός 5: Γεινίαση-κατά-8

Δυο εικονοστοιχεία A και B με τιμές που εκμαιεύονται από ένα σύνολο τιμών V , εμφανίζουν γεινίαση-κατά-8 αν το B ανήκει στο σύνολο $N_8(A)$ [5]

Προφανώς, όπως και στην περίπτωση της γειτονιάς, η γεινίαση μπορεί να επεκταθεί σε έναν αυθαίρετο (q) αριθμό διαστάσεων



Εικόνα 11: Από αριστερά προς τα δεξιά
Απεικόνιση της γεινίασης-κατά-4 και της γεινίασης-κατά-8.

1.2.3 Η έννοια της Συνδεσιμότητας στα εικονοστοιχεία

Τόσο η *γειτονιά* όσο και η *γεινίαση* έθεσαν μέχρι στιγμής τα θεμέλια για τον εντοπισμό των στοιχείων που περιβάλλουν ένα υπό εξέταση pixel. Απώτερος σκοπός της παρούσας υποενότητας είναι να φωτίσει τις έννοιες εκείνες που ορίζουν την έννοια της συνδεσιμότητας μεταξύ των εικονοστοιχείων μιας προκείμενης περιοχής.

Η έννοια της συνδεσιμότητας των εικονοστοιχείων εμφανίζει τις «ρίζες» της στον χαρακτηρισμό της πορείας που ακολουθούν τα γειτονικά εικονοστοιχεία σε σχέση με το κεντρικό. Τι όμως ορίζεται ως πορεία από το ένα εικονοστοιχείο στο άλλο;

Ορισμός 6: Πορεία

Ως ψηφιακή πορεία από το pixel A στο pixel B ορίζεται η αλληλουχία των διακεκριμένων γειτνιακών εικονοστοιχείων που οδηγούν στην συνένωση των δυο προηγούμενων [5]

Ας επιτραπεί να εξεταστεί η *Εικόνα 11* στην οποία εντοπίζεται η *γεινίαση-κατά-4* και η *γεινίαση-κατά-8* μιας συγκεκριμένης γειτονιάς εικονοστοιχείων, όπου το κέντρο αυτών αποτελεί το pixel με τιμή φωτεινότητας έντασης ίση με 62. Παρατηρείται ότι στην *Εικόνα 11A)* η πορεία που ακολουθείται από το κεντρικό μέχρι το τελικό pixel, εδώ 59, είναι μια ανοιχτή διαδρομή, δηλαδή ένα μονοπάτι το οποίο ξεκινά από το κεντρικό σημείο αναφοράς και δεν ολοκληρώνεται σε αυτό. Αντίστροφα, στην *Εικόνα 11B)* παρατηρείται ότι το μονοπάτι που διαγράφεται, όπως κι αν αυτό ιδωθεί, ξεκινά και ολοκληρώνεται στο κεντρικό εικονοστοιχείο, διαγράφοντας την επονομαζόμενη κλειστή διαδρομή.

Τα προηγούμενα δημιουργούν προσοδοφόρο έδαφος για την έγερση του παρακάτω ορισμού:

Ορισμός 7: Σύνδεση εικονοστοιχείων

Εστω ότι το S αντιπροσωπεύει ένα υποσύνολο εικονοστοιχείων σε μια εικόνα. Δύο εικονοστοιχεία A και B λέγεται ότι συνδέονται στο S εάν υπάρχει μονοπάτι μεταξύ τους που αποτελείται εξ ολοκλήρου από εικονοστοιχεία του S . Για κάθε εικονοστοιχείο p στο S , το σύνολο των εικονοστοιχείων που συνδέονται με αυτό στο S ονομάζεται συνδεδεμένη συνιστώσα του S . [5]

1.3 Βασικές πράξεις στην Επεξεργασία Εικόνας

Η προηγούμενη ενότητα αποτέλεσε το εφαλτήριο για την ανάδειξη των σχέσεων που εμφανίζουν τα εικονοστοιχεία μεταξύ τους με την εισαγωγή των εννοιών της *γειτονιάς*, της *γεινίασης* και της *συνδεσιμότητας*. Τα προηγούμενα, αν και θεμελιώδη, δεν αρκούν για να μπορέσουν να αποδώσουν κάποια αξιοποιήσιμη μορφή πληροφορίας η οποία να αποδίδει νόημα ή να σχηματίζει κάποιο εννοιολογικό πλαίσιο στην εκάστοτε υπό εξέταση εικόνα. Θα πρέπει να επιστρατευτούν εργαλεία τα οποία θα μπορέσουν να αποτυπώσουν αλλά και να εξάγουν αυτήν την ουσιώδη υποβόσκουσα πληροφορία. Στόχος της ενότητας αυτής είναι να αποτυπώσει κάποια από τα βασικά μαθηματικά αξιώματα, η χρήση των οποίων είναι κομβική όχι μόνο για την επεξεργασία εικόνων αλλά και για τον τρόπο με τον οποίο τα ίδια τα υπολογιστικά συστήματα επεξεργάζονται την πληροφορία.

1.3.1 Πράξεις στοιχείο-προς-στοιχείο

Όπως έχει εκτενώς αναφερθεί μέχρι στιγμής, οι ψηφιακές εικόνες μπορούν να γίνουν ορθώς αντιληπτές ως πίνακες αυθαίρετων διαστάσεων. Η Γραμμική Άλγεβρα είναι εκείνος ο κλάδος της επιστήμης των μαθηματικών που παρέχει όλα τα απαραίτητα, αλλά και όχι μόνο, αξιώματα που αναφέρονται στην επεξεργασία και εφαρμογή πράξεων μεταξύ πινάκων, οδηγώντας σε μια αβίαστη και εύλογη συνειδητοποίηση ότι όταν κανείς ασχολείται με μήτρες ή διανύσματα θα πρέπει οι διαδικασίες που επιλέγει να εφαρμόσει να υπάγονται ρητά στους νόμους που πρεσβεύει η

προαναφερθείσα. Πέρα από τις τυπικές πράξεις της πρόσθεσης, αφαίρεσης, πολλαπλασιασμού και διαίρεσης που μπορούν να εφαρμοστούν χωρίς κάποιο ιδιαίτερο πρόβλημα σε όλες τις εικόνες⁹, ίσως μια από τις πιο διαδεδομένες πράξεις είναι αυτές των πράξεων *στοιχείο-προς-στοιχείο* σύμφωνα με τις οποίες το κάθε στοιχείο του ενός πίνακα πολλαπλασιάζεται με το αντίστοιχο στοιχείο στην αντίστοιχη θέση του άλλου πίνακα. Ένα ενδεικτικό παράδειγμα αποτελεί το παρακάτω:

Έστω δυο πίνακες με τα ακόλουθα στοιχεία:

$$\begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{bmatrix} \text{ και } \begin{bmatrix} y_{11} & y_{12} \\ y_{21} & y_{22} \end{bmatrix}$$

Τότε το γινόμενο στοιχείο-προς-στοιχείο των δυο πινάκων δίνεται από την διαδικασία:

$$\begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{bmatrix} \odot \begin{bmatrix} y_{11} & y_{12} \\ y_{21} & y_{22} \end{bmatrix} = \begin{bmatrix} x_{11}y_{11} & x_{12}y_{12} \\ x_{21}y_{21} & x_{22}y_{22} \end{bmatrix}$$

1.3.2 Γραμμικές και μη-Γραμμικές πράξεις

Ένα σημείο εξαιρετικού ενδιαφέροντος είναι η αξιολόγηση της γραμμικότητας, ή μη, μιας πράξης. Έστω ότι υπάρχει ο γενικός τελεστής \mathcal{L} , η εφαρμογή του οποίου σε μια εικόνα $I(x, y)$ δίνει ως αποτέλεσμα μια επεξεργασμένη εικόνα $O(x, y)$:

$$O(x, y) = \mathcal{L}[I(x, y)]$$

Συμπληρωματικά της παραπάνω υπόθεσης, έστω ότι υπάρχουν και δυο σταθερές α και β καθώς και δυο εντελώς διαφορετικές εικόνες $I_1(x, y)$ και $I_2(x, y)$. Τα παραπάνω δίνουν την δυνατότητα σχηματισμού του παρακάτω ορισμού:

Ορισμός 8: Γραμμικοί τελεστές

Ένας τελεστής είναι γραμμικός αν ισχύει η παρακάτω σχέση:

$$\begin{aligned} \mathcal{L}[\alpha I_1(x, y) + \beta I_2(x, y)] &= \alpha \mathcal{L}[I_1(x, y)] + \beta \mathcal{L}[I_2(x, y)] \\ &= \alpha O_1(x, y) + \beta O_2(x, y) \end{aligned}$$

⁹ Μια απαραίτητη διευκρίνιση που πρέπει να γίνει στο προκείμενο σημείο είναι ότι οι πράξεις που αναφέρονται απαιτούν οι πίνακες να έχουν τις ίδιες διαστάσεις.

Γίνεται εμφανές δηλαδή ότι το αποτέλεσμα ενός γραμμικού τελεστή που εφαρμόζεται στο άθροισμα δυο εισόδων είναι το ίδιο με την εφαρμογή του τελεστή ξεχωριστά στις εισόδους με την διαδικασία να ολοκληρώνεται από την εφαρμογή του αθροίσματος στα εκάστοτε στοιχεία [5].

Προκειμένου όμως να επιτευχθεί πληρότητα στην απόδοση των εννοιών, θα ήταν πρέπον να αναφερθεί κι ένα παράδειγμα μη γραμμικότητας. Έστω ότι υπάρχουν δυο εικόνες που ορίζονται από τους τετραγωνικούς πίνακες f_1 και f_2 :

$$I_1 = \begin{bmatrix} 0 & 1 \\ 2 & 3 \end{bmatrix} \quad \text{και} \quad I_2 = \begin{bmatrix} 9 & 3 \\ 2 & 0 \end{bmatrix}$$

Έστω επίσης ξανά δυο τελεστές $\alpha = 1$ και $\beta = -2$. Εφαρμόζοντας την σχέση του Ορισμού 7 και χρησιμοποιώντας την εύρεση της μέγιστης τιμής λαμβάνεται ότι:

$$\max \left\{ (1) \begin{bmatrix} 0 & 1 \\ 2 & 3 \end{bmatrix} + (-2) \begin{bmatrix} 9 & 3 \\ 2 & 0 \end{bmatrix} \right\} = \max \left\{ \begin{bmatrix} -18 & -5 \\ -2 & 3 \end{bmatrix} \right\} = 3$$

Αλλάζοντας τώρα προσέγγιση και εφαρμόζοντας τους τελεστές ξεχωριστά σε κάθε πίνακα, λαμβάνεται ότι:

$$(1) \max \left\{ \begin{bmatrix} 0 & 1 \\ 2 & 3 \end{bmatrix} \right\} + (-2) \max \left\{ \begin{bmatrix} 9 & 3 \\ 2 & 0 \end{bmatrix} \right\} = 3 - 18 = -15$$

Παρατηρείται ότι το αριστερό μέλος της σχέσης στον Ορισμό 7 δεν είναι το ίδιο με το δεξί, αποδεικνύοντας ότι η εύρεση της μέγιστης τιμής δεν είναι μια γραμμική πράξη. Κλιμακώνοντας το προηγούμενο συμπέρασμα μπορεί να υποστηριχθεί ότι όταν δεν ικανοποιείται η σχέση του Ορισμού 7, τότε αυτόματα ο τελεστής προς εφαρμογή είναι μη γραμμικός.

1.3.3 Χωρικές Πράξεις: Γεωμετρικοί Μετασχηματισμοί

Στην Υποενότητα 1.2.1 παρουσιάστηκαν οι βασικές σχέσεις που συνδέουν τα εικονοστοιχεία μεταξύ τους σε μια εικόνα αναφοράς θέτοντας τα θεμέλια για την περαιτέρω επεξεργασία σε μακροσκοπικό επίπεδο. Η σειρά με την οποία παρατίθενται οι πληροφορίες δεν αποτελούν προϊόν τυχαιότητας καθώς θα ήταν αβάσιμο να παρουσιαστούν μέθοδοι επεξεργασίας μιας εικόνας χωρίς πρώτα να έχουν γίνει αναφορές στους δομικά στοιχεία που την αποτελούν. Έχοντας ως γνώμονα το τελευταίο και με την γνώση των προαναφερθέντων, αποτελεί πλέον εύκολο να παρατεθούν κάποιες από τις βασικές χωρικές πράξεις που μπορούν να εφαρμοστούν στις εικόνες. Αν και οι τελευταίες διακρίνονται στις παρακάτω τρεις διακριτές κατηγορίες [5]:

1. Τις πράξεις σε συγκεκριμένα εικονοστοιχεία
2. Τις πράξεις σε γειτονιές
3. Τους χωρικούς γεωμετρικούς μετασχηματισμούς

Η παρούσα υποενότητα υπογραμμίζει μόνο τους χωρικούς γεωμετρικούς μετασχηματισμούς, χωρίς αυτό να μειώνει την σημαντικότητα των προηγούμενων. Οι γεωμετρικοί μετασχηματισμοί είναι μια διαδικασία (ή κι ένα σύνολο διαδικασιών) που έχουν ως στόχο τους την επιβολή σκόπιμης αλλαγής της χωροταξικής θέσης των εικονοστοιχείων μιας εικόνας που παρέχεται ως είσοδος σε ένα σύστημα επεξεργασίας. Για να αποδοθεί πληρότητα στον ορισμό τους όμως χρειάζονται επιπρόσθετα στοιχεία τα οποία μπορούν να εκμαιευτούν από μια απλή πορεία σκέψης. Στην *Υποενότητα 1.1.1* δομήθηκε το βασικό πλαίσιο όπου πραγματευόταν η δημιουργία της ψηφιακής εικόνας ως το συναπάντημα τόσο της χωρικής πληροφορίας όσο και των τιμών της έντασης της φωτεινότητας σε συγκεκριμένες συντεταγμένες, «γεννώντας» άμεσα την έννοια των εικονοστοιχείων. Η ίδια μεθοδολογία υιοθετείται και στο παρών κομμάτι, όπου εντοπίζεται ότι οι γεωμετρικοί μετασχηματισμοί μιας ψηφιακής εικόνας αποτελούνται από δυο βασικές διαδικασίες [5]:

- A. Τον χωροταξικό μετασχηματισμό των συντεταγμένων της εικόνας
- B. Την παρεμβολή (*interpolation*) στην ένταση όπου ανατίθενται τιμές έντασης στα χωρικά μετασχηματιζόμενα εικονοστοιχεία.¹⁰

Στην επεξεργασία εικόνας, το ενδιαφέρον στρέφεται κυρίως στους μετασχηματισμούς της κλιμάκωσης (*scaling*), της μετατόπισης (*translation*), της περιστροφής (*rotation*) και της διάτμησης (*shearing*) οι οποίοι στην διεθνή βιβλιογραφία είναι γνωστοί και ως *affine* μετασχηματισμοί. Ο λόγος για τον οποίο οι *affine* μετασχηματισμοί γνώρισαν τόσο έντονη φήμη είναι λόγω της δυνατότητας που παρέχουν να διατηρούν τα σημεία, τις ευθείες γραμμές καθώς και τα επίπεδα μιας εικόνας αναλλοίωτα στις μεταβολές που επιβάλλονται λόγω αυτών. Και παρ' όλο που το τελευταίο φαίνεται αντιφατικό κανείς θα μπορούσε χονδροειδώς να το αναλογιστεί πως παρόλο που η εικόνα μεταβάλλεται στην μορφολογία της, τα πηγαία χαρακτηριστικά της παραμένουν αναλλοίωτα [5], [25], [26].

Μέχρι το παρών σημείο έχει εμπεριστατωθεί ότι οι *affine* μετασχηματισμοί αποτελούν ένα σύνολο τεσσάρων διακριτών διαδικασιών, οδηγώντας σε μια απόλυτα λογική σκέψη ότι οι μαθηματικές σχέσεις που περιγράφουν τις προαναφερθείσες είναι κι αυτές με την σειρά τους ξεχωριστές η μια από την άλλη. Κι ενώ η προηγούμενη θεώρηση είναι έγκυρη, αυτό το οποίο είναι πραγματικά εντυπωσιακό είναι πως με την χρήση της Γραμμικής Άλγεβρας δίνεται η δυνατότητα το σύστημα


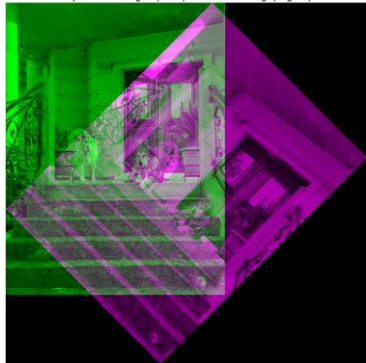
¹⁰ Η έννοια της παρεμβολής στην ένταση των εικονοστοιχείων είναι ένα θέμα το οποίο διαπερνά κατά πολύ τις προοπτικές αυτής της διπλωματικής εργασίας τόσο από την άποψη του περιεχομένου της όσο κι από την άποψη της συγγραφικής έκτασης που θα λάμβανε η εργασία. Αν ο αναγνώστης επιθυμεί να σχηματίσει πλήρη δομημένη άποψη επί του θέματος, μπορεί να καταφύγει στην ενότητα 2.4 της πηγής [5].



των τεσσάρων μετασχηματισμών να αποδίδεται με έναν αποκλειστικό πίνακα διαστάσεων 3×3 ομογενών συντεταγμένων. Ο αναφερόμενος πίνακας είναι της μορφής [5]:

$$\begin{bmatrix} X \\ Y \\ 1 \end{bmatrix} = A \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

Όπου με κεφαλαία έντονα γράμματα υποδηλώνονται οι νέες χωρικές συντεταγμένες των σημείων ενώ με μικρά γράμματα υποδηλώνονται οι αρχικές χωρικές συντεταγμένες των εικονοστοιχείων προς επεξεργασία και ο μετασχηματισμός που επιβάλλεται εξαρτάται από τις τιμές που επιλέγονται στον πίνακα A . Ο παρακάτω πίνακας υποδεικνύει τους χωρικούς μετασχηματισμούς, τις μήτρες που τους εφαρμόζουν, τις εξισώσεις που εφαρμόζονται στα εικονοστοιχεία και το αποτέλεσμα του μετασχηματισμού [5], [25], [26], [27].

Πίνακας 1: Συγκεντρωτικός πίνακας απεικόνισης των affine μετασχηματισμών.

Μετασχηματισμός	Affine Πίνακας (A)	Εξισώσεις συντεταγμένων	Παράδειγμα
Κλιμάκωση (Scaling)	$\begin{bmatrix} s_x & 0 & 0 \\ 0 & s_y & 0 \\ 0 & 0 & 1 \end{bmatrix}$	$X = s_x x$ $Y = s_y y$	<small>Depiction of the original (Green) and the scaled image (Magenta)</small> 
Περιστροφή (Rotation)	$\begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}$	$X = x\cos\theta - y\sin\theta$ $Y = x\sin\theta + y\cos\theta$	<small>Depiction of the original (Green) and the rotated image (Magenta)</small> 

<p>Μετατόπιση (Translation)</p>	$\begin{bmatrix} 1 & 0 & t_x \\ 0 & 1 & t_y \\ 0 & 0 & 1 \end{bmatrix}$	$X = x + t_x$ $Y = y + t_y$	<p style="text-align: center; font-size: small;">Depiction of the original (Left) and the translated image (Right)</p> 
<p>Οριζόντια Διάτμηση (Horizontal Shearing)</p>	$\begin{bmatrix} 1 & 0 & 0 \\ S_h & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$	$X = x$ $Y = S_x x + y$	<p style="text-align: center; font-size: small;">Depiction of the original (Left) and the sheared image (Right)</p> 

1.4 Επεξεργασία Ιστογράμματος

Η μέχρι τώρα πορεία της εργασίας αποτύπωσε τις βασικές έννοιες που κρύβονται πίσω από την ψηφιακή μορφή μιας εικόνας, φωτίζοντας τα στοιχεία που την αποτελούν και υπογράμμισε με έμφαση τα κύρια μαθηματικά εργαλεία που μπορεί κανείς να χρησιμοποιήσει για να επιβάλλει μεταβολές στα εγγενή δεδομένα της εικόνας. Το παρών κομμάτι εστιάζει σε ένα ακόμη κομβικό σημείο που είναι τα ιστογράμματα, μια στατιστική μορφή πληροφορίας η οποία περιγράφει την κατανομή των τιμών έντασης των εικονοστοιχείων μιας ψηφιακής εικόνας. Δεδομένης της υπολογιστικής τους απλότητας και της ευελιξίας που προσφέρουν, τα ιστογράμματα αποτελούν πόλο έλξης για εφαρμογές επεξεργασίας εικόνας σε πραγματικό χρόνο.

1.4.1 Τι είναι Ιστόγραμμα

Πλέον έχει αδιαμφισβήτητα εμπειρισταωθεί ότι τα διακριτά επίπεδα φωτεινότητας που λαμβάνει τυπικά μια *grayscale* εικόνα είναι στο εύρος $[0,255]$ (βλέπε υποενότητα 1.1.1) δίνοντας την δυνατότητα απόδοσης 256 διακριτών επιπέδων του γκρι σε μια ψηφιακή εικόνα. Οι διακριτές αυτές τιμές φωτεινότητας δεν ακολουθούν μια δομημένη σειρά κατά μήκος και πλάτος της απεικόνισης, αλλά εμφανίζονται στις χωρικές συντεταγμένες που ορίζονται μέσω της διαδικασίας της δειγματοληψίας και του κβαντισμού, όπως προαναφέρθηκε και σε προηγούμενη ενότητα. Έστω λοιπόν ότι επιλέγεται να εξεταστούν οι τιμές έντασης που εμφανίζονται κατά μήκος μιας στήλης της *Εικόνας 1B*) διατηρώντας σταθερή την σειρά. Δεδομένου ότι οι διαστάσεις της υπό εξέταση εικόνας είναι 4000×3000 αυτό σημαίνει ότι το παραγόμενο μέγεθος θα είναι ένα διάνυσμα μήκους 3000 στηλών, δηλαδή 1×3000 , και θα περιλαμβάνει ένα πλήθος από διακριτές τιμές φωτεινότητας.

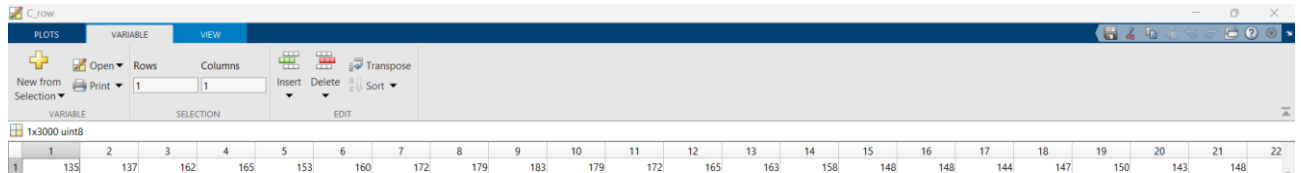
Μεταφέροντας την θεωρία στην πράξη και χρησιμοποιώντας το προγραμματιστικό περιβάλλον του MATLAB, η παρακάτω εικόνα αποδεικνύει την εν λόγω θεώρηση:

Name	Value	Class
I_gray	4000x3000 uint8	uint8
C_row	1x3000 uint8	uint8

Εικόνα 12: Από επάνω προς τα κάτω

- A) Απεικόνιση της grayscale εικόνας αναφοράς με την μορφή πίνακα στο MATLAB.
 B) Απεικόνιση του διανύσματος των 3000 στηλών.

Αξιοποιώντας την δυνατότητα που δίνει το MATLAB για προεπισκόπηση των τιμών που περιέχουν οι μεταβλητές και επιλέγοντας την C_row που αναφέρεται στο διάνυσμα στήλης, εντοπίζεται ένα μέρος του υπό εξέταση πλήθους των τιμών έντασης:



Εικόνα 13: Απεικόνιση μέρους των τιμών έντασης του διανύσματος στήλης.

Με προσεκτική μελέτη στην επεξεργασμένη Εικόνα 14 εντοπίζεται ότι στο πεπερασμένο αυτό πλήθος εμφανίζονται επανειλημμένα κάποιες τιμές φωτεινότητας με μια συχνότητα που διαφέρει ανά δείγμα. Σε ένα πολύ χονδροειδές επίπεδο, αυτό αποτελεί το ιστόγραμμα, δηλαδή μια απόθεση σε «δοχεία» (bins) της συχνότητας εμφάνισης των διακριτών τιμών έντασης των επιπέδων που εμφανίζονται σε μια ψηφιακή εικόνα.



Εικόνα 14: Απεικόνιση της εύρεσης τριών διακριτών τιμών έντασης στο πεπερασμένο πλήθος του διανύσματος στήλης.

Υιοθετώντας μια πιο επιστημονική προσέγγιση, αποτυπώνεται ότι:

Ορισμός 9: Ιστόγραμμα

Αν με r_k υποδηλώνονται τα $k = 0, 1, 2 \dots L-1$ επίπεδα φωτεινότητας μιας ψηφιακής εικόνας, τότε το μη-κανονικοποιημένο ιστόγραμμα ορίζεται ως [5]:

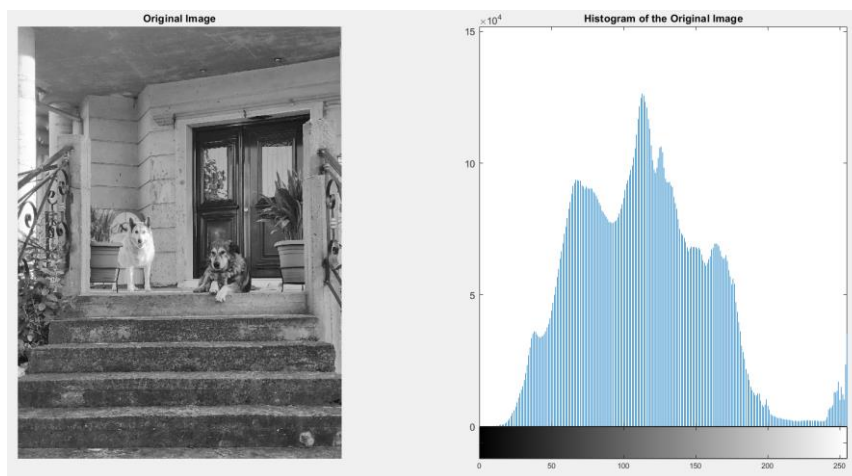
$$h(r_k) = n_k \text{ για } k = 0, 1, \dots, L - 1$$

Όπου με n_k υποδηλώνεται ο αριθμός εμφάνισης των εικονοστοιχείων στην εικόνα I με τιμές έντασης φωτεινότητας r_k . Τέλος ως ιστογραμματικά «δοχεία - bins» συμβολίζονται οι υποδιαιρέσεις της κλίμακας των τιμών φωτεινότητας [5].

Επειδή όμως συχνά πυκνά επιζητείται οι πληροφορίες, όποιες φύσης κι αν είναι αυτές, να βρίσκονται στην ίδια κλίμακα αναφοράς, τα ιστογράμματα μπορούν να υποστούν μια διαδικασία κανονικοποίησης σύμφωνα με την οποία η ιστογραμμιακές τιμές διαιρούνται με το γινόμενο των διαστάσεων των σειρών και των στηλών της εικόνας, δηλαδή:

$$p(r_k) = \frac{h(r_k)}{RC} = \frac{n_k}{RC}$$

Όπου με R συμβολίζονται οι διαστάσεις των σειρών και με C οι διαστάσεις των στηλών. Παρατηρείται ότι η σημειογραφία αλλάζει προσδίδοντας στο παραγόμενο πηλίκο την έννοια της πιθανότητας με μια εύλογη δικαιολόγηση καθώς πλέον οι ιστογραμμιακές τιμές αξιολογούνται με βάση την συχνότητα εμφάνισής τους σε σχέση με το σύνολο των διαθέσιμων τιμών στην εικόνα. Για τον προηγούμενο λόγο, το άθροισμα των τιμών αυτών οριοθετείτε μέχρι την μονάδα. Από φυσικής/μαθηματικής σημασίας, η απόδοση των ιστογραμμιακών τιμών με την μορφή πιθανοτήτων αντικατοπτρίζει μια εκτίμηση του *κατά πόσο είναι πιθανό μια συγκεκριμένη τιμή φωτεινότητας να είναι παρούσα στην υπό εξέταση εικόνα.*



Εικόνα 15: Από αριστερά προς τα δεξιά
A) Η αρχική εικόνα προς επεξεργασία.
B) Το ιστόγραμμα της αρχικής εικόνας.

1.5 Εξαγωγή Χαρακτηριστικών

Οι μέθοδοι που έχουν αναφερθεί μέχρι το παρόν σημείο είναι γενικές διαδικασίες που εντάσσονται στον πρώτο άξονα της Ψηφιακής Επεξεργασίας Εικόνας, όπου το σήμα εισόδου είναι μια εικόνα και το αποτέλεσμα της επεξεργασίας είναι μια εικόνα¹¹. Εστιάζοντας πλέον την προσοχή στον δεύτερο άξονα σύμφωνα με τον οποίο το σήμα εισόδου είναι μια εικόνα ενώ το αποτέλεσμα της επεξεργασίας της είναι ένα νέο είδος πληροφορίας, η πρώτη, και ίσως μια από τις βασικότερες έννοιες στην επεξεργασία εικόνας, έννοια που συναντά κανείς είναι τα *χαρακτηριστικά* μιας εικόνας.

Η επιστημονική κοινότητα, μέχρι και σήμερα, δεν έχει καταφέρει επιτυχώς να αποδώσει μια γενικώς περιγραφική έννοια για το τι είναι χαρακτηριστικό μιας εικόνας, αλλά αν κανείς επιζητούσε να δώσει μια περιγραφική ετυμολογία θα μπορούσε να υποστηρίξει ότι ένα χαρακτηριστικό είναι ένα σύνολο διακριτών ιδιοτήτων ή περιγραφών που εμφανίζει μια εικόνα (ή ακόμη και περιοχές μιας εικόνας), ώστε να αποδοθούν ετικέτες που να καταστούν εφικτό τον διαχωρισμό τόσο ανάμεσα σε διάφορες περιοχές μια εικόνας όταν αυτές είναι παρούσες, όσο και ανάμεσα σε διαφορετικές εικόνες μεταξύ τους. Στόχος λοιπόν είναι να φωτιστούν οι τρόποι με τους οποίους μπορούν να εξαχθούν αυτά τα ξεχωριστά χαρακτηριστικά που φέρει η κάθε εικόνα και να επισημανθούν οι μεθοδολογίες που μπορούν να επιστρατευτούν για να γίνει περιγραφή αυτών.

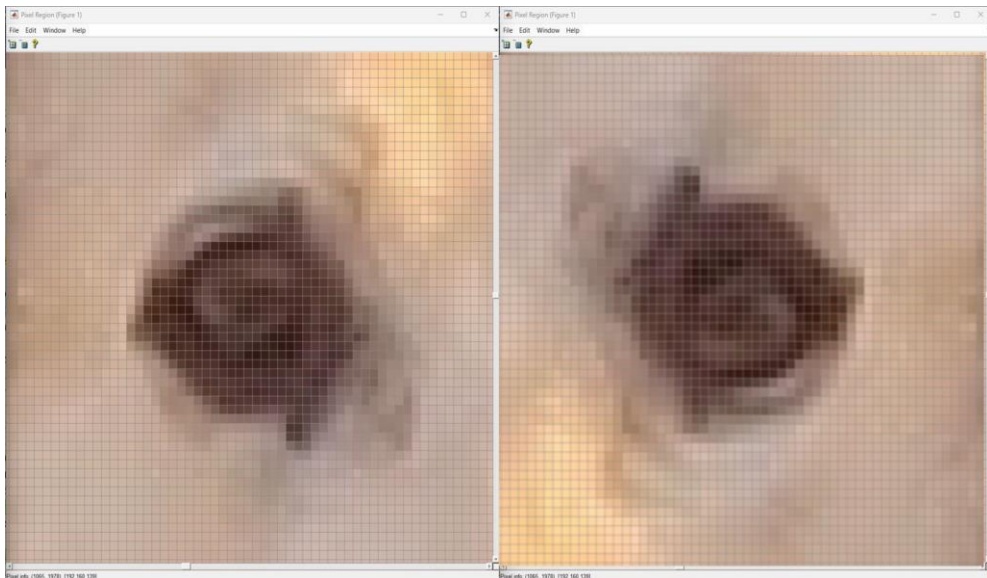
Αναφέρθηκε ότι τα χαρακτηριστικά είθισται να λειτουργούν ως ο στυλοβάτης της ανάδειξης διαφορών ανάμεσα σε εικόνες μέσω των ξεχωριστών ιδιοτήτων που φέρουν. Μια πολύ σημαντική ερώτηση που εγείρεται όμως είναι: «Ποιες είναι εκείνες οι ιδιότητες που πρέπει να φέρουν τα χαρακτηριστικά στον χώρο της εικόνας που να τα καταστούν αξιόπιστα σε οποιαδήποτε μεταβολή;» Η χρήση της λέξης «μεταβολή» συνυφαίνεται πολύ εύκολα με την λέξη «μετασχηματισμός», σχηματίζοντας μια άμεση σύνδεση με τους *affine* μετασχηματισμούς που μελετήθηκαν στην υποενότητα 1.3.3 προσδίδοντας ένα γενικό γνώρισμα σύμφωνα με το οποίο τα χαρακτηριστικά πρέπει να είναι ανεξάρτητα από μετασχηματισμούς κλιμάκωσης, περιστροφής και μετατόπισης [5].

Έχοντας εντοπίσει την ιδιότητα που καθιστά μοναδικά τα χαρακτηριστικά, κανείς τώρα αναρωτιέται: «Τι είναι αυτό που προσδιορίζεται ως ανεξαρτησία;». Στο πλαίσιο των χαρακτηριστικών, η ανεξαρτησία ορίζεται ως η αμεταβλητότητα (invariance), δηλαδή η διατήρηση των αρχικών καταστάσεων των ιδιοτήτων που εντοπίζονται σε συγκεκριμένα μέρη της εικόνας που περιγράφουν τα εξαγόμενα χαρακτηριστικά, παρά την επιβολή μιας οποιασδήποτε ομάδας μετασχηματισμών που μπορούν να εφαρμοστούν στην εικόνα υπό εξέταση. Ένα παράδειγμα που θα μπορούσε να δοθεί για

¹¹ Ίσως να προκαλεί σύγχυση στον αναγνώστη, αλλά ακόμη και το ιστόγραμμα είναι μια διαδικασία που έχει ως αποτέλεσμα μια νέα εικόνα. Στην προκειμένη περίπτωση επιλέχθηκε να γίνει αναφορά μονάχα στην έννοια του προηγούμενου κι όχι στις εφαρμογές που το τελευταίο βρίσκει πάνω στις εικόνες. Πάραυτα, ως ένδειξη καλής πρακτικής, αξίζει να σημειωθεί ότι μια από τις πιο γνωστές μεθόδους ιστογραμματικής επεξεργασίας εικόνας είναι η εξισορρόπηση ιστογράμματος για την αντιμετώπιση της άνισης διάταξης των τιμών φωτεινότητας σε μια εικόνα [5].

να γίνει πιο εύπεπτη η προηγούμενη πληροφορία είναι με μια εκ νέου επίσπευση της *Εικόνας 4Γ*). Δεδομένου ότι το σημείο ενδιαφέροντος είναι το μάτι του υποκειμένου, δηλαδή μια ελλειπτική περιοχή που ορίζεται από τα πιο σκουρόχρωμα εικονοστοιχεία, κανείς θα μπορούσε να εφαρμόσει την παρακάτω σειρά βημάτων:

- 1) Να εξάγει χαρακτηριστικά στην ελλειψοειδή περιοχή.
- 2) Να εφαρμόσει τον μετασχηματισμό της *περιστροφής*.
- 3) Να εξάγει τα νέα χαρακτηριστικά στην περιστρεφόμενη εικόνα.
- 4) Να επαναλάβει τα βήματα 2) και 3) εφαρμόζοντας όμως τον μετασχηματισμό της *μετατόπισης*.



Εικόνα 16: Από αριστερά προς τα δεξιά

A) Απεικόνιση λεπτομέρειας της αρχικής εικόνας του ματιού της Εικόνας 4Α

B) Περιστροφή της αρχικής εικόνας κατά 180 μοίρες.

Συγκρίνοντας τα χαρακτηριστικά που θα εξάγονταν από την σειρά βημάτων που περιγράφεται παραπάνω, κανείς θα διαπίστωνε ότι τα χαρακτηριστικά που θα προέκυπταν από την αρχική εξαγωγή θα ήταν ακριβώς τα ίδια με αυτά που θα προέκυπταν από την εφαρμογή των μετασχηματισμών. Αυτή ακριβώς είναι η ιδιότητα της αμεταβλητότητας στα χαρακτηριστικά που εξάγονται από μια εικόνα! Η αντίστροφη διαδικασία θα ίσχυε για την περίπτωση της κλιμάκωσης, όπως αυτή μπορεί να εντοπιστεί στις *Εικόνες 4Α*) και *4Γ*) όπου η αλλαγή της περιοχής ενδιαφέροντος θα επιφέρει και αλλαγή στα χαρακτηριστικά που θα εξαχθούν, συνιστώντας την έννοια της συμμεταβλητότητας (*covariance*). Το προηγούμενο παράδειγμα δίνει το πάτημα για να υπογραμμιστούν δυο βασικά πορίσματα σχετικά με τα χαρακτηριστικά, όπως αυτά παρουσιάζονται παρακάτω:

- 1) Είναι αντιληπτό ότι εφόσον τα χαρακτηριστικά δεν μπορούν να είναι εντελώς αμετάβλητα σε όλους τους μετασχηματισμούς ταυτόχρονα, είναι απαραίτητο να εφαρμόζονται σε πρωτότερο

στάδιο τα κατάλληλα βήματα κανονικοποίησης με σκοπό την άρση εμφάνισης όσο το δυνατόν περισσότερων καταστάσεων συμμεταβλητότητας [5].

- 2) Στο παράδειγμα με την Λίζα, τα χαρακτηριστικά μπορούν να εντοπιστούν σε διαφορετικές περιοχές εντός της εικόνας. Δηλαδή, αν ο στόχος της εφαρμογής ήταν να μπορεί να διακριθεί το υποκείμενο από κάποιο άλλο σκυλί, θα μπορούσε να εφαρμοστεί εξαγωγή χαρακτηριστικών σε όλη την αρχική εικόνα και στην συνέχεια να εφαρμοστεί ξεχωριστή εξαγωγή χαρακτηριστικών μόνο στο μάτι προσδίδοντας έτσι πιο έντονα περιγραφικές πληροφορίες για την προαναφερθείσα ενισχύοντας την μοναδικότητα της εικόνας ως σύνολο¹². Εφαρμόζοντας την ίδια διαδικασία για έναν άλλο σκύλο, όπως για παράδειγμα ο κύριος που της κρατάει συντροφιά στην *Εικόνα 1*, η επιτυχής αναγνώριση την κυρίας θα ήταν αδιαμφισβήτητη. Αυτό προσφέρει την περαιτέρω ενίσχυση του ορισμού με την ταξινόμηση των χαρακτηριστικών σε τοπικά, όταν η εξαγωγή εφαρμόζεται σε συγκεκριμένα σύνολα της εικόνας και σε ολικά όταν η εξαγωγή εφαρμόζεται σε όλη την εικόνα [5].

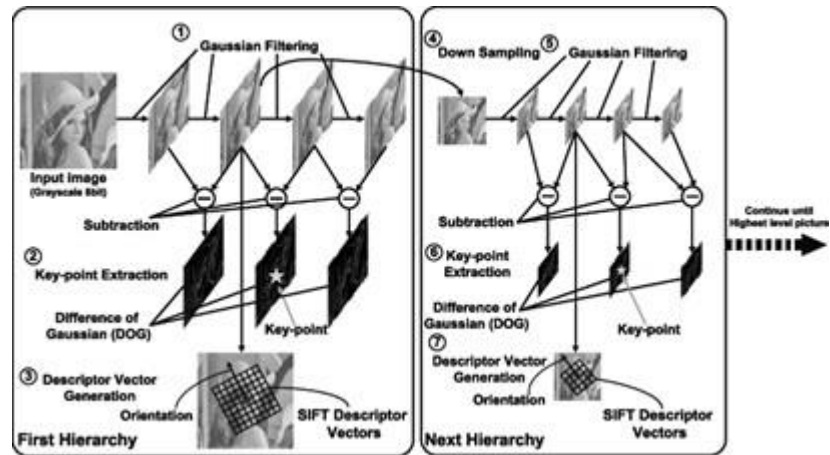
Το μοναδικό κομμάτι που υπολείπεται για να ολοκληρωθεί η ανάλυση των χαρακτηριστικών είναι ο τρόπος αναπαράστασής τους. Τα χαρακτηριστικά, όπως κι άλλα πολλά πράγματα που έχουν αναφερθεί μέχρι πρότινος σε αυτήν την εργασία, λαμβάνουν την μορφή πίνακα. Συγκεκριμένα κάθε ξεχωριστός *περιγραφέας*, δηλαδή κάθε ξεχωριστό και μοναδικό σημείο που εντοπίζεται πάνω στην εικόνα και στο οποίο θα εφαρμοστούν μια πληθώρα μαθηματικών διεργασιών, παράγει ένα *n-διάστατο* διάνυσμα που συμπυκνώνει την ωφέλιμη πληροφορία. Η συνένωση¹³ των *k* διακεκριμένων διανυσμάτων σχηματίζει τον συνολικό πίνακα χαρακτηριστικών που περιγράφει την υπό εξέταση εικόνα.

Αξίζει να γίνει μια σύντομη αναφορά σε έναν από τους πιο διάσημους αλγορίθμους εξαγωγής εύρωστων και αμετάβλητων χαρακτηριστικών που είναι ο *Scale Invariant Feature Transformation – SIFT*. Το προϊόν παραγωγής του Lowe αποτελεί την κορωνίδα των αλγορίθμων εξαγωγής χαρακτηριστικών σε μια εικόνα, καθώς η ευρετική (*heuristic*) φύση του επιτρέπει την άμεση λήψη αποδεκτών και διαισθητικών αποτελεσμάτων που προσεγγίζουν πολύ τις βέλτιστες λύσεις. Τα χαρακτηριστικά που εντοπίζονται από τον SIFT ονομάζονται σημεία-κλειδιά (*keypoints*) και είναι αμετάβλητα στην κλιμάκωση και την περιστροφή, με επιπρόσθετη ευρωστία σε μια πληθώρα *affine* παρεμβολών, σε αλλαγές στην τρισδιάστατη οπτική, τον θόρυβο καθώς

¹² Η διαδικασία που περιγράφεται απέχει πολύ από τις πραγματικές εφαρμογές εξαγωγής δεδομένων από την άποψη ότι δεν έχει νόημα να εξαχθούν δεδομένα μόνο από ιδιόζοντα χαρακτηριστικά ενός αντικειμένου (εκτός κι αν η εφαρμογή το επιζητεί), όπως εδώ το μάτι της Λίζας. Αντιθέτως μια μέθοδος που συνήθως εφαρμόζεται είναι η «σάρωση» της εικόνας με τετραγωνικά παράθυρα προκαθορισμένου μήκους εικονοστοιχείων (επικαλυπτόμενα ή μη) τα οποία σταδιακά εξάγουν και εν τέλει συνενώνουν όλα τα χαρακτηριστικά σε έναν κοινό πίνακα.

¹³ Η έννοια της συνένωσης εδώ αντιστοιχεί στην αγγλική σημειογραφία *concatenation* και μπορεί να εφαρμοστεί είτε οριζόντια είτε κάθετα, ανάλογα από τον τρόπο που σχηματίζεται το διάνυσμα.

και αλλαγές στον φωτισμό. Κλείνοντας, τονίζεται ότι η είσοδος στον αλγόριθμο είναι η υπό εξέταση εικόνα ενώ η έξοδος αποτελείται από ένα n -διάστατο διάνυσμα χαρακτηριστικών¹⁴ τα στοιχεία του οποίου είναι οι αμετάβλητοι περιγραφείς των χαρακτηριστικών [5], [27].



Εικόνα 17: Απεικόνιση της ροής των διαδικασιών που λαμβάνουν μέρος για την υλοποίηση του SIFT¹⁵.

2 ΚΕΦΑΛΑΙΟ 2^ο : Αλγόριθμοι Μάθησης

Σε αυτό το κεφάλαιο πραγματοποιείται μια επισκόπηση των κοινότυπων αλγορίθμων που χρησιμοποιούνται για να αποδοθεί η έννοια της μάθησης στις υπολογιστικές μηχανές. Σε πρώτο στάδιο υπογραμμίζονται τα χαρακτηριστικά που εμφανίζουν οι δυο προσεγγίσεις μάθησης, σημειώνοντας επιφανειακά τα πιο γνωστά είδη κι από τους δυο κόσμους. Στην συνέχεια, με γνώμονα τις πειραματικές μεθοδολογίες που χρησιμοποιήθηκαν, από τους αλγορίθμους επιβλεπόμενης μάθησης που έχουν αναφερθεί παρατίθεται το γενικευμένο θεωρητικό υπόβαθρο για τον αλγόριθμο των *Μηχανών Υποστήριξης Διανυσμάτων (Support Vector Machines)*. Το κεφάλαιο ολοκληρώνεται υιοθετώντας την ακριβώς ίδια πορεία σκέψης με προηγουμένως αλλά με αλλαγή θεματολογίας καθώς επιλέγεται να γίνει μνεία στον τρόπο λειτουργίας των Γκαουσιανών Μοντέλων Μίξης (*Gaussian Mixture Models - GMM*), ένα μοντέλο μη Επιβλεπόμενης Μάθησης, με την συμπληρωματική αποτύπωση μιας παραλλαγής τους επαναληπτικού αλγορίθμου (*Expectation Maximization - EM*) που φέρει την ονομασία *βηματικός-EM*, η σύγκλιση του οποίου αποδεικνύεται ότι επιτυγχάνεται ταχύτερα με τον παράλληλο εντοπισμό καλύτερων αποτελεσμάτων σε σχέση με τον συμβατικό EM.

¹⁴ Στην πλειονότητα των περιπτώσεων το n -διάστατο διάνυσμα που παράγεται από τον SIFT είναι 128-διαστάσεων.

¹⁵ Η παρούσα εικόνα προέρχεται από τον ακόλουθο [σύνδεσμο](#).

2.1 Βασικές κατηγορίες Αλγορίθμων Μάθησης

Είναι αδιαμφισβήτητο ότι ο άνθρωπος εμπνέεται από διάφορες πηγές μετουσιώνοντας αυτήν την έμπνευση στο καθημερινό του έργο. Δεν αποκλίνει από την πραγματικότητα να ειπωθεί ότι και η εκμάθηση των υπολογιστικών μηχανών είναι ένας τομέας που επωφελείται από την μίμηση καθημερινών δραστηριοτήτων του ανθρώπου, με τον αναφερόμενο να κεντροποιείται στον τρόπο με τον οποίο ο άνθρωπος συλλέγει και επεξεργάζεται ένα σύνολο δεδομένων ώστε να εξάγει την ωφέλιμη πληροφορία. Κανείς θα μπορούσε να αναπολήσει πίσω στην παιδική του ηλικία, εκεί που το ταξίδι της αναζήτησης για γνώση ξεκινά από το σχολείο. Σε αυτή τη στιγμή στο χρόνο και λόγω του νεαρού της ηλικίας, η εξαγωγή της γνώσης από τα παρεχόμενα δεδομένα¹⁶ δεν είναι μια διαδικασία που μπορεί να επιτευχθεί χωρίς βοήθεια. Αρωγοί σε αυτό το έργο αποτελούν οι δάσκαλοι και οι καθηγητές, οι οποίοι αναλαμβάνουν να φωτίσουν τον δρόμο και να υποδείξουν τους τρόπους με τους οποίους τα παιδιά θα εξάγουν το νόημα πίσω από τις λέξεις και θα αποκτήσουν την γνώση.

Το παραπάνω ενδεικτικό παράδειγμα ενσαρκώνει με την μορφή αναλογίας ότι τα παιδιά είναι τα υπολογιστικά συστήματα τα οποία πρέπει να μάθουν πως να εξάγουν την γνώση από τα δεδομένα που τους παρέχονται και οι *επιβλέποντες* είναι ο άνθρωπος ο οποίος θα πρέπει να υποδείξει τους τρόπους για να επιτελέσουν το ζητούμενο, θέτοντας ένα απλοϊκό προσχέδιο για την έννοια της *Επιβλεπόμενης Μάθησης*. Υιοθετώντας μια πιο επιστημονική σκοπιά, ορίζεται ότι η *Επιβλεπόμενη Μάθηση αποτελεί ένα παράδειγμα Μηχανικής Μάθησης σύμφωνα με το οποίο δυο ειδών δεδομένα, το σύνολο των δεδομένων εισόδου και το σύνολο των επιθυμητών ετικετών (labels) εξόδου χρησιμοποιούνται για να εκπαιδεύσουν ένα μοντέλο*. Όμως ερωτήματα συνεχίζουν να πλανώνται όπως: «Ποια είναι η μορφή που έχουν ή θα πρέπει να έχουν αυτά τα δεδομένα εισόδου;» και «Ποιες είναι οι διεργασίες που επιτελούνται εσωτερικά που οδηγούν στην εξαγωγή νοήματος οδηγώντας στην γνώση;» [28].

Οι παραπάνω ερωτήσεις αποτελούν το εφελτήριο για την απόθεση πολύ σημαντικών εννοιών στον τομέα της Μηχανικής Μάθησης. Ξεκινώντας από τα δεδομένα, σύμφωνα με τους *Whickham* και *Codd*: «Τα δεδομένα θεωρούνται ότι έχουν τακτοποιημένη μορφή» χωρίς αυτό να θέτει κάποιο περιορισμό στη μορφή με την οποία αυτά συνοδεύονται αλλά να εξαρτάται εξ ολοκλήρου από την αρχιτεκτονική με την οποία τα μοντέλα δημιουργούνται. Για παράδειγμα ένα μοντέλο μπορεί να δέχεται ως δεδομένα εκπαίδευσής του μια αλληλουχία λέξεων ενώ ένα άλλο μοντέλο μπορεί να δέχεται ως δεδομένα εκπαίδευσής του ακουστικά σήματα. Αν όμως χρειαζόταν να τεθεί μια ευρέως αποδεκτή μορφή που είθισται να ακολουθούν τα δεδομένα, αυτή θα ήταν ξανά με την μορφή διανυσμάτων ή κατ' επέκτασιν πινάκων [29].

¹⁶ Στο παράδειγμα που δίνεται, η γνώση σχηματίζεται από την κατανόηση των λέξεων του βιβλίου και το νόημα που κρύβεται πίσω από αυτές.

Δεδομένου ενός συνόλου N παραδειγμάτων εκπαίδευσης, τα οποία ακολουθούν την μορφή:

$$X = \begin{cases} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix}, & \text{αν τα δεδομένα είναι διανύσματα} \\ \text{ή} \\ \begin{bmatrix} x_{11} & \cdots & x_{1n} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nn} \end{bmatrix}, & \text{αν τα δεδομένα είναι πίνακες} \end{cases}$$

Και ενός συνόλου δεδομένων με ενσωματωμένη γνώση των κλάσεων στις οποίες ανήκουν τα x_i δεδομένα, φέροντας την τυπική ονομασία *ετικέτες (labels)*:

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}$$

Η διαδικασία της εκπαίδευσης κεντρίζεται στην προσπάθεια ανάκτησης μιας συνάρτησης, έστω f , που θα μπορεί να δημιουργήσει μια εννοιολογική αντιστοιχία ανάμεσα στον χώρο που συνιστούν τα δεδομένα εκπαίδευσης και το χώρο που συνιστούν τα δεδομένα εξόδου:

$$f: X \rightarrow Y \quad (1)$$

Το μέτρο της εννοιολογικής κάλυψης που μπορεί να προσφέρει η αντιστοιχία που επιφέρει η συνάρτηση υπό μελέτη περιγράφεται αποδοτικότερα όταν η (1) λαμβάνει την μορφή *συνάρτησης βαθμολόγησης (scoring function)* [28],[31], [32], [33]:

$$f: X \times Y \rightarrow \mathbb{R} \quad (2)$$

Το πραγματικό μέτρο αξιολόγησης της αποδοτικότητας που μπορεί να προσφέρει η συνάρτηση ώστε να περιγράψει όσο τον δυνατόν καλύτερα την κατανομή των δεδομένων, είναι ο εντοπισμός της απώλειας σε σχέση με τα ήδη γνωστά δεδομένα. Χωρίς να γίνει εκτενής αναφορά, εντοπίζεται ότι υπάρχουν δυο είδη απωλειών/ σφαλμάτων τα οποία επιζητούνται να ελαχιστοποιηθούν:

1. Το εμπειρικό ρίσκο (empirical risk), όπου επιζητείται να βρεθεί εκείνη η συνάρτηση που θα μπορέσει να προσαρμοστεί καλύτερα στα δεδομένα και
2. Το δομικό ρίσκο (structural risk), όπου συμπεριλαμβάνεται μια συνάρτηση επιβολής ποινής με σκοπό τον έλεγχο της αντιστάθμισης ανάμεσα στην μεροληψία (*bias*) και την διακύμανση (*variance*).

Στον αντίποδα βρίσκεται η διαδικασία της *μη-Επιβλεπόμενης* μάθησης η οποία, υιοθετώντας πάλι το παράδειγμα με το σχολείο, μπορεί να παρομοιαστεί σαν το κομμάτι του πανεπιστημίου, όπου τα παιδιά αφήνονται να εξάγουν την γνώση μόνα τους ανάλογα με την δική τους οπτική και τρόπο σκέψης. Οι βασικές διαφορές που εμφανίζει η προσέγγιση της μη-Επιβλεπόμενης μάθησης με την Επιβλεπόμενη, είναι η έλλειψη δεδομένων με ετικέτες, όπου ο στόχος είναι οι αλγόριθμοι να μπορέσουν να εξάγουν το εγγενές μοτίβο που κρύβεται στα δεδομένα καθώς κι ότι η διαδικασία της ίδιας της μάθησης περιστρέφεται γύρω από μια προσπάθεια μίμησης των δεδομένων καθώς και της χρήσης του σφάλματος στα μιμούμενα δεδομένα ώστε να επιτευχθεί η απαιτούμενη διόρθωση [30], [31], [32], [33].

Έχοντας παραθέσει τις θεμελιώδεις διαφορές ανάμεσα στις δυο προσεγγίσεις μάθησης, σειρά έχει να μελετηθούν δυο αλγόριθμοι, ένας σε κάθε κατηγορία, που αποτελούν τους θεμέλιους λίθους πάνω στους οποίους δομήθηκε η παρούσα εργασία.

2.2 Μηχανές Υποστήριξης Διανυσμάτων (*Support Vector Machines*)

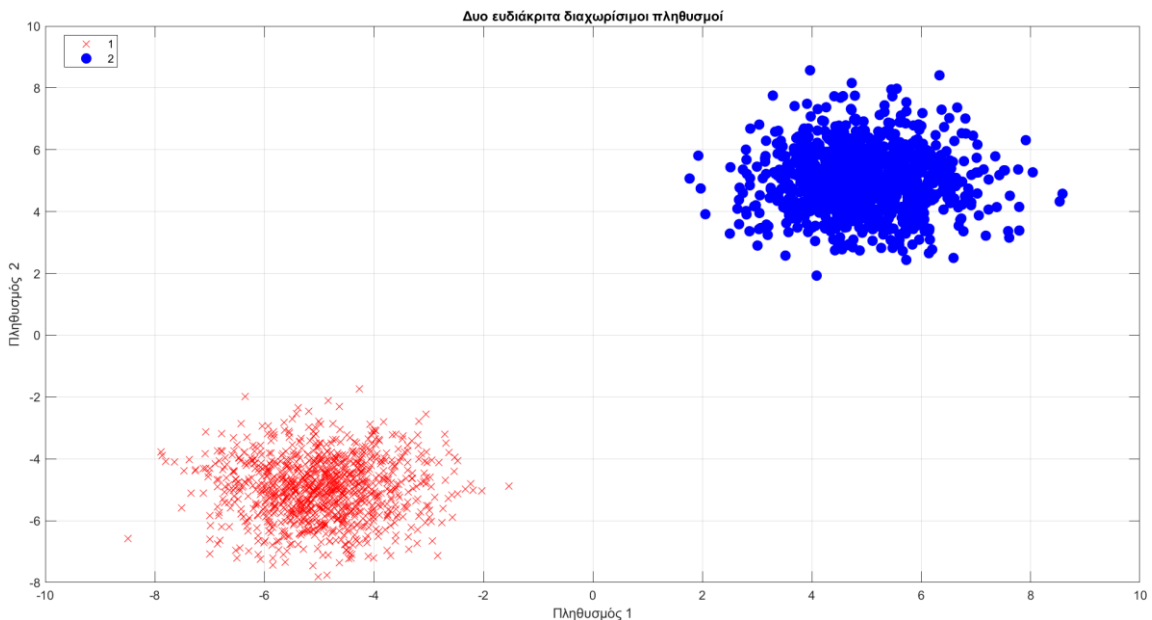
Οι Μηχανές Διανυσμάτων Υποστήριξης (SVM) αποτελούν ένα ισχυρό και ευέλικτο σύνολο αλγορίθμων επιβλεπόμενης μάθησης, που χρησιμοποιούνται ευρέως για διαδικασίες δυαδικής ταξινόμησης, παλινδρόμησης και ανίχνευσης παρεκκλίσεων. Βασίζονται κατά κόρον στις αρχές της στατιστικής θεωρίας μάθησης, που αναπτύχθηκε κυρίως από τον Vladimir Vapnik και τους συναδέλφους του τη δεκαετία του 1990 με την βασική ιδέα να είναι η εύρεση του βέλτιστου υπερεπιπέδου που μεγιστοποιεί το περιθώριο μεταξύ των διαφόρων κλάσεων σε ένα σύνολο δεδομένων, το οποίο εγγενώς ενισχύει την ικανότητα γενίκευσης του μοντέλου. Η έννοια του περιθωρίου είναι κρίσιμη, καθώς ορίζει την απόσταση μεταξύ των πλησιέστερων σημείων (διανύσματα υποστήριξης) των διαφορετικών κλάσεων στο υπερεπίπεδο. Με τη μεγιστοποίηση αυτού του περιθωρίου, ο αλγόριθμος διασφαλίζει ότι το όριο απόφασης όχι μόνο διαχωρίζει αποτελεσματικά τις κλάσεις αλλά και παραμένει ανθεκτικό σε νέα, ανεπίγνωστα σημεία δεδομένων μειώνοντας την πιθανότητα υπερπροσαρμογής (*overfitting*) [34], [35].

Πριν όμως γίνει αναφορά στις διαδικασίες που λαμβάνουν μέρος για την επίτευξη της λειτουργικότητας του αλγορίθμου, είναι πρόπον να παρατεθούν κάποιες σημαντικές πληροφορίες. Ξεκινώντας, θα πρέπει να διευκρινιστεί ότι όταν κάποιος αναφέρεται στην έννοια της *δυαδικής ταξινόμησης*, κάνει λόγο στην ύπαρξη αποκλειστικά μόνο δυο κλάσεων/ομάδων με τις πιθανές τιμές να λαμβάνονται μόνο μέσα σε αυτό το δυαδικό εύρος. Λόγω λοιπόν την ελευθερίας που προσφέρει η απλότητα της μορφής που μπορούν να λάβουν οι ετικέτες, παρατηρείται ότι η παγκόσμια βιβλιογραφία, καθώς και διαφορετικές προγραμματιστικές υλοποιήσεις, υιοθετούν μια πληθώρα από πιθανές τιμές που μπορούν να λάβουν τα αποτελέσματα της ταξινόμησης με τις επικρατέστερες να είναι είτε $\{-1,1\}$ είτε $\{0,1\}$. Εισάγοντας την απαραίτητη μαθηματική σημειογραφία και

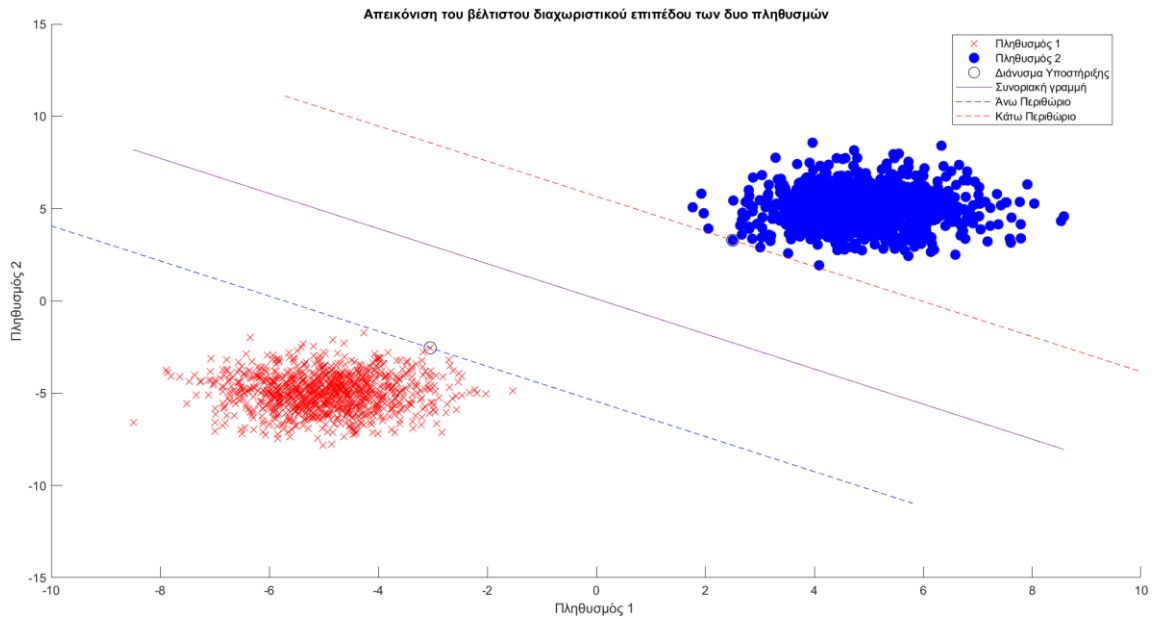
διευκρινίζοντας ότι ένα δεδομένο, έστω x_n , αναπαρίσταται ως ένα διάνυσμα D διαστάσεων που αντιστοιχεί σε ένα σημείο στον D -διάστατο χώρο, τότε οι αποκρίσεις από τις προβλέψεις αποτυπώνονται ως:

$$f: \mathbb{R}^D \rightarrow \{-1, 1\} \quad (3)$$

Κλιμακώνοντας το πρόβλημα σε ένα γενικότερο πλαίσιο, δεδομένου ενός συνόλου εκπαίδευσης που αποτελείται από ζευγάρια δεδομένων-ετικετών $\{(x_1, y_1), \dots, (x_n, y_n)\}$ επιζητείται να εκτιμηθούν οι παράμετροι των Μηχανών Υποστήριξης Διανυσμάτων που να μπορούν να παράγουν το χαμηλότερο δυνατό σφάλμα ταξινόμησης. Για να γίνει καλύτερα αντιληπτό, ας επιτραπεί η επισκόπηση της *Εικόνας 18*, όπου εντοπίζονται δυο ευδιάκριτα διαχωρίσιμοι πληθυσμοί: Ο πρώτος πληθυσμός φέρει ετικέτες που αναπαρίστανται από κόκκινους σταυρούς ενώ ο δεύτερος πληθυσμός φέρει ετικέτες που αναπαρίστανται από μπλε τελείες. Τα SVM στην πράξη αναζητούν το βέλτιστο υπερεπίπεδο, δηλαδή έναν *affine* υποχώρο διαστάσεων $D-1$, το οποίο να διαχωρίζει τους δυο αυτούς πληθυσμούς με τον βέλτιστο δυνατό τρόπο. Στην προκειμένη περίπτωση εφόσον ο χώρος είναι δισδιάστατος, $\mathbb{R}^D \xrightarrow{D=2} \mathbb{R}^2$, το υπερεπίπεδο αναμένεται να είναι μονοδιάστατο, δηλαδή να αναπαρίσταται ως μια ευθεία. Στην *Εικόνα 19* αποτυπώνεται το τελικό αποτέλεσμα από την εύρεση του διαχωριστικού επιπέδου που βρέθηκε ότι ήταν το βέλτιστο με τις επιπρόσθετες πληροφορίες της ανάδειξης των διανυσμάτων υποστήριξης καθώς και των άνω και κάτω περιθωρίων. Όλες οι έννοιες που αναφέρονται προηγουμένως αναλύονται στις επόμενες υποενότητες σε βάθος.



Εικόνα 18: Απεικόνιση δυο ευδιάκριτα διαχωρίσιμων πληθυσμών.



Εικόνα 19: Απεικόνιση του βέλτιστου διαχωριστικού επιπέδου, των διανυσμάτων υποστήριξης καθώς και των άνω και κάτω περιθωρίων.

2.2.1 Διαχωριστικό Υπερεπίπεδο

Από το εισαγωγικό σημείωμα έγινε αντιληπτό ότι το υπερεπίπεδο, όποιας μορφής κι αν είναι αυτό, λειτουργεί ως ένας διαχωριστής ανάμεσα στις κλάσεις του δυαδικού προβλήματος ταξινόμησης. Ποιο όμως είναι εκείνο το κριτήριο που δίνει την δυνατότητα να επιλεγθεί σε ποιο τμήμα του υπερεπιπέδου ανήκει ένα δεδομένο υπό εξέταση; Μια διαισθητική απάντηση, που δεν αποκλίνει καθόλου από την πραγματικότητα, είναι ο έλεγχος του μέτρου της ομοιότητας που εμφανίζουν τα σημεία του χώρου μεταξύ τους. Εφόσον είναι γνωστό ότι τα σημεία αυτά αναπαρίστανται με την μορφή διανυσμάτων, τότε αποτελεί φυσικό επακόλουθο ότι το μαθηματικό εργαλείο που χρησιμοποιείται για να περιγράψει αυτή τη σχέση ομοιότητας είναι το *εσωτερικό γινόμενο*, καθώς οι τιμές που λαμβάνονται από το τελευταίο εξαρτώνται σε μεγάλο βαθμό από το μήκος (ή αλλιώς την *νόρμα*) του κάθε διανύσματος, δίνοντας έναν απόλυτο ορισμό γεωμετρικών εννοιών όπως η ορθογωνιότητα και η προβολή.

Υιοθετώντας ξανά την ίδια υπόθεση ότι έστω ένα σημείο x το οποίο ανήκει στον D -διάστατο χώρο των δεδομένων, $x \in \mathbb{R}^D$, μπορεί να θεωρηθεί ότι υπάρχει μια συνάρτηση, f , η οποία δημιουργεί μια άμεση αντιστοιχία ανάμεσα στον χώρο των δεδομένων με τον γενικευμένο χώρο των αριθμών, $f: \mathbb{R}^D \rightarrow \mathbb{R}$, με την συνάρτηση αυτή να ορίζεται από την σχέση:

$$f(x) := \langle \mathbf{w}, x \rangle + b \quad (4)$$

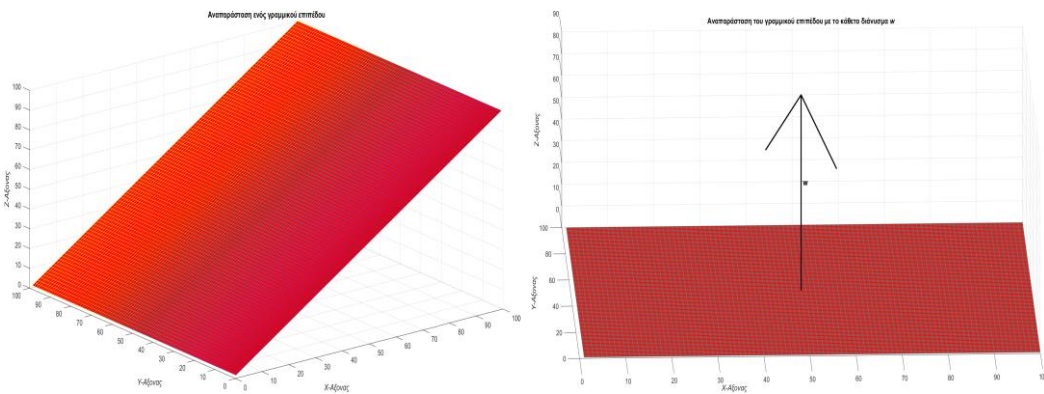
Όπου:

- Το $w \in \mathbb{R}^D$ είναι ένα διάνυσμα το οποίο είναι κάθετο στο υπερεπίπεδο¹⁷ και
- Το $b \in \mathbb{R}$ είναι η τομή του διανύσματος w με το υπερεπίπεδο.

Ορισμός 10: Υπερεπίπεδο

Ορίζεται ότι το υπερεπίπεδο, που λειτουργεί ως affine υποχώρος, που διαχωρίζει τις δυο κλάσεις ενός δυαδικού προβλήματος δίνεται από τον τύπο [36]:

$$\{x \in \mathbb{R}^D: f(x) = 0\} \quad (5)$$



Εικόνα 20: Από αριστερά προς τα δεξιά

A) Απεικόνιση ενός γραμμικού υπερεπιπέδου.

B) Απεικόνιση ενός διανύσματος κάθετο στο γραμμικό υπερεπίπεδο.

Η πραγματική αξία του Ορισμού 10 ξεπερνά τον απλό σχηματισμό του υπερεπιπέδου. Η καθετότητα του διανύσματος w επί του επιπέδου, προσθέτει την έννοια της κατεύθυνσης που υποδεικνύει σε ποια κλάση (σε ποιο κομμάτι του χώρου δηλαδή) πραγματικά ανήκει το εξεταζόμενο σημείο. Προφανώς λόγω της φύσης του προβλήματος εμφανίζονται δυο διακριτές περιπτώσεις [36]:

$$f(x) := \begin{cases} \langle w, x_n \rangle + b \geq 0, & \text{για } y_n = +1 \\ \langle w, x_n \rangle + b < 0, & \text{για } y_n = -1 \end{cases} \quad (6)$$

Η σχέση (6) μπορεί να λάβει μια πιο συμπυκνωμένη μορφή, η οποία εξαρτάται εξ ολοκλήρου από την τιμή που παίρνει η μεταβλητή y_n :

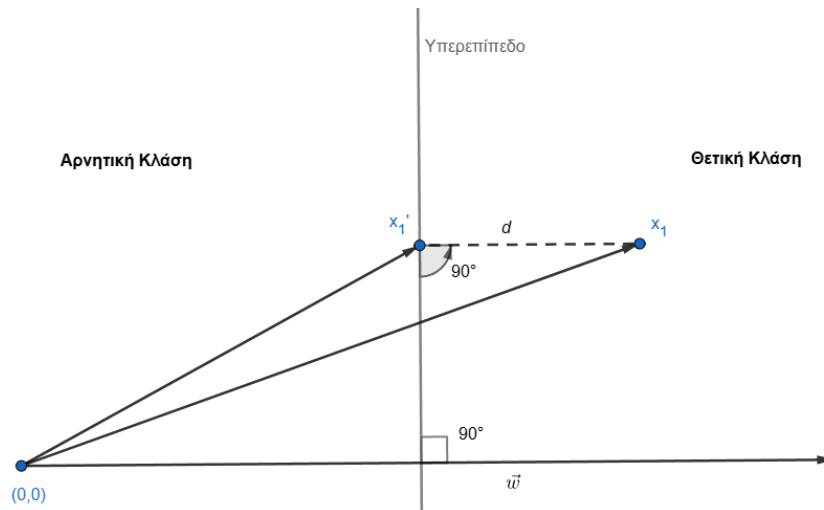
$$y_n(\langle w, x_n \rangle + b) \geq 0 \quad (7)$$

¹⁷ Είναι λογικό κανείς να καταλήξει στο συμπέρασμα πως οποιοδήποτε σημείο πάνω στο υπερεπίπεδο εμφανίζει εσωτερικό γινόμενο με το διάνυσμα w που ισούται με το μηδέν. Το φαινόμενο αυτό προσδίδει την ιδιότητα της ορθογωνιότητας του διανύσματος w ως προς όλα τα πιθανά σημεία πάνω στο υπερεπίπεδο.

2.2.2 Η έννοια του Περιθωρίου

Η ετυμολογία της λέξης του Περιθωρίου προδίδει άμεσα την φύση που πραγματεύεται, με το τελευταίο να αποτελεί την απόσταση του διαχωριστικού υπερεπιπέδου από το κοντινότερο δεδομένο (ή δεδομένα). Η ίδια η έννοια της απόστασης όμως φαίνεται πως φέρει ένα σημαντικό τροχοπέδη καθώς για να μπορέσει κανείς να την ορίσει με τον πρέπον τρόπο θα πρέπει να λάβει υπόψιν του ποια θα είναι η κλίμακα αναφοράς της μέτρησης.

Η προσέγγιση του προβλήματος, ορίζοντας την κλίμακα συναρτήσει της εξίσωσης (5) του υπερεπιπέδου, θέτει ως βασικός άξονας ενδιαφέροντος την κατεύθυνση που έχει το διάνυσμα w . Η ενασχόληση αποκλειστικά με την κατεύθυνση προσδίδει ότι το μήκος του διανύσματος θεωρείται πως είναι μονάδα, δηλαδή ότι $\|w\| = 1^{18}$, καθώς κι ότι τα δεδομένα είναι γραμμικώς διαχωρίσιμα. Έστω λοιπόν ένα υπερεπίπεδο $\langle w, x \rangle + b$ καθώς κι ενός δεδομένου το οποίο εντοπίζεται για παράδειγμα στο κομμάτι της θετικής κλάσης, όπως φαίνονται στην *Εικόνα 21*.



Εικόνα 21: Αποτύπωση της απόστασης του σημείου x_1 από το υπερεπίπεδο.

Στόχος είναι να μπορέσει να βρεθεί ένας τρόπος εκπεφρασμού της μεγιστοποίησης της απόστασης $d > 0$, έτσι ώστε να ισχύει ότι το εκάστοτε δεδομένο x_1 ικανοποιεί την σχέση $\langle w, x_1 \rangle + b > 0$. Για να μπορέσει να γίνει αυτό, γίνεται χρήση της ορθογωνιότητας του διανύσματος w με το υπερεπίπεδο και εντοπίζεται η ορθογώνια προβολή του x_1 πάνω σε αυτό, δηλαδή εδώ το x_1' . Δεδομένης της διανυσματικής φύσης της προβολής, είναι εύκολο να συναχθεί ότι το διάνυσμα w μπορεί να ιδωθεί ως μια κλιμάκωση της προβολής του x_1 , οδηγώντας σε μια νέα εξίσωση:

$$x_a = x_a' + d \frac{w}{\|w\|} \quad (8)$$

¹⁸ Όπου ο τελεστής $\| \cdot \|$ υποδεικνύει την Ευκλείδεια νόρμα $\|w\| = \sqrt{w^T w}$.

Ο τύπος (8) χρησιμοποιεί την απόσταση d ως έναν παράγοντα κλιμάκωσης του μοναδιαίου διανύσματος που δείχνει προς την κατεύθυνση του w ($\frac{w}{\|w\|}$) για να αποτυπώσει την απόσταση του σημείου x_1 από το υπερεπίπεδο. Αν δοθεί η ελευθερία να ειπωθεί ότι το σημείο υπο εξέταση είναι και το κοντινότερο στο υπερεπίπεδο, τότε η απόσταση d οριοθετεί το περιθώριο, μετασχηματίζοντας τον τύπο (7) σε [36]:

$$y_n(\langle w, x_n \rangle + b) \geq d \quad (9)$$

Το τελικό πρόβλημα βελτιστοποίησης προς επίλυση αποτυπώνεται:

$$\begin{aligned} & \max_{w,b,d} d \\ & \text{Δεδομένου ότι } \underbrace{y_n(\langle w, x_n \rangle + b) \geq d}_{\text{προσαρμογή στα δεδομένα}} \text{ , } \underbrace{\|w\| = 1}_{\text{κανονικοποίηση}} \text{ , } d > 0 \end{aligned} \quad (10)$$

Κι ενώ το πρόβλημα φαίνεται συμπαγές, ο μη-κυρτός περιορισμός του $\|w\| = 1$ αποτελεί σημαντικό τροχοπέδη για όλα γνωστά λογισμικά επίλυσης βελτιστοποίησης. Τι θα γινόταν λοιπόν αν άλλαζε αυτό; Έστω ότι διαιρούμε την ποσότητα κλιμάκωσης d με το προβληματικό μέρος των περιορισμών, ώστε να μεγιστοποιηθεί πλέον ο όρος $\frac{d}{\|w\|}$, τότε το πρόβλημα θα ανακατασκευαζόταν ως:

$$\begin{aligned} & \max_{w,b,d} \frac{d}{\|w\|} \\ & \text{Δεδομένου ότι } y_n(\langle w, x_n \rangle + b) \geq 1, \text{ για όλα τα } n = 1, 2, \dots, N \end{aligned}$$

Το οποίο και πάλι εμφανίζει ένα μη-κυρτό πρόβλημα, εδώ το $\frac{d}{\|w\|}$. Με τις παρατηρήσεις ότι:

- 1) Οποιαδήποτε κλιμάκωση των όρων w, b δεν επιφέρει καμιά απολύτως αλλαγή, οδηγώντας στην μορφοποίηση του περιορισμού ότι το μέγιστο περιθώριο είναι ίσο με την μονάδα, δηλαδή $d = 1$.
- 2) Η επίλυση της μεγιστοποίησης του $\frac{d}{\|w\|} = \frac{1}{\|w\|}$ είναι το ίδιο με την ελαχιστοποίηση του $\|w\|^2$.

Το τελικό πρόβλημα βελτιστοποίησης προς επίλυση αποτυπώνεται ως [37]:

$$\begin{aligned} & \min_{w,b,d} \frac{1}{2} \|w\|^2 \\ & \text{Δεδομένου ότι } y_n(\langle w, x_n \rangle + b) \geq 1, \text{ } n = 1, 2, \dots, N \end{aligned} \quad (11)$$

Ο τύπος (11) είναι γνωστός ως *αυστηρό περιθώριο του SVM*.

2.2.3 Από το Πρωτεύον στο Δυαδικό πρόβλημα :Κυρτή δυαδικότητα μέσω πολλαπλασιαστών Lagrange

Όπως φαίνεται, μεγάλο κομμάτι των SVM αποτελούν οι προσπάθειες εύρεσης περιγραφικότερων και ευκολότερα εφαρμόσιμων αντικειμενικών συναρτήσεων προς επίλυση. Το πρώτο είδος προβλήματος που εμφανίζεται είναι το *πρωτεύον πρόβλημα*:

$$\begin{aligned} & \min_w f(w) \\ & \text{Δεδομένου ότι: } g_i(w) \leq 0, i = 1, \dots, k \\ & h_i(w) = 0, i = 1, \dots, l \end{aligned} \quad (12)$$

Για να λυθεί το άνωθεν, αποδίδεται η γενικευμένη μορφή της Lagrangian ως:

$$\mathcal{L}(w, \alpha, \beta) = f(w) + \sum_{i=1}^k \alpha_i g_i(w) + \sum_{i=1}^l \beta_i h_i(w) \quad (13)$$

Όπου τα α_i και β_i είναι οι **πολλαπλασιαστές Lagrange**. Η χρήση των τελευταίων γίνεται:

- Ο α μόνο όταν παρουσιάζονται ανισότητες.
- Ο β μόνο όταν παρουσιάζονται ισότητες.

Θεωρώντας την ποσότητα

$$\theta_P(w) = \max_{\alpha, \beta: \alpha_i \geq 0} \mathcal{L}(w, \alpha, \beta) \quad (14)$$

Με αντικατάσταση του τύπου (14) στον τύπο της ελαχιστοποίησης (12) μπορεί να γραφεί ότι:

$$\min_w \theta_P(w) = \min_w \max_{\alpha, \beta: \alpha_i \geq 0} \mathcal{L}(w, \alpha, \beta) \quad (15)$$

Αλλάζοντας τώρα οπτική, ας θεωρηθεί το παρακάτω πρόβλημα:

$$\theta_D(\alpha, \beta) = \min_w \mathcal{L}(w, \alpha, \beta) \quad (16)$$

Παρατηρείται ότι σε αντίθεση με τον τύπο (14) αντί να επιζητείται να μεγιστοποιηθεί το $\theta_P(w)$ ως προς α, β , εδώ επιζητείται να ελαχιστοποιηθεί ως προς w . Ορίζεται λοιπόν ότι το *δυαδικό πρόβλημα βελτιστοποίησης* είναι το:

$$\max_{\alpha, \beta: \alpha_i \geq 0} \theta_D(\alpha, \beta) = \max_{\alpha, \beta: \alpha_i \geq 0} \min_w \mathcal{L}(w, \alpha, \beta) \quad (17)$$

¹⁹Γενικά ισχύει πως $\max \min \leq \min \max$.

Αν οριστεί ότι η βέλτιστη τιμή του πρωτεύοντος και η βέλτιστη τιμή του δευτερεύοντος προβλήματος δίνονται ως:

$$p^* = \min_w \theta_P(w) \text{ και } d^* = \max_{\alpha, \beta: \alpha_i \geq 0} \theta_D(w)$$

Τότε κάτω από συγκεκριμένες συνθήκες μπορεί να εμφανιστεί το φαινόμενο κατά το οποίο οι βέλτιστες τιμές είναι ίσες μεταξύ τους, δηλαδή $p^* = d^*$. Οι συνθήκες αυτές είναι γνωστές ως **Karush-Kuhn-Tucker** και συνοψίζονται ως εξής [37]:

Συνθήκες Karush-Kuhn-Tucker

Πρέπει να υπάρχουν w^*, α^* και β^* έτσι ώστε το w^* να είναι η λύση στο δυαδικό πρόβλημα, τα α^* και β^* να είναι οι λύσεις του δυαδικού προβλήματος και επιπρόσθετα $p^* = d^* = L(w^*, \alpha^*, \beta^*)$. Ακόμη, τα w^*, α^*, β^* θα πρέπει να ικανοποιούν τις συνθήκες KKT που είναι:

$$\frac{\partial \theta}{\partial w_i} \mathcal{L}(w^*, \alpha^*, \beta^*) = 0, \quad i = 1, 2, \dots, n \quad (18)$$

$$\frac{\partial \theta}{\partial \beta_i} \mathcal{L}(w^*, \alpha^*, \beta^*) = 0, \quad i = 1, 2, \dots, l \quad (19)$$

$$\alpha_i^* g_i(w^*) = 0, \quad i = 1, 2, \dots, k \quad (20)$$

$$g_i(w^*) \leq 0, \quad i = 1, 2, \dots, k \quad (21)$$

$$\alpha^* \geq 0, \quad i = 1, 2, \dots, k \quad (22)$$

Αν κάποια από τα w^*, α^*, β^* ικανοποιεί τις παραπάνω συνθήκες τότε είναι λύση και στο πρωτεύον αλλά και στο δυαδικό πρόβλημα. Δίνετε ιδιαίτερη έμφαση στην συνθήκη (20) η οποία ονομάζεται **KKT συνθήκη Δυαδικής Συμπληρωματικότητας (KKT dual complementarity condition)**. Η προκειμένη εκφράζει ότι **αν το $\alpha_i^* > 0$ τότε το $g_i(w^*) = 0$** , δηλαδή ο περιορισμός $g_i(w) \leq 0$, τύπος (12), είναι ενεργός, κάτι που σημαίνει **ότι ικανοποιεί την ισότητα κι όχι την ανισότητα**. Αυτό είναι πολύ σημαντικό, **γιατί δείχνει ότι το SVM έχει έναν μικρό αριθμό διανυσμάτων υποστήριξης** όπως μπορεί να εντοπιστεί και από την *Εικόνα 19* [37].

2.2.4 Ο ταξινομητής βέλτιστου Περιθωρίου

Μια βασική ιδέα στο σχηματισμό του δυαδικού προβλήματος, είναι να μπορέσει να εκφραστεί ο αλγόριθμος συναρτήσει μόνο του εσωτερικού γινομένου $\langle x^{(i)}, x^{(j)} \rangle$, κάτι το οποίο μπορεί να γίνει αντιληπτό ως $(x^{(i)})^T x^{(j)}$ ανάμεσα στα σημεία του χώρου χαρακτηριστικών. Σχηματίζεται λοιπόν εκ νέου η *Lagrangian* ως:

$$\mathcal{L}(w, b, a) = \frac{1}{2} \|w\|^2 - \sum_{i=1}^m a_i [y^{(i)} (w^T x^{(i)} + b) - 1] \quad (23)$$

Ελαχιστοποιώντας την (23) ως προς w και b μέσω της εύρεσης του πού η παράγωγος ισούται με το μηδέν θα ισχύει ότι:

$$\nabla_D \mathcal{L}(w, b, a) = w - \sum_{i=1}^m a_i y^{(i)} x^{(i)} = \mathbf{0} \quad (24)$$

Ενώ για την παράγωγο συναρτήσεως του b θα ισχύει:

$$\frac{\partial}{\partial b} \mathcal{L}(w, b, a) = \sum_{i=1}^m a_i y^{(i)} = 0 \quad (25)$$

Λύνοντας την (24) προκύπτει ότι:

$$w = \sum_{i=1}^m a_i y^{(i)} x^{(i)} \quad (26)$$

Και με την γνώση που προκύπτει από την (25), ο τελευταίος όρος της *Lagrangian* μηδενίζεται:

$$\begin{aligned} \mathcal{L}(w, a, b) &= \sum_{i=1}^m a_i - \frac{1}{2} \sum_{i,j=1}^m y^{(i)} y^{(j)} a_i a_j (x^{(i)})^T x^{(j)} - b \sum_{i=1}^m a_i y^{(i)} \leftrightarrow \\ \mathcal{L}(w, a, b) &= \sum_{i=1}^m a_i - \frac{1}{2} \sum_{i,j=1}^m y^{(i)} y^{(j)} a_i a_j (x^{(i)})^T x^{(j)} \end{aligned} \quad (27)$$

Προσθέτοντας και τους περιορισμούς, η τελική μορφή του δυαδικού προβλήματος βελτιστοποίησης με την παρουσία εσωτερικού γινομένου, διαμορφώνεται ως:

$$\max_{\alpha} W(\alpha) = \sum_{i=1}^m \alpha_i - \frac{1}{2} \sum_{i,j=1}^m y^{(i)} y^{(j)} \alpha_i \alpha_j \langle x^{(i)} x^{(j)} \rangle \quad (28)$$

Δεδομένου ότι: $\alpha_i \geq 0, \quad i = 1, \dots, m$ και

$$\sum_{i=1}^m \alpha_i y^{(i)} = 0$$

Αν λοιπόν το πρόβλημα λύνεται, τότε μπορεί να χρησιμοποιηθεί ο τύπος (26) ώστε να ανακτηθούν τα βέλτιστα w συναρτήσεως του a . Αν με την σειρά τους βρεθούν και τα w τότε τα b εντοπίζονται μέσω της σχέσης:

$$b_* = -\frac{\max_{i:y^{(i)}=-1} w^{*T} x^{(i)} + \min_{i:y^{(i)}=1} w^{*T} x^{(i)}}{2} \quad (29)$$

Με την υπόθεση ότι έχουν βρεθεί και υπολογιστεί οι παράμετροι που μπορούν να οδηγήσουν το μοντέλο ώστε να διαχωρίζει με τον βέλτιστο δυνατό τρόπο τα δεδομένα, ο επόμενος στόχος είναι να αξιολογηθεί η αποδοτικότητά του με ένα καινούργιο δεδομένο x . Για να γίνει αυτό, θα χρειαζόταν να υπολογιστούν τα $w^T x + b$ και να πραγματοποιηθεί μια πρόβλεψη ότι το $y = 1$ **αν και μόνο αν αυτή η ποσότητα είναι μεγαλύτερη από το μηδέν**. Έχοντας τα προηγούμενα υπόψιν, η παραγωγή προβλέψεων θα γινόταν μέσω του τύπου:

$$w^T x + b = \left(\sum_{i=1}^m a_i y^{(i)} x^{(i)} \right)^T x + b \quad (30)$$

$$= \sum_{i=1}^m a_i y^{(i)} \langle x^{(i)}, x^{(j)} \rangle + b \quad (31)$$

2.2.5 Πυρήνες (Kernels)

Μέχρι αυτό το σημείο, ήταν επικρατούσα η υπόθεση ότι τα δεδομένα που είναι διαθέσιμα είναι γραμμικώς διαχωρίσιμα. Τι γίνεται όμως όταν αυτό δεν μπορεί να λεχθεί με απόλυτη βεβαιότητα ή αν τα δεδομένα δεν είναι όντως γραμμικώς διαχωρίσιμα; Η λύση σε αυτό το τροχοπέδη βρέθηκε το 1992 από την ομάδα του Vapnik ξανά, όπου προτάθηκε να δημιουργηθούν μη-γραμμικοί ταξινομητές εφαρμόζοντας το κόλπο των πυρήνων (*kernel trick*), το οποίο προτάθηκε αρχικά από την ομάδα του Aizerman [38].

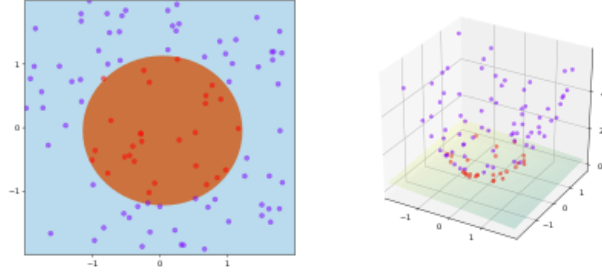
Ένας πυρήνας²⁰ είναι ουσιαστικά μια συνάρτηση που υπολογίζει την ομοιότητα μεταξύ δύο σημείων δεδομένων, με την μαθηματική σημειογραφία:

$$K(x_i, x_j) = \langle \phi(x_i), \phi(x_j) \rangle \quad (32)$$

Εδώ, το ϕ δηλώνει μια συνάρτηση απεικόνισης που μετασχηματίζει τα αρχικά σημεία δεδομένων σε έναν χώρο υψηλότερων διαστάσεων $\phi: \mathcal{X} \rightarrow \mathcal{V}$. Η ομορφιά της χρήσης πυρήνων είναι ότι δεν χρειάζεται να γνωρίζουμε ή να υπολογίζουμε ρητά την απεικόνιση ϕ δίνοντας την δυνατότητα οι προσπάθειες να επικεντρωθούν μόνο στη συνάρτηση πυρήνα K , γεγονός που απλοποιεί σημαντικά την υπολογιστική διαδικασία. Η μεταφορά και η ενασχόληση σε έναν χώρο υψηλότερων διαστάσεων είναι κρίσιμη καθώς μπορεί στον αρχικό χώρο των δεδομένων να μην υπάρχει κάποιο διαχωριστικό υπερεπίπεδο που να ταξινομεί τα δεδομένα αποτελεσματικά, αλλά μέσω της μη-γραμμικής

²⁰ Στην επιστήμη των μαθηματικών, η λέξη «πυρήνας» χρησιμοποιείται για να υποδηλώσει μια συνάρτηση στάθμησης για ένα σταθμισμένο άθροισμα ή ολοκλήρωμα [39].

αντιστοίχισης στον μετασχηματισμένο χώρο χαρακτηριστικών υψηλότερων διαστάσεων, το υπερεπίπεδο να υπάρχει, όπως μπορεί να εντοπιστεί στην *Εικόνα 22*!



Εικόνα 22: Το SVM με έναν πυρήνα που δίνεται από την σχέση $\varphi(\mathbf{a}, \mathbf{b}) = (\mathbf{a}, \mathbf{b}, \mathbf{a}^2 + \mathbf{b}^2)$ δίνοντας ότι $\mathbf{k}(\mathbf{x}, \mathbf{y}) = \mathbf{x} \cdot \mathbf{y} + \|\mathbf{x}\|^2 \|\mathbf{y}\|^2$. Τα δεδομένα εκπαίδευσης αντιστοιχίζονται σε έναν τρισδιάστατο χώρο όπου το υπερεπίπεδο μπορεί να εντοπιστεί εύκολα ²¹.

Πως όμως είναι σίγουρο ότι μια συνάρτηση K είναι όντως ένας έγκυρος πυρήνας, δηλαδή ότι όντως μπορεί να γεφυρώσει τον αρχικό με τον μετασχηματισμένο χώρο των χαρακτηριστικών; Ας υποθεθεί για αρχή ότι ο πυρήνας K είναι όντως έγκυρος, ότι υπάρχουν κάποια πεπερασμένα σημεία, έστω m , καθώς κι ότι ο K αναπαρίσταται ως ένας πίνακας. Αν όντως ο πίνακας K είναι έγκυρος, τότε θα ισχύει ότι:

$$K_{ij} = K(x^{(i)}, x^{(j)}) = \varphi(x^{(i)})^T \varphi(x^{(j)}) = \varphi(x^{(j)})^T \varphi(x^{(i)}) = K(x^{(j)}, x^{(i)}) = K_{ji}$$

υποδεικνύοντας ότι ο K είναι συμμετρικός. Αν με $\varphi_k(x)$ αναπαρίσταται η k -οστή συντεταγμένη του διανύσματος $\varphi(x)$ και ισχύει ότι $x, z \in \mathbb{R}^n$, τότε αποδεικνύεται ότι για κάθε διάνυσμα δεδομένων z ισχύει ότι:

$$\begin{aligned} z^T K z &= \sum_i \sum_j z_i K_{ij} z_j \\ &= \sum_i \sum_j z_i \varphi(x^{(i)})^T \varphi(x^{(j)}) z_j \\ &= \sum_i \sum_j z_i \sum_k \varphi_k(x^{(i)}) \varphi_k(x^{(j)}) z_j \\ &= \sum_i \sum_j \sum_k z_i \varphi_k(x^{(i)}) \varphi_k(x^{(j)}) z_j \\ &= \sum_k \left(\sum_i z_i \varphi_k(x^{(i)}) \right)^2 \geq 0 \end{aligned}$$

Εφόσον λοιπόν το z είναι αυθαίρετο, ο K είναι ένας θετικός ημι-ορισμένος πίνακας καθώς το $K \geq 0$. Τα παραπάνω αποτελούν μια επαλήθευση του θεωρήματος του Mercer, το οποίο διατυπώνει πως αν

²¹ Η εικόνα λήφθηκε από τον [σύνδεσμο](#).

ο K είναι όντως ένας έγκυρος πυρήνας (*Mercer*), είναι ικανή και αναγκαία συνθήκη ο αντίστοιχος πίνακας του πυρήνα να είναι συμμετρικός θετικά ημι-ορισμένος [37], [40].

Η σφραγίδα που διασφαλίζει ότι ο αλγόριθμος μπορεί να διαχωρίσει μη-γραμμικά δεδομένα καθώς κι ότι δεν είναι τόσο επιρρεπής σε ακραίες τιμές (*outliers*), είναι ο επαναπροσδιορισμός του πρωτεύοντος προβλήματος βελτιστοποίησης ως:

$$\min_{r,w,b} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^m \xi_i \quad (33)$$

Δεδομένου ότι $y^{(i)}(w^T x^{(i)} + b) \geq 1 - \xi_i, i = 1, 2, \dots, m$ και

$$\xi_i \geq 0, i = 1, 2, \dots, m$$

Η παραπάνω μορφή του πρωτεύοντος προβλήματος άρει την αυστηρότητα που εμφανίζεται στον τύπο (11) επιτρέποντας πλέον στα δεδομένα να έχουν περιθώριο το οποίο είναι μικρότερο της μονάδας. Αν όντως λοιπόν κάποιο από τα δεδομένα εμφανίζεται σε απόσταση $1 - \xi_i$ του περιθωρίου, τότε επιβάλλεται ποινή που οδηγεί στην αύξηση της αντικειμενικής συνάρτησης κατά $C \xi_i$, όπου του C ελέγχει την στάθμιση της ποινής με τέτοιο τρόπο ώστε να επιτυγχάνεται και το μέγιστο $\|w\|^2$ αλλά και να διασφαλίζεται ότι τα περισσότερα δεδομένα εμφανίζουν περιθώριο το οποίο είναι τουλάχιστον 1.

Το αντίστοιχο δυαδικό πρόβλημα βελτιστοποίησης θα είναι:

$$\max_a W(a) \sum_{i=1}^m a_i - \frac{1}{2} \sum_{i,j=1}^m y^{(i)} y^{(j)} a_i a_j \langle x^{(i)}, x^{(j)} \rangle \quad (34)$$

Δεδομένου ότι $0 \leq a_i \leq C, i = 1, 2, \dots, m$ και

$$\sum_{i=1}^m a_i y^{(i)} = 0$$

2.3 Αλγόριθμοι Μη Επιβλεπόμενης Μάθησης

Οι μη επιβλεπόμενοι αλγόριθμοι μάθησης είναι ένας συναρπαστικός κλάδος της μηχανικής μάθησης που λειτουργεί χωρίς την καθοδήγηση δεδομένων με ετικέτες. Σε αντίθεση με την επιβλεπόμενη μάθηση, ο πρωταρχικός στόχος της μάθησης χωρίς επίβλεψη είναι η αποκάλυψη της υποκείμενης δομής των δεδομένων ή του μοτίβου που αυτά εμφανίζουν. Μια από τις πιο γνωστές μεθόδους στη μάθηση χωρίς επίβλεψη είναι η ομαδοποίηση (*clustering*)²². Οι αλγόριθμοι ομαδοποίησης συγκεντρώνουν τα σημεία δεδομένων με βάση τις ομοιότητές τους (ή τις διαφορές τους) σε κοινές ομάδες, με στόχο να διασφαλίσουν ότι τα σημεία εντός της ίδιας συστάδας είναι περισσότερο όμοια μεταξύ τους από ό,τι εκείνα σε άλλες συστάδες [41].

Ένας απλός και ευρέως διαδεδομένος αλγόριθμος ομαδοποίησης είναι ο *k-means*. Με τον τελευταίο να αποτελεί μια άμεση μέθοδο διανυσματικής κβάντισης, στόχος είναι η τμηματοποίηση του χώρου των δεδομένων σε έναν νέο χώρο²³ που περιγράφεται από έναν διακριτό αριθμό ομάδων/συστάδων εκπροσωπούμενες από τον αντίστοιχο αριθμό κέντρων. Ο αλγόριθμος ξεκινά με την τυχαία τοποθέτηση k σημείων, τα οποία ονομάζονται κεντροειδή (*centroids*), εντός του χώρου δεδομένων. Στη συνέχεια, κάθε σημείο δεδομένων αντιστοιχίζεται στο πλησιέστερο κεντροειδές χρησιμοποιώντας μια συγκεκριμένη απόσταση, με την πιο συχνά χρησιμοποιούμενη να είναι η Ευκλείδεια, σχηματίζοντας συστάδες. Στη συνέχεια, τα κεντροειδή υπολογίζονται εκ νέου ως ο μέσος όρος των σημείων σε κάθε συστάδα και η διαδικασία επαναλαμβάνεται έως ότου τα κεντροειδή δεν μετακινούνται πλέον σημαντικά. Μαθηματικά, ο στόχος του *k-means* είναι η ελαχιστοποίηση του αθροίσματος των τετραγωνικών αποστάσεων μεταξύ των σημείων δεδομένων και των αντίστοιχων κεντροειδών τους (*within-cluster sum of squares- WCSS*), δηλαδή της *διακύμανσης*.

Δοσμένου ενός συνόλου δεδομένων (x_1, x_2, \dots, x_n) όπου κάθε δεδομένο αναπαρίσταται ως ένας d -διάστατο διάνυσμα πραγματικών τιμών, ο *k-means* στοχεύει στην τμηματοποίηση των n δεδομένων σε k σύνολα $S = \{S_1, S_2, \dots, S_k\}$, με $k \leq n$, ώστε να ελαχιστοποιήσει την διακύμανση σύμφωνα με το πρόβλημα βελτιστοποίησης που ορίζεται από τον τύπο [42], [43], [44], [45]:

$$\arg \min_S \sum_{i=1}^k \sum_{x \in S_i} \|x - \mu_i\|^2 = \arg \min_S \sum_{i=1}^k |S_i| \text{Var } S_i \quad (35)$$

Όπου

²² Συχνά κανείς μπορεί να συναντήσει την έννοια της συσταδοποίησης και ως συσταδική ανάλυση (*cluster analysis*).

²³ Ο νέος τμηματοποιημένος χώρος είναι γνωστός ως Voronoi κελιά.

- Το μ_i είναι η μέση τιμή ή εδώ τα *κεντροειδή* των πολυδιάστατων σημείων S_i :

$$\mu_i = \frac{1}{|S_i|} \sum_{x \in S_i} x$$

με το $|S_i|$ να υποδηλώνει το μέγεθος του S_i

- Το $|\cdot|$ να είναι η Ευκλείδεια νόρμα.

Τα βήματα που λαμβάνονται για την επιτυχή υλοποίηση του αλγορίθμου είναι:

Algorithm 1 *K-Means Clustering*

- 1: **Input:** data, numClusters, maxIter
- 2: **Output:** Centroids
- 3: Initialize centroids randomly from the data points
- 4: **while** not converged and $iter \leq maxIter$ **do**
- 5: Assign each point to the nearest centroid

$$S_i^{(t)} = \{x_p : \|x_p - m_i^{(t)}\|^2 \leq \|x_p - m_j^{(t)}\|^2 \forall j, 1 \leq j \leq k\}$$

- 6: Update centroids based on current cluster assignments

$$m_i^{(t)} = \frac{1}{|S_i^{(t)}|} \sum_{x_j \in S_i^{(t)}} x_j$$

- 7: Check for convergence (e.g., change in centroids is below a threshold)
 - 8: $iter \leftarrow iter + 1$
 - 9: **end while**
 - 10: **return** Centroids
-

Αλγόριθμος 1: Συσταδοποίηση κατά k-Means.

Ενώ ο *k-means* είναι εύκολος στην κατανόηση και την εφαρμογή, έχει περιορισμούς ειδικά σε πολύπλοκες κατανομές δεδομένων. Με την υπολογιστική του πολυπλοκότητα να εκφράζεται ως $\mathcal{O}(n \cdot d \cdot k \cdot i)^{24}$, όπου:

- n : είναι ο αριθμός των δεδομένων (συνήθως αναφέρεται στις σειρές του πίνακα δεδομένων).
- d : είναι οι διαστάσεις των δεδομένων (συνήθως αναφέρεται στις στήλες του πίνακα δεδομένων).
- k : είναι ο αριθμός των κέντρων που ορίζεται από τον χρήστη.
- i : είναι ο αριθμός των επαναλήψεων.

Γίνεται αντιληπτό ότι η εύρεση της βέλτιστης λύσης αποτελεί μια μη-τετριμμένη διαδικασία κατατάσσοντας τον αλγόριθμο σε ένα *NP-hard* πρόβλημα. Επίσης, η ευαισθησία του στην επιλογή των αρχικών κεντροειδών συναρτήσκει του γεγονότος ότι συχνά οδηγείται από την υπόθεση σφαιρικής δομής των συστάδων, προσθέτει επιπρόσθετα μελανά σημεία στα αποτελέσματα που εκλαμβάνονται. Αυτό που αποτελεί το σημαντικότερο τροχοπέδη, αλλά σε άλλες περιπτώσεις και τρομερό πλεονέκτημα, του *k-means* είναι η επιβολή *αυστηρής ανάθεσης* (*Hard-Assignment*), ή, όπως

²⁴ Η πραγματική υπολογιστική πολυπλοκότητα του *k-means* μπορεί να θεωρηθεί ως ένα πρόβλημα που εξαρτάται σε πολύ μεγάλο βαθμό από την προσέγγιση που επιλέγει ο χρήστης. Για παράδειγμα, ένα επιπρόσθετο υπολογιστικό φορτίο θα ήταν να επιβληθεί στον αλγόριθμο να επαναλαμβάνει την διαδικασία της συσταδοποίησης χρησιμοποιώντας καινούργιες τοποθεσίες κέντρων κάθε φορά (*replicates*) [42].

θα εντοπιστεί και στην μετέπειτα ανάλυση, της *αυστηρής κωδικοποίησης (Hard-Coding)* των δεδομένων στον χώρο. Οι προαναφερθείσες έννοιες υπογραμμίζουν την διαδικασία εναπόθεσης των εξεταζόμενων σημείων του πολυδιάστατου χώρου σε ένα και μόνο κέντρο του διαμελισμένου χώρου χαρακτηριστικών που συνιστά η ανάθεση του αλγορίθμου.

Σε αυτό το σημείο έρχονται στο προσκήνιο τα Μοντέλα Γκαουσιανής Μίξης (GMM), τα οποία προσφέρουν μια πιο ευέλικτη και ισχυρή προσέγγιση στην διαδικασία της συσταδοποίησης.

2.3.1 Αλγόριθμοι Συσταδοποίησης: Γκαουσιανά Μοντέλα Μίξης (GMM)

Το σημαντικό εμπόδιο που παρουσιάζει ο *k-means* στον υπολογιστικό χρόνο που απαιτείται εξαιτίας της πολυπλοκότητάς του, αποτελεί ένα εφιαλτήριο για την προσπάθεια εύρεσης ενός εναλλακτικού τρόπου στην προσέγγιση της επεξεργασίας των δεδομένων. Τίθεται λοιπόν το ερώτημα: «*Θα μπορούσε κανείς να χρησιμοποιήσει κάποιου είδους αντιστοιχίας ανάμεσα στα πολυδιάστατα δεδομένα και μια πολύ καλή εκτίμηση αυτών για να μειώσει τον υπολογιστικό χρόνο;*» Η απάντηση σε αυτό το ερώτημα δίνεται από τα Γκαουσιανά Μοντέλα Μίξης, ένα σύνολο ισχυρών πιθανοτικών μοντέλων που χρησιμοποιούνται στην ανάλυση των δεδομένων αλλά και την Μηχανική Μάθηση για τον εντοπισμό κρυμμένων μοτίβων και δομών στα δεδομένα.

2.3.2 Γκαουσιανές Κατανομές (Gaussian Distributions)

Δεν ξεφεύγει από την πραγματικότητα να ειπωθεί ότι οι Γκαουσιανές κατανομές (ή αλλιώς *Κανονικές κατανομές*) είναι από τις πιο μελετημένες και πιο συχνά αξιοποιήσιμες συνεχείς πιθανοτικές κατανομές πραγματικών τυχαίων μεταβλητών. Η σημαντικότητά τους οφείλεται σε ένα μέρος από το *Θεώρημα Κεντρικού Ορίου (Central Limit Theorem)*, το οποίο δηλώνει πως κάτω υπό συγκεκριμένες συνθήκες, ο μέσος όρος πολλών δειγμάτων μιας τυχαίας μεταβλητής με πεπερασμένη μέση τιμή και διακύμανση είναι κι αυτός μιας τυχαία μεταβλητή του οποίου η κατανομή συγκλίνει σε μια κανονική κατανομή όσο το πλήθος των δειγμάτων αυξάνεται [46].

Η Κανονική κατανομή εμφανίζει δυο διακεκριμένες μορφές που συμβαδίζουν με την διαστατικότητα των μεταβλητών που περιγράφουν. Για μια *μονοσήμαντη τυχαία μεταβλητή (univariate random variable)*, η Γκαουσιανή κατανομή έχει **πυκνότητα** που δίνεται από τον τύπο:

$$p(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right) \quad (36)$$

Όπου:

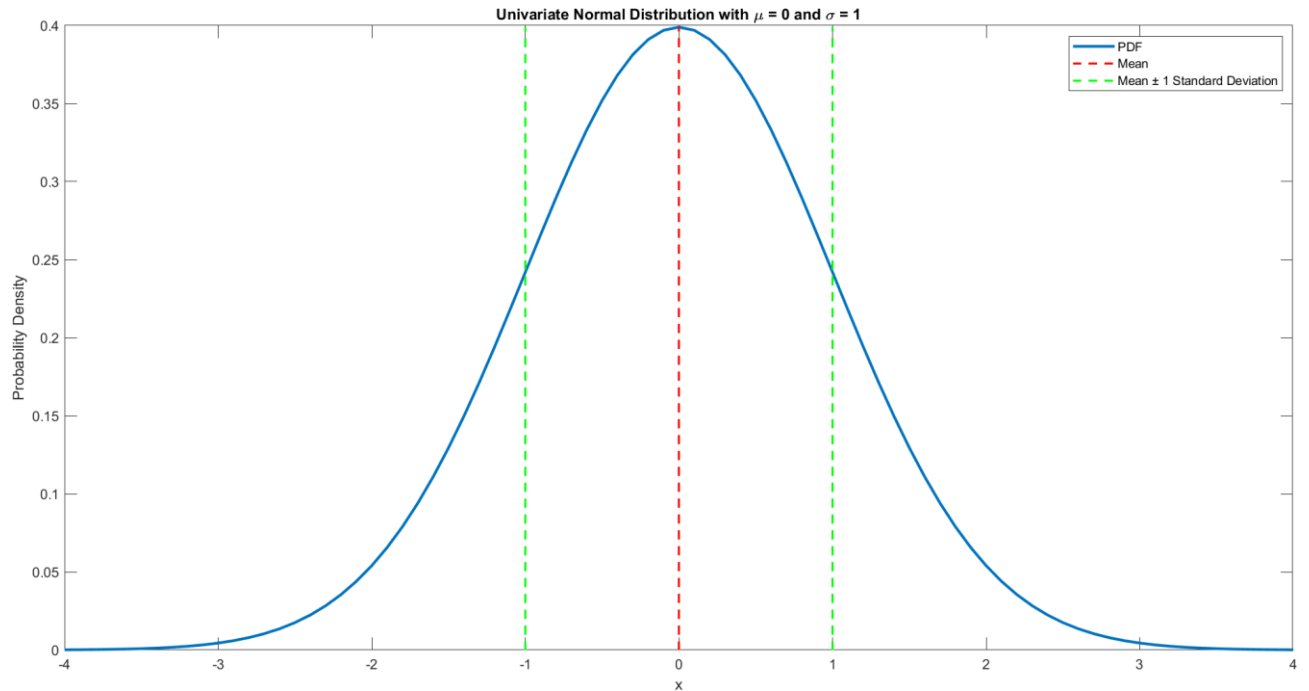
- μ : η μέση τιμή (*mean*).
- σ : η διακύμανση (*variance*).

Αντίστοιχα μια *πολυσήμαντη τυχαία μεταβλητή (multivariate random variable)* χαρακτηρίζεται πλήρως από το διάνυσμα της μέσης τιμής, μ , και πλέον έναν πίνακα συνδιακύμανσης, Σ , του οποίου

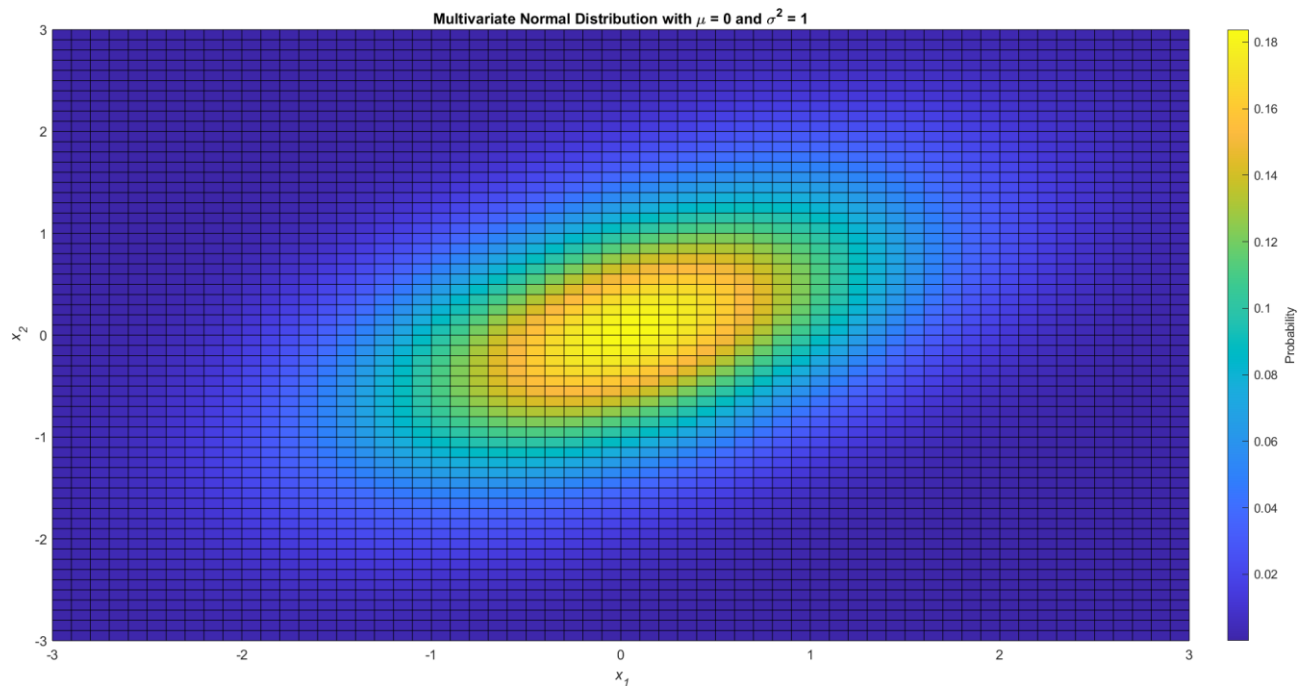
η κύρια διαγώνιος περιγράφει την διακύμανση. Η σημειογραφία που χρησιμοποιείται για την προηγούμενη είναι [36]:

$$p(x|\mu, \Sigma) = (2\pi)^{-\frac{D}{2}} |\Sigma|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(x - \mu)^T \Sigma^{-1}(x - \mu)\right) \quad (37)$$

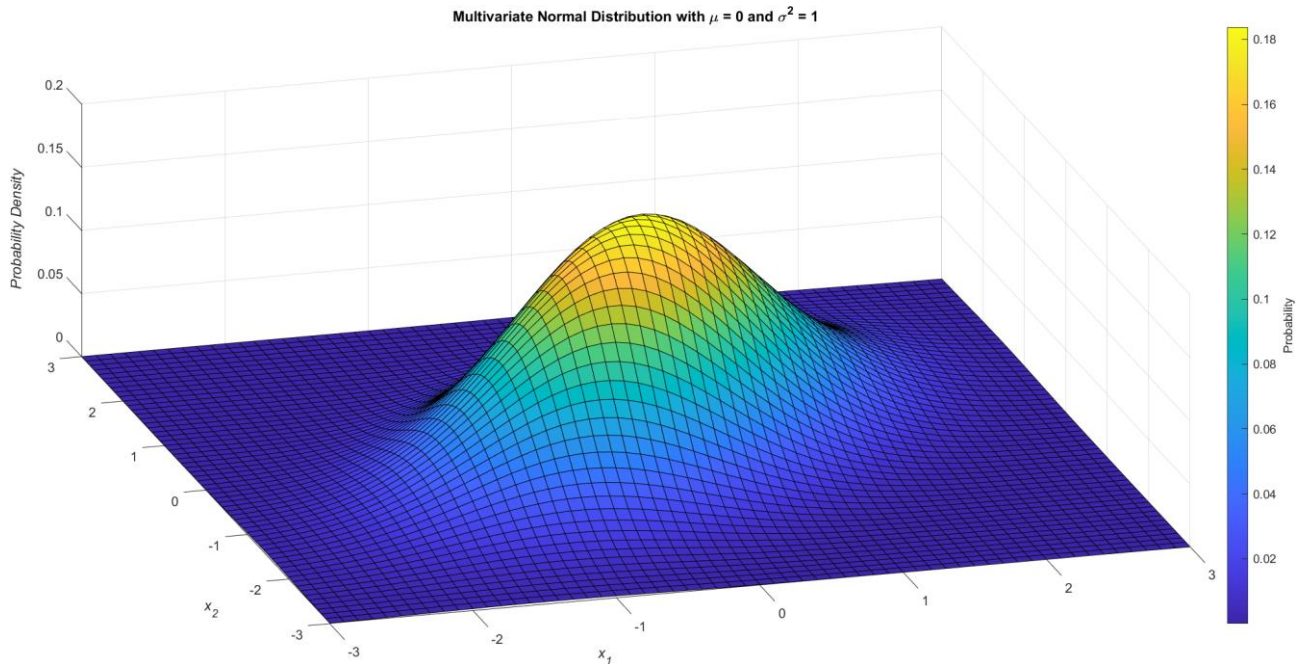
Όπου το $x \in \mathbb{R}$.



Εικόνα 23: Απεικόνιση της γκαουσιανής κατανομής πυκνότητας πιθανότητας μιας μονοσήμαντης τυχαίας μεταβλητής.



Εικόνα 24: Δισδιάστατη αναπαράσταση της γκαουσιανής κατανομής πυκνότητας πιθανότητας πολυσήμαντων τυχαίων (εδώ δύο) μεταβλητών.



Εικόνα 25: Τρισδιάστατη αναπαράσταση της γκαουσιανής κατανομής πυκνότητας πιθανότητας της [Εικόνας 24](#).

Η στατιστική φύση των Γκαουσιανών κατανομών, και πιο συγκεκριμένα η **εκτίμηση της πυκνότητας**, αποτελεί μια λύση στην άρση της αναπαράστασης των δεδομένων μέσω των ίδιων των δεδομένων, καθώς η δειγματοληψία από τις κατανομές αποτελεί έναν συμπαγή και αναλογικό τρόπο λήψης της πληροφορίας κατευθείαν από τα δεδομένα με την εξαίρεση ότι πλέον το μόνο που επιζητείται είναι να βρεθούν η μέση τιμή και η διακύμανση, ή όπως θα είναι από εδώ και πέρα πλέον γνωστές ως **παράμετροι**, μειώνοντας κατά πολύ τον υπολογιστικό χρόνο.

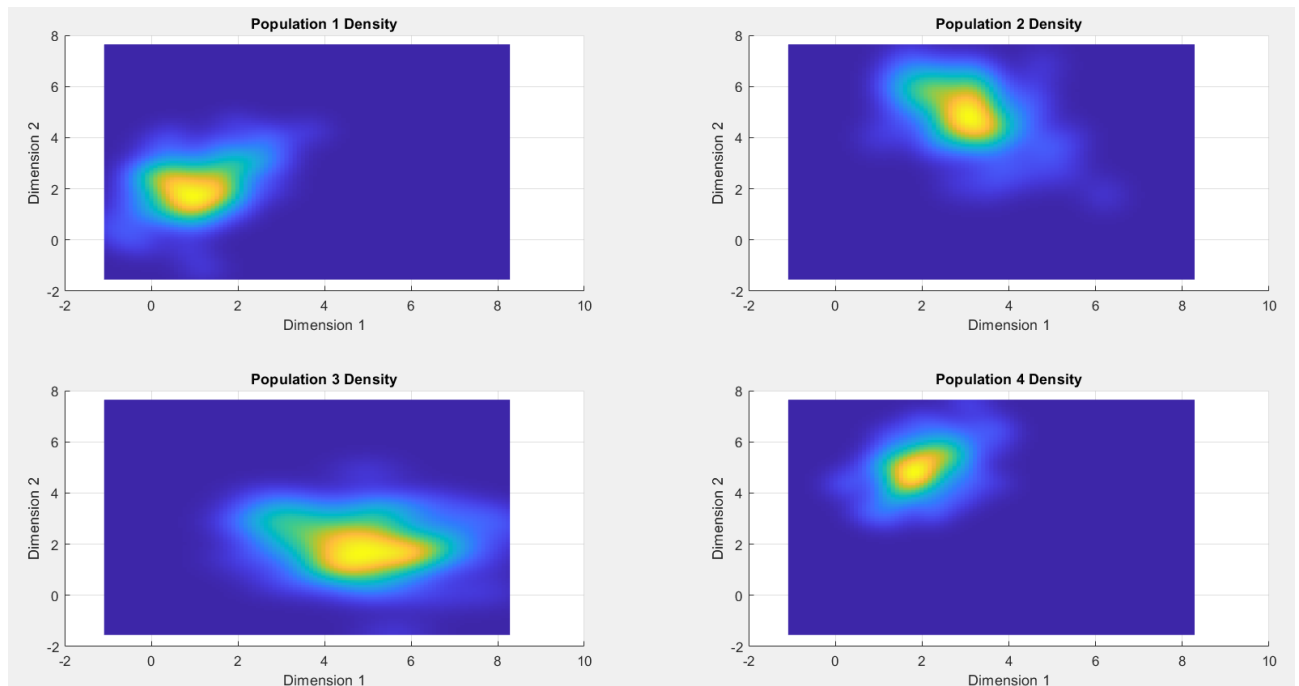
2.3.3 Γκαουσιανά Μοντέλα Μίξης

Η χρησιμότητα της στατιστικής φύσης των κανονικών κατανομών αποδείχθηκε ότι είναι μείζονος σημασίας σε μια αλλαγή της οπτικής της επεξεργασίας των δεδομένων, αλλά δεν παύει να εμφανίζει κάποια μειονεκτήματα. Ας επιτραπεί να μελετηθεί το παρακάτω παράδειγμα, το οποίο απεικονίζει ένα σύνολο από τέσσερις πληθυσμούς κατανεμημένους στον d -διάστατο χώρο περιγραφής των δεδομένων τους, όπως αυτοί απεικονίζονται στην [Εικόνα 26](#). Με μια πρώτη ματιά είναι εύκολο να εντοπιστεί ότι ο κάθε πληθυσμός φέρει διαφορετικό χρωματικό κώδικα, με την πιο σημαντική βέβαια παρατήρηση να είναι ότι οι πληθυσμοί εμφανίζουν επικαλύψεις σε κάποια σημεία του χώρου.

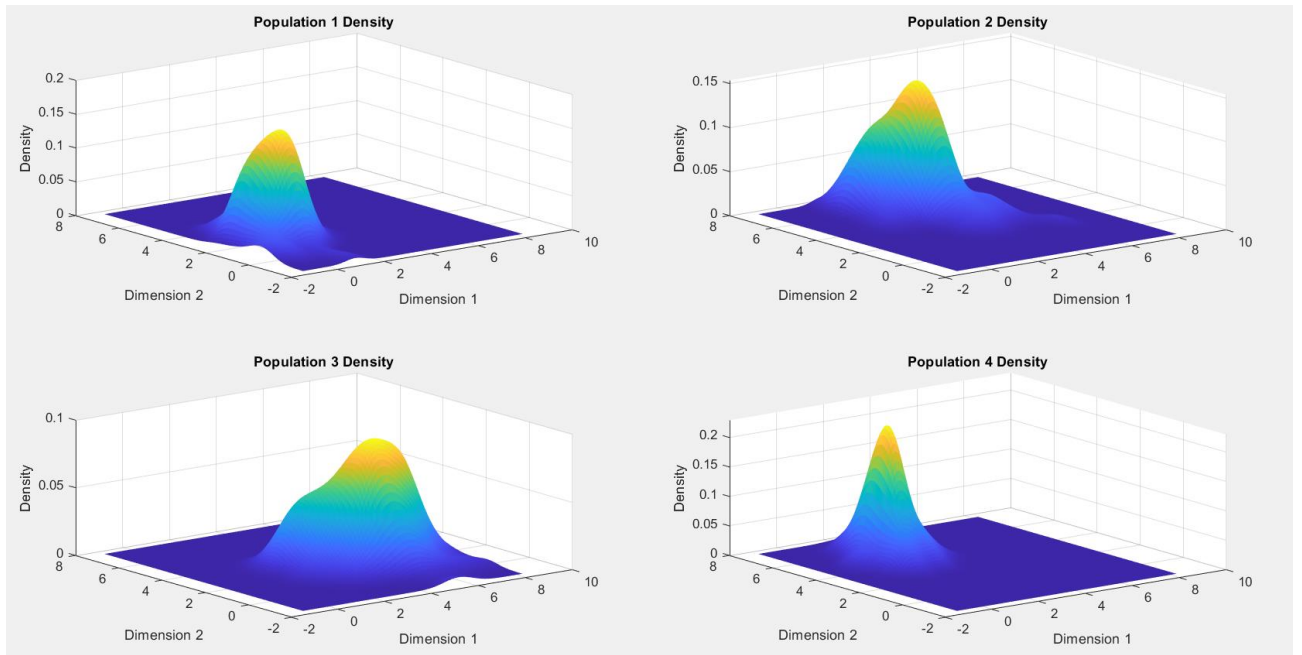


Εικόνα 26: Απεικόνιση των τεσσάρων πληθυσμών προς εξέταση.

Το τροχοπέδη λοιπόν σε αυτές τις καταστάσεις είναι ότι μια και μόνο Κανονική κατανομή δεν είναι αρκετή ώστε να μπορέσει να περιγράψει αποδοτικά όλους τους παρόντες πληθυσμούς, το οποίο μπορεί εύκολα να εξαχθεί κι από τον υπολογισμό των πυκνοτήτων των πληθυσμών που εντοπίζονται σε διαφορετικά σημεία του χώρου, όπως φαίνεται και στην *Εικόνα 27*.



Εικόνα 27: Δισδιάστατη αναπαράσταση της πυκνότητας των τεσσάρων πληθυσμών.



Εικόνα 28: Τρισδιάστατη αναπαράσταση της πυκνότητας των τεσσάρων πληθυσμών.

Μια πιο περιγραφική οικογένεια κατανομών είναι τα *Μοντέλα Μίξεων*, τα οποία προσπαθούν να περιγράψουν μια συνολική κατανομή δεδομένων μέσω ενός συνδυασμού k -Κανονικών κατανομών. Η μαθηματική σημειογραφία που περιγράφει το άνωθεν δίνεται από τους τύπους [36]:

$$p(x) = \sum_{k=1}^K \pi_k p_k(x) \quad (38)$$

Με:

$$0 \leq \pi_k \leq 1$$

$$\sum_{k=1}^K \pi_k = 1$$

Όπου p_k είναι η κατανομή που χρησιμοποιείται για να περιγραφεί ο πληθυσμός και π_k είναι τα βάρη μίξης. Φέρνοντας λοιπόν τα πράγματα στο μέτρο των Γκαουσιανών κατανομών, ο τύπος (38) μπορεί να γραφεί ως [36]:

$$p(x|\theta) = \sum_{k=1}^K \pi_k \mathcal{N}(x|\mu_k, \theta_k)(x) \quad (39)$$

Με:

$$0 \leq \pi_k \leq 1$$

$$\sum_{k=1}^K \pi_k = 1$$

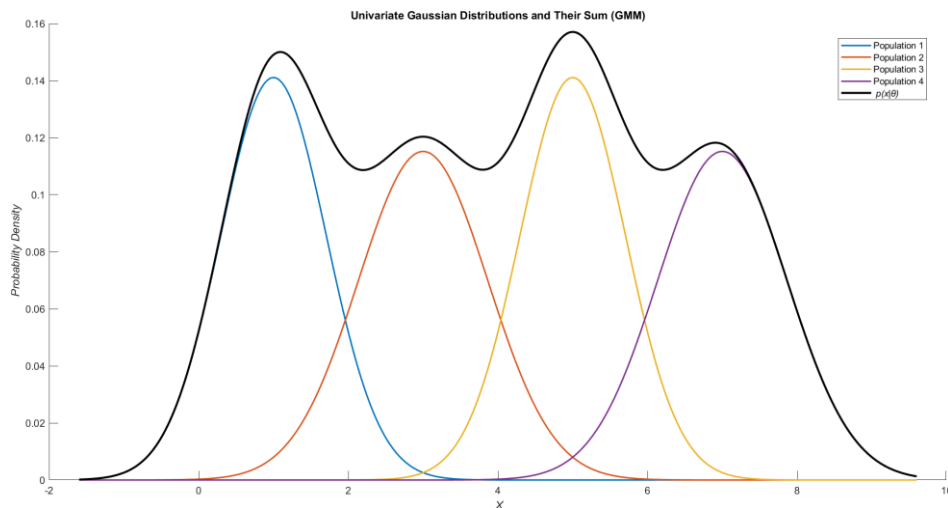
Όπου $\theta := \{\mu_k, \Sigma_k, \pi_k: k = 1, 2, \dots, K\}$ είναι το σύνολο των παραμέτρων του μοντέλου. Τι σημαίνει όμως αυτό; Στην πράξη, δίνεται η δυνατότητα περιγραφής των δεδομένων ως ένας γραμμικός

συνδυασμός k διαφορετικών Κανονικών κατανομών. Κι όντως, αν γνωρίζουμε ότι οι τέσσερις πληθυσμοί έχουν μέσες τιμές, διακυμάνσεις και ισόποσα αρχικοποιημένα βάρη, όπως αυτά παρατίθενται στον Πίνακα 2, τότε το μείγμα γκαουσιανών αποτελείται από έναν κυρτό συνδυασμό²⁵ των εκάστοτε μονοσήμαντων Κανονικών κατανομών αυξάνοντας την περιγραφικότητα που θα μπορούσε να είχε ο κάθε ξεχωριστός παράγοντας που το αποτελεί, όπως φαίνεται και στην *Εικόνα 29* οδηγώντας στην αποτύπωση της συνολικής πιθανότητας ως [36]:

$$p(x|\theta) = 0.25\mathcal{N}(x|1,0.5) + 0.25\mathcal{N}(x|3,0.75) + 0.25\mathcal{N}(x|5,0.5) + 0.25\mathcal{N}(x|7,0.75)$$

Πίνακας 2: Ενδεικτικές τιμές των παραμέτρων των τεσσάρων πληθυσμών.

Πληθυσμός	Μέση τιμή	Διακύμανση	Βάρος μίξης
1	1	0.5	0.25
2	3	0.75	0.25
3	5	0.5	0.25
4	7	0.75	0.25



Εικόνα 29: Αποτύπωση του συνδυασμού των τεσσάρων μονοσήμαντων γκαουσιανών κατανομών σε ένα συνολικό μείγμα κατανομών.

Το αποτέλεσμα που είναι εμφανές στην *Εικόνα 29* αποτελεί μια ιδανική περίπτωση καθώς όλα τα δεδομένα του προβλήματος αποτελούν ένα επίπλαστο παράδειγμα που δημιουργήθηκε για την κατανόηση του πως λειτουργούν χονδροειδώς τα GMM. Όμως τα πράγματα στον πραγματικό κόσμο δεν είναι ιδανικά, καθώς δεν μπορεί να είναι *a priori* γνωστό ποια είναι εκείνη η κατανομή που περιγράφει τα δεδομένα σε έναν αρκετά καλό βαθμό. Στόχος λοιπόν από το παρόν σημείο και μέχρι το τέλος αυτού του κεφαλαίου είναι να υπογραμμιστεί η σειρά των διαδικασιών που λαμβάνουν μέρος ώστε τα μοντέλα των Γκαουσιανών Μίξεων να μπορούν να εξάγουν τις υποκείμενες κατανομές

²⁵ Με την χρήση της λέξης κυρτός συνδυασμός, περιγράφεται το φαινόμενο αποτύπωσης ενός τυπικού σταθμισμένου αθροίσματος όπου τα βάρη είναι μη αρνητικά βάρη και αθροίζονται στην μονάδα [47].

που περιγράφουν τα δεδομένα με όσο το δυνατόν καλύτερο τρόπο μέσω μιας επαναληπτικής διαδικασίας εκμάθησης των παραμέτρων τους.

2.3.4 Εκτίμηση Μέγιστης Πιθανοφάνειας (*Maximum Likelihood Estimation*)

Το πρώτο βήμα ξεκινά με τον σχηματισμό της συνάρτησης που θα επιτρέψει την προσέγγιση των δεδομένων. Εφόσον τα GMM λειτουργούν βάσει τριών παραμέτρων, είναι φυσικό αλλά και επόμενο η συνάρτηση αυτή να εκφράζεται συναρτήσει των θ . Ο κορμός στο πρόβλημα της Εκτίμησης της Μέγιστης Πιθανοφάνειας, είναι η προσπάθεια ελαχιστοποίησης του αρνητικού λογαρίθμου της πιθανοφάνειας. Για δεδομένα που αναπαρίστανται από μια τυχαία μεταβλητή, x , και για μια οικογένεια πυκνοτήτων πιθανοτήτων $p(x|\theta)$ που παραμετροποιούνται από το θ , ο αρνητικός λογάριθμος της πιθανοφάνειας δίνεται από τον τύπο [36]:

$$\mathcal{L}_x(\theta) = -\log p(x|\theta) \quad (40)$$

Η σημειογραφία της σχέσης (40) δίνει έμφαση στο γεγονός ότι η παράμετρος θ μεταβάλλεται και τα δεδομένα x παραμένουν σταθερά. Σε μια ρύθμιση επιβλεπόμενης μάθησης λοιπόν, όπου λαμβάνονται ζευγάρια δεδομένων και ετικετών ως $(x_1, y_1), \dots (x_N, y_N)$ με το $x_n \in \mathbb{R}^D$ και το $y_n \in \mathbb{R}$, άμεσα εξάγεται ότι το x_n είναι ένας πίνακας (π.χ. ένας πίνακας χαρακτηριστικών *SIFT*), ενώ το y_n είναι ένα διάνυσμα ετικετών με τον απώτερο σκοπό να είναι η διάρθρωση ενός πιθανοτικού παραγωγικού μοντέλου κατανομών. Στο προηγούμενο εγχείρημα απόδοσης της μορφής των δεδομένων πρέπει να γίνει μια διευκρίνιση για την φύση αυτών. Γενικά σε τέτοιες περιπτώσεις είναι επικρατής η υπόθεση ότι το σύνολο των παραδειγμάτων είναι *ανεξάρτητο και πανομοιότυπα κατανεμημένο* (*independent and identically distributed – i.i.d.*), κάτι που μεταφράζεται ως ότι:

1. Κάθε όρος του γινομένου προέρχεται από την ίδια κατανομή και ότι όλοι τους μοιράζονται τις ίδιες παραμέτρους.
2. Ότι η πιθανοφάνεια που περιλαμβάνει όλο το σύνολο δεδομένων, παραγοντοποιείται σε ένα γινόμενο της πιθανοφάνειας του εκάστοτε παραδείγματος ως:

$$p(\mathcal{Y}|\mathcal{X}, \theta) = \prod_{n=1}^N p(y_n|x_n, \theta)$$

Με βάση τα τελευταία σχόλια, η συνάρτηση της αρνητικής λογαριθμικής πιθανοφάνειας μπορεί να προσεγγιστεί μέσω μιας υπολογιστικά αποδοτικότερης υλοποίησής της ως:

$$\mathcal{L}(\theta) = -\log(p(\mathcal{Y}|\mathcal{X}, \theta)) = -\sum_{n=1}^N \log(p(y_n|x_n, \theta)) \quad (41)$$

Ως μια τελευταία σημείωση, αξίζει να σημειωθεί ότι και σε αυτήν την περίπτωση το αρνητικό σύμβολο είναι μια σύμβαση που προέρχεται από το πεδίο της βελτιστοποίησης ώστε να ανάγει το πρόβλημα σε κυρτή μορφή καθώς κι ότι η πρακτικότητα της μείωσης της λογαριθμικής πιθανοφάνειας έγκειται στην βλέψη του φαινομένου ως την αποτύπωση της μείωσης του σφάλματος της συνάρτησης με το πέρασ των βημάτων εκπαίδευσης. **Η λήψη λοιπόν χαμηλών τιμών στην αρνητική λογαριθμική πιθανοφάνεια ισοδυναμεί με μείωση του σφάλματος της αντικειμενικής συνάρτησης προς ελαχιστοποίηση.** Κλείνοντας, είναι πρόπον να σημειωθεί ότι η εκτίμηση της μέγιστης Πιθανοφάνειας μπορεί να υποφέρει από υπερπροσαρμογή (*overfitting*) [48].

2.3.5 Εκμάθηση παραμέτρων μέσω της Εκτίμησης Μέγιστης Πιθανοφάνειας

Όπως έχει γίνει μέχρι τώρα γνωστό, προκειμένου να μπορέσει να εντοπιστεί η ελάχιστη τιμή που μπορεί να λάβει μια αντικειμενική συνάρτηση, θα πρέπει η κλίση συναρτήσε των παραμέτρων της να ισούται με μηδέν. Η ίδια μεθοδολογία εφαρμόζεται και στην προκείμενη περίπτωση όπου, με μια αναδιάταξη των όρων της σχέσης (41), ο τύπος της αρνητικής λογαριθμικής πιθανοφάνειας γράφεται ως:

$$\mathcal{L} = \sum_{n=1}^N \log \sum_{k=1}^K \pi_{k\mathcal{N}}(x_n | \mu_k, \Sigma_k) \quad (42)$$

Οι απαιτήσεις λοιπόν για την εύρεση του τοπικού ελαχίστου στην λογαριθμική πιθανοφάνεια είναι:

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial \mu_k} = 0^T &\Leftrightarrow \sum_{n=1}^N \frac{\partial \log p(x_n | \theta)}{\partial \mu_k} = 0^T \\ \frac{\partial \mathcal{L}}{\partial \Sigma_k} = 0^T &\Leftrightarrow \sum_{n=1}^N \frac{\partial \log p(x_n | \theta)}{\partial \Sigma_k} = 0, \\ \frac{\partial \mathcal{L}}{\partial \pi_k} = 0^T &\Leftrightarrow \sum_{n=1}^N \frac{\partial \log p(x_n | \theta)}{\partial \pi_k} = 0 \end{aligned}$$

Ενώ η προσπάθεια εύρεσης μηδενισμού των μερικών παραγώγων των παραμέτρων φαίνεται πως είναι μια απλή υπόθεση, στην πραγματικότητα θα αποδειχθεί ότι η διαδικασία δεν είναι τελικά τετριμμένη. Και για τις τρεις συνθήκες, απαιτείται οι μερικές παράγωγοι να έχουν την μορφή:

$$\frac{\partial \log p(x_n | \theta)}{\partial \theta} = \frac{1}{p(x_n | \theta)} \frac{\partial p(x_n | \theta)}{\partial \theta}$$

Συμπληρωματικά, αξίζει να μελετηθεί ένα ακόμη πολύ σημαντικό στατιστικό μέγεθος: τα **βάρη ευθυνών**.

2.3.6 Βάρη Ευθυνών (*Responsibilities*)

Προς ευκολία περάτωσης της ερευνητικής διαδικασίας που ακολουθείται σε όλο το κεφάλαιο, θα οριστεί μια νέα ποσότητα η οποία ονομάζεται *βάρος ευθύνης* και αναφέρεται στον k -οστό συντελεστή μίξης (δηλαδή της τριάδας: *μέση τιμή, πίνακας συνδιακύμανσης και βάρη συμμετοχής*) για το n -οστό δεδομένο ως:

$$r_{nk} = \frac{\pi_k \mathcal{N}(x_n | \mu_k, \Sigma_k)}{\sum_{j=1}^K \pi_j \mathcal{N}(x_n | \mu_j, \Sigma_j)} \quad (43)$$

με το $r_{nk} \in \mathbb{R}^{N \times K}$, όπου:

- Με N αντιπροσωπεύεται το πλήθος των δεδομένων (συνήθως είναι οι σειρές του πίνακα που περιγράφει τα δεδομένα) ενώ
- Το K αντιπροσωπεύει τον αριθμό των κέντρων που ορίζει ο χρήστης.

Προσοχή εφίσταται στο γεγονός ότι οι παράμετροι των GMM, ειδικά στην περίπτωση διψήφιου αριθμού κέντρων, είναι πίνακες και όχι διανύσματα²⁶, όπου η κάθε στήλη τους εκφράζει το ποσοστό συμμετοχής του κάθε κέντρου στο συνολικό μείγμα! Επιπροσθέτως, το πιθανοτικό κανονικοποιημένο διάνυσμα $r_n := [r_{n1}, \dots, r_{nK}]^T$ θα πρέπει να ικανοποιεί τους περιορισμούς ότι:

- $\sum_k r_{nk} = 1$
- $r_{nk} \geq 0$

Η σχέση ανάμεσα στο εκάστοτε βάρος ευθύνης του k -οστού συντελεστή μίξης και την πιθανοφάνεια των συντελεστών μίξης των δεδομένων είναι άμεση, καθώς αντικατοπτρίζει την πιθανότητα ότι το x_n δείγμα έχει παραχθεί από την k -οστή συνιστώσα μίξης, σχηματίζοντας την έννοια της «ελαφριάς ανάθεσης» και οδηγώντας στο συμπέρασμα ότι **υψηλότερο βάρος ευθύνης αποδίδεται σε εκείνο το σημείο του χώρου (δεδομένο) που θα μπορούσε να είναι ένα πιθανό δείγμα από αυτή τη μίξη** [36].

2.3.7 Ενημέρωση των παραμέτρων

Στην προηγούμενη υποενότητα έγινε αναφορά στο γεγονός ότι η διαδικασία εύρεσης του μηδενισμού της κλίσης των παραμέτρων είναι μια μη-τετριμμένη διαδικασία. Στο σημείο αυτό επιβεβαιώνονται οι παραπάνω ισχυρισμοί, παραθέτοντας τρία βασικά θεωρήματα που αποτυπώνουν την διαδικασία ανανέωσης των παραμέτρων του GMM.

Ξεκινώντας με τις μέσες τιμές, αποτυπώνεται ότι:

²⁶ Κανείς θα μπορούσε να υιοθετήσει μια διαφορετική οπτική στην οποία μπορεί να δει ότι το κάθε κέντρο παράγει ένα διάνυσμα τιμών, το οποίο στην συνέχεια συνενώνεται οριζόντια με το προηγούμενο. Το αποτέλεσμα καταλήγει να είναι πάλι ένας πίνακας.

Ορισμός 11: Θεώρημα ενημέρωσης των μέσων τιμών του GMM

Η ενημέρωση των παραμέτρων της μέσης τιμής $\mu_k, k = 1, \dots, K$ των GMM δίνεται από την σχέση [36]:

$$\mu_k^{new} = \frac{\sum_{n=1}^N r_{nk} x_n}{\sum_{n=1}^N r_{nk}} \quad (44)$$

Η εμφάνιση και μόνο των βαρών ευθύνης που φέρουν μαζί τους την πληροφορία όλων των παραμέτρων των μοντέλων, δείχνει πως δεν είναι δυνατή η λήψη μιας λύσης κλειστής μορφής για όλα τα μ_k απευθείας από το θεώρημα [36]. Συνεχίζοντας με τους πίνακες συνδιακύμανσης, αποτυπώνεται ότι:

Ορισμός 12: Θεώρημα ενημέρωσης των πινάκων συνδιακύμανσης του GMM

Η ενημέρωση των παραμέτρων των πινάκων συνδιακύμανσης $\Sigma_k, k = 1, \dots, K$ των GMM δίνεται από την σχέση [36]:

$$\Sigma_k^{new} = \frac{1}{N_k} \sum_{n=1}^N r_{nk} (x_n - \mu_k)(x_n - \mu_k)^T \quad (45)$$

Το ίδιο πρόβλημα που υπήρχε στην ανανέωση των μέσων τιμών φαίνεται πως εμφανίζεται και στους πίνακες συνδιακύμανσης. Τέλος, για τα βάρη μίξης αποτυπώνεται ότι:

Ορισμός 13: Θεώρημα ενημέρωσης των βαρών μίξης

Η ενημέρωση των παραμέτρων των βαρών μίξης των GMM δίνεται από την σχέση [36]:

$$\pi_k^{new} = \frac{N_k}{N}, \quad k = 1, \dots, K \quad (46)$$

Ενώ με μια πρώτη ματιά φαίνεται πως η ανανέωση των βαρών μίξης δεν επηρεάζεται από τα βάρη ευθυνών, αρκεί κανείς να θυμηθεί ότι $N_k = \sum_{i=1}^N r_{nk}$, οδηγώντας δυστυχώς πάλι στο ίδιο πρόβλημα. Οι αποδείξεις για το πως οι τύποι (44) (45) και (46) οδηγούνται σε ανοικτής μορφής μπορεί να εντοπιστεί στο Κεφάλαιο 12 του βιβλίου "Mathematics for Machine Learning".

2.3.8 Ο αλγόριθμος Μεγιστοποίησης Προσδοκιών (Expectation Maximization)

Η παραπάνω υποενοότητα εμπεριστατώνει ότι η απομονωμένη προσπάθεια εύρεσης του μηδενισμού των κλίσεων των παραμέτρων είναι ατελέσφορη. Λύση σε αυτό το τροχοπέδη αποτελεί η υλοποίηση του αλγορίθμου της Μεγιστοποίησης Προσδοκιών (Expectation Maximization), ο οποίος μέσω του επαναληπτικού σχηματικού εκτέλεσής του άρει την ανοικτή μορφή των εξισώσεων ανανέωσης των παραμέτρων.

Τα βασικά στάδια του αλγορίθμου είναι δυο.

- Στο πρώτο, το οποίο είναι το στάδιο της υλοποίησης των προσδοκιών (*E-step*), υπολογίζονται τα βάρη ευθυνών r_{nk}
- Στο δεύτερο, το οποίο αφορά την μεγιστοποίηση των προσδοκιών (*M-step*), ανανεώνονται τα βάρη ευθυνών για να επανεκτιμηθούν οι παράμετροι θ .

Ο λεπτομερής ψευδοκώδικας που περιγράφει την διαδικασία του EM μόνο για διαγώνιους πίνακες συνδιακύμανσης, παρατίθεται παρακάτω:

Algorithm 1 Gaussian Mixture Model (only for Diagonal Covariances)

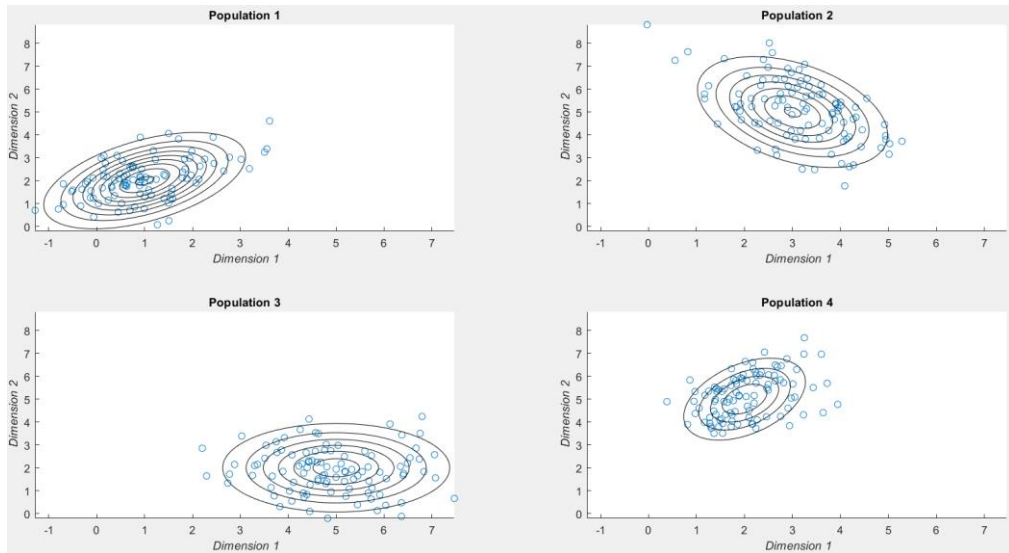
```

1: Input: data, numClusters
2: Output: GMM structure containing weights ( $\pi_k$ ), means ( $\mu_k$ ), variances ( $\Sigma_k$ ), logLikelihood
3: Initialize logLikelihoodOld, bestLogLikelihood, bestLogLikelihoodOld
4: Add a constant term as a bias to avoid numerical overflow/underflow
5: Initialize  $\pi_k$ ,  $\mu_k$  using kmeans and  $\Sigma_k$ 
6: for each iteration do
7:   E-step:
8:   for each cluster j do
9:     Compute log prior
10:    Compute logDetSigma
11:    Compute Log Likelihood for cluster j
12:    Update Log Likelihood with log prior and constTerm
13:   end for
14:   Compute MaxLogLikelihood, Responsibilities, Density, Logpdf
15:   Compute total logLikelihood
16:   Normalize Responsibilities
17:   M-step:
18:   for each cluster j do
19:     Compute Responsibilities.j
20:     Update  $\mu_k$  for cluster j
21:     Compute the centered data
22:     Update  $\Sigma_k$  for cluster j
23:     Update  $\pi_k$  for cluster j
24:   end for
25:   Check for convergence
26:   if converged or last iteration then
27:     Break
28:   end if
29:   Update logLikelihoodOld
30: end for
31: Check for significant improvement in bestLogLikelihood
32: Set final  $\pi_k$ ,  $\mu_k$ ,  $\Sigma_k$  to best found
33: Create GMM structure with final parameters and metrics

```

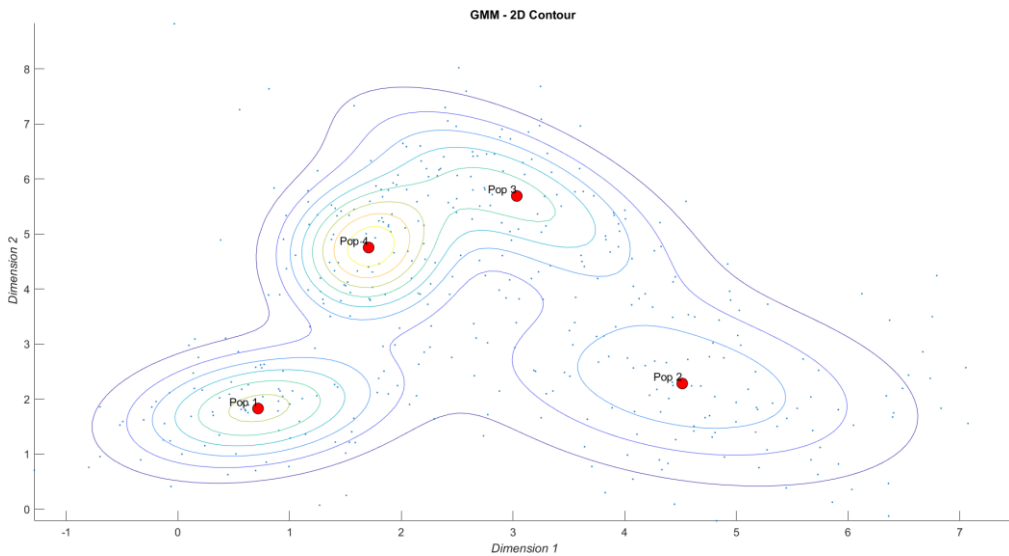
Αλγόριθμος 2: Αποτύπωση του EM αλγορίθμου μόνο για διαγώνιους πίνακες συνδιακύμανσης.

Πριν κλείσει η υποενότητα αυτή, για λόγους καλής πρακτικής θα ήταν πρόπον να γίνει μια επιστροφή στο αρχικό παράδειγμα που είχε τεθεί ώστε να αποτυπωθεί η καινοτομία που επιφέρουν τα GMM. Έχοντας ως οδηγό την *Εικόνα 26* όπου αναπαρίσταται ο αρχικός χώρος διάταξης των δεδομένων, με την εφαρμογή της συνάρτησης [ksdensity](#) υπολογίζεται μια αρχική εκτίμηση της πυκνότητας πιθανότητας των πληθυσμών αποδίδοντας την παρακάτω απεικόνιση:

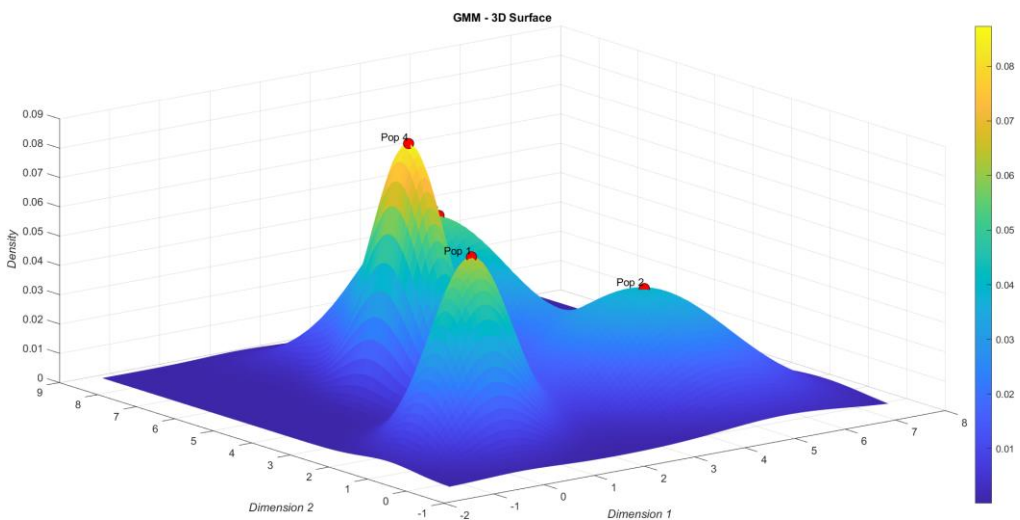


Εικόνα 30: Αποτύπωση των συστάδων όπως αυτές προκύπτουν από την χρήση της συνάρτησης *ksdensity*.

Εφαρμόζοντας τώρα τον EM στα δεδομένα για 10 διαφορετικές αρχικοποιήσεις του και για 100 επαναλήψεις, τα παραγόμενα αποτελέσματα είναι:

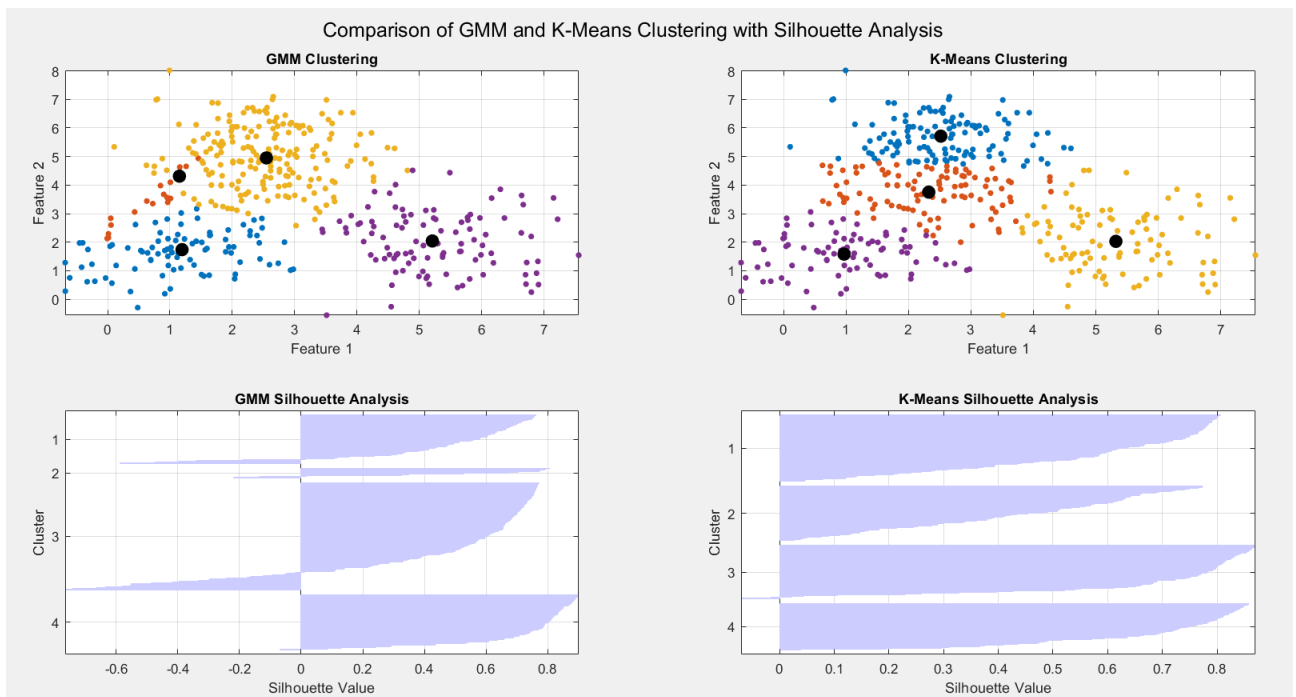


Εικόνα 31: Αποτύπωση της διδιάστατης κατανομής των συστάδων με βάση τα αποτελέσματα του GMM.



Εικόνα 32: Αποτύπωση της τριδιάστατης κατανομής των συστάδων με βάση τα αποτελέσματα του GMM.

Μέσω μια οπτικής επίσπευσης και σύγκρισης των *Εικόνων 26, 30 και 31* κανείς μπορεί να εκλάβει ότι το GMM δεν κατάφερε να ταξινομήσει επιτυχώς τα δεδομένα καθώς διαφαίνεται ότι ο πληθυσμός 2 με τον πληθυσμό 3 θα έπρεπε να έχουν διαφορετική θέση. Ο αναγνώστης καλείται να λάβει υπόψιν του ότι η αποτελεσματικότητα και η δυνατότητα ερμηνείας των αλγορίθμων ομαδοποίησης, όπως τα GMM, εξαρτώνται σε μεγάλο βαθμό από τη δομή και τη φύση των δεδομένων. Τα τυχαία παραγόμενα δεδομένα ενδέχεται να μην παρουσιάζουν αξιόλογες ομαδοποιήσεις, συγκριτικά με τα δομημένα δεδομένα όπως τα χαρακτηριστικά που εξάγονται από εικόνες με τη χρήση του SIFT καθώς τα τελευταία περιέχουν εγγενή μοτίβα που μπορούν να ομαδοποιηθούν αποτελεσματικά. Έτσι παρέχοντας την ελαστικότητα ότι στα τυχαία παραγόμενα δεδομένα ο κάθε χρήστης θα μπορούσε να υιοθετήσει διαφορετικό σύστημα για την ονομασία των συστάδων και επικεντρώνοντας το ενδιαφέρον απλά στον εντοπισμό τους, μπορεί να ειπωθεί με βεβαιότητα ότι τα GMM κατάφεραν να παράγουν αξιόλογα αν όχι ιδανικά αποτελέσματα για τα παρεχόμενα δεδομένα που τους δόθηκαν. Επιπρόσθετα, με την χρήση της μετρικής *silhouette* εφαρμόζεται μια υποτυπώδη σύγκριση ανάμεσα στον *k-means* και στα *GMM* αποτυπώνοντας στην πρώτη περίπτωση την αυστηρή ανάθεση (*hard-assignment*) και στην δεύτερη το χαρακτηριστικό της *ήπιας/ελαστικής* ανάθεσης (*soft-assignment*), δηλαδή του φαινομένου κατά το οποίο η ανάθεση στα κέντρα των πληθυσμών γίνεται με πιθανοτικό τρόπο.



Εικόνα 33: Αποτίπωση της ανάθεσης των δεδομένων στα κέντρα των πληθυσμών καθώς και η ανάλυση με χρήση της μετρικής *Silhouette* τόσο για τον *k-means* όσο και για τα *GMM*.

2.3.9 Ο αλγόριθμος Βηματικής Μεγιστοποίησης Προσδοκιών (*stepwise-Expectation Maximization*)

Η μοναδικότητα της λειτουργίας του EM είναι απaráμιλλης σημασίας στον κόσμο της συσταδοποίησης δεδομένων. Ακόμη όμως κι ένας τόσο σημαντικός αλγόριθμος υποφέρει από κάποια σοβαρά μειονεκτήματα με το σημαντικότερο αυτών να είναι η αργή σύγκλιση σε κάποιο αποτέλεσμα: ένα φαινόμενο που οφείλεται κατά κόρον στην τροφοδότηση όλου του εύρους των δεδομένων απευθείας²⁷, κάτι που σημαίνει ότι οι παράμετροι θ ενημερώνονται μόνο όταν έχουν προσπελαστεί όλα τα δεδομένα. Το πρόβλημα αυτό έρχεται να άρει μια ομάδα ερευνητών του πανεπιστημίου του Berkeley, ο *Percy Liang* και ο *Dan Klein*, οι οποίοι στην δημοσίευσή τους με τίτλο: «*Online EM for Unsupervised Models*» [49] παραθέτουν μια σημαντική οικογένεια παραλλαγών του EM υπογραμμίζοντας τα πλεονεκτήματα και τα μειονεκτήματα μεταξύ τους. Προτού παρατεθεί το σχηματικό που ακολουθήθηκε για την υλοποίηση μιας από τις παραλλαγές που προτείνουν, είναι πρέπον να οριστεί το τι σημαίνει *online* εκδοχή του EM αλγορίθμου.

Σύμφωνα με την δημοσίευση, προκειμένου να μπορέσει να ληφθεί μια *online* εκδοχή του EM, θα πρέπει να αποθηκευτεί ένα μόνο σύνολο επαρκών στατιστικών στοιχείων (δηλαδή των παραμέτρων θ) και η ανανέωσή τους να πραγματοποιηθεί εφόσον έχει επεξεργαστεί κάθε δείγμα. Έτσι για το *i*-οστό δεδομένο υπολογίζονται τα επαρκή στατιστικά στοιχεία, εδώ με την ονομασία s_i' . Με γνώμονα τα προηγούμενα, δημιουργείται το γόνιμο έδαφος για να παρατεθεί ο τρόπος λειτουργίας μιας παραλλαγής του EM αλγορίθμου που ονομάζεται *βηματικός-EM*. Ο υπο μελέτη αλγόριθμος έχει τις ρίζες του στην στοχαστική προσεγγιστική φύση των αλγορίθμων. Συγκρίνοντας την διαδικασία με αυτή του προκατόχου του που χρειάζεται να διατρέξει όλα τα δεδομένα και μετά να εφαρμόσει τις ενημερώσεις, ο *sEM* τμηματοποιεί τα δεδομένα σε επιμέρους κομμάτια τα οποία διατρέχονται με βάση ένα συγκεκριμένο βήμα που μπορεί να καθοριστεί από τον χρήστη σε πρωτύερο χρόνο, οδηγώντας στην αποτύπωση μιας οπτικής που θυμίζει την στοχαστική κλίση (*stochastic gradient*) αλλά στον χώρο των επαρκών στατιστικών μεγεθών. Πως ακριβώς όμως λειτουργεί όλο αυτό; Το μυστικό της λειτουργίας του κρύβεται πίσω από τις λέξεις «*κομμάτια δεδομένων*» και «*μέγεθος βήματος*».

Προσεγγίζοντας τα πράγματα από το τέλος προς την αρχή, πρώτο μέλημα είναι ο καθορισμός της ισχύς μείωσης του μεγέθους του βήματος (*stepsize reduction power*) που συμβολίζεται με το γράμμα α . Δοσμένου ενός βήματος η_k , η σχέση που συνδέει το βήμα με την δύναμη μείωσης είναι [49]:

$$\eta_k = (k + 2)^{-\alpha} \quad (47)$$

²⁷ Το φαινόμενο της τροφοδότησης των δεδομένων με αυτόν τον τρόπο φέρει την αγγλική σημειογραφία *batch*.

Χρησιμοποιώντας την βιβλιογραφία που αναφέρεται στις στοχαστικές προσεγγίσεις, οι δυο ικανές και αναγκαίες συνθήκες που πρέπει να ισχύουν προκειμένου να εξασφαλιστεί σύγκλιση σε ένα τοπικό βέλτιστο είναι [49]:

$$\sum_{k=0}^{\infty} \eta_k = \infty$$

και

$$\sum_{k=0}^{\infty} \eta_k^2 < \infty$$

Αν λοιπόν ισχύει η σχέση (47), τότε οποιαδήποτε τιμή λαμβάνεται στο εύρος $0.5 < \alpha \leq 1$ είναι έγκυρη, ορίζοντας ποιοτικά πως όσο πιο μικρή είναι η εκθετική δύναμη α , τόσο μεγαλύτερες είναι οι ενημερώσεις που εφαρμόζονται οδηγώντας όμως σε ταχύτερη απόσβεση (decay) των παλαιών επαρκών στατιστικών μεγεθών. Αυτό με τη σειρά του οδηγεί σε γρηγορότερη σύγκλιση με την παράλληλη παραγωγή αστάθειας [49].

Η εμφάνιση αστάθειας είναι κάτι το οποίο αποτελεί ιδιαίτερα ζημιογόνο για καταστάσεις καταβολής βελτιστοποίησης μιας ποσότητας. Οπότε γεννάται το ερώτημα: «*Τι θα μπορούσε να προστεθεί ώστε να αντικρούσει αυτό το εμπόδιο;*» Δοσμένου ότι η ισχύς μείωσης του μεγέθους του βήματος είναι παρούσα, το μόνο που θα μπορούσε να προσθέσει σταθερότητα στην συνολική διαδικασία θα ήταν ένας τρόπος μέσω του οποίου προσπελούνται τα δεδομένα. Αυτό λοιπόν μπορεί να επιτευχθεί με τον «τεμαχισμό» τους σε μικρές παρτίδες (*mini batch sizes*), τα οποία μπορούν να συμβολιστούν με την αγγλική σημειογραφία m . Ο διαχωρισμός των n δεδομένων σε m παρτίδες μικρότερου μήκους δίνει την δυνατότητα στον αλγόριθμο να συμπεριφέρεται σε κάθε «κομμάτι» δεδομένων ως ένα ξεχωριστό δεδομένο, όπου στο καθένα [49]:

1. Υπολογίζονται τα επαρκή στατιστικά μεγέθη $s_{i+1}', s_{i+2}', \dots, s_{i+m}'$.
2. Ανανεώνονται οι παράμετροι χρησιμοποιώντας τα $s_{i+1}', s_{i+2}', \dots, s_{i+m}'$.

Με αυτό τον τρόπο, όσο μεγαλύτερο είναι το m τόσο λιγότερο θα εφαρμόζονται ενημερώσεις αλλά τόσο πιο σταθερές θα είναι. Κανείς μπορεί να παρατηρήσει ότι αναλόγως την επιλογή του m ο sEM ακροβατεί ανάμεσα σε μια καθαρή online μορφή ($m=1$) και σε μια batch μορφή ($m=n$) [49].

Algorithm 1 Stochastic Expectation Maximization for GMM

```
1: Input: data, numClusters
2: Output: GMM parameters (NegLogLikelihood, Log_Likelihood,  $\mu$ ,  $\Sigma$ ,  $\pi$ )
3: Initialize parameters: MaxIterations, Tolerance, Alpha, BatchSize
4: Calculate numPoints, dimFeatures from data
5: Initialize oldLogLikelihood to  $-\infty$ 
6: Initialize weights, mus, Sigmas
7: Compute constant term for log-likelihood calculation
8: for each iteration  $i$  up to MaxIterations do
9:   Shuffle data to create random order
10:  Partition data into batches based on BatchSize
11:  Shuffle batches to randomize order
12:  for each batch  $j$  do
13:    Calculate learning rate  $\eta_k$  based on iteration count  $k$  and Alpha
14:    Extract current batch of responsibilities
15:    Apply one-hot encoding to current batch responsibilities
16:    Update responsibilities with stochastic gradient step
17:    Increment  $k$ 
18:  end for
19:  Calculate log-likelihood for the current iteration
20:  if convergence criterion met then
21:    Update GMM parameters (NegLogLikelihood, Log_Likelihood)
22:    break
23:  end if
24:  Update oldLogLikelihood
25: end for
26: function ONE_HOT_ENCODING(Responsibilities)
27:   Initialize one-hot matrix to zeros
28:   for each data point do
29:     Set the maximum responsibility class to 1
30:   end for
31:   return one-hot matrix
32: end function
```

*Αλγόριθμος 3: Αποτύπωση του sEM όπως έχει διαρθρωθεί
βασίζόμενος στον ψευδοκώδικα που διατίθεται στην δημοσίευση [49].*

3 ΚΕΦΑΛΑΙΟ 3^ο : Μαθηματικό Υπόβαθρο

Αξιοποιώντας την γνώση που έχει συλλεχθεί μέχρι στιγμής, τόσο από την παγκόσμια βιβλιογραφία όσο και (ευελπιστούμε) από την παρούσα εργασία, τα δεδομένα που παρέχονται, είτε αυτά είναι ανεπεξέργαστα είτε έχουν μετασχηματιστεί σε μια πιο συμπαγή και συνάμα περιγραφική μορφή (για παράδειγμα μέσω της εξαγωγής SIFT χαρακτηριστικών), εδρεύουν σε χώρους υψηλών διαστάσεων με την μορφή σημείων σε αυτούς τους χώρους καθιστώντας την απεικόνιση μια μη τετριμμένη, πόσο μάλλον αδύνατη, καταβολή προσπάθειας²⁸. Τα μαθηματικά μπορούν να παραμερίσουν αυτό το τροχοπέδι προσφέροντας μετρικές που αποτυπώνουν την συσχέτιση των δεδομένων μεταξύ τους με τρόπο πλουραλιστικό, ουσιώδη, αδιαμφισβήτητο οπότε εν τέλει και κοινώς αποδεκτό. Με γνώμονα τα προηγούμενα, αξίζει να υπογραμμιστεί ότι στον τομέα της ταξινόμησης εικόνων²⁹, μια μετρική³⁰ μείζονος σημασίας είναι η **απόσταση** ανάμεσα σε αυτά τα σημεία του χώρου. Με την Ευκλείδεια απόσταση να είναι ο στυλοβάτης της περιγραφής τόσο του υποκείμενου χώρου των χαρακτηριστικών όσο και της μετρικής με την οποία εξοπλίζεται, η επιστημονική κοινότητα δόμησε ένα μεγάλο πλήθος ταξινομητών που αξιοποιούν την προαναφερθείσα ώστε να μπορούν να διαχωρίζουν τοπικούς πληθυσμούς που σχηματίζονται από δεδομένα που παρουσιάζουν έντονα κοινά χαρακτηριστικά (όπως για παράδειγμα τα SVM που αναλύθηκαν στο *Κεφάλαιο 2.2*).

Η περαιτέρω όμως ενασχόληση με την φύση των δεδομένων αποδεικνύει περίτρανα και επανειλημμένα ότι τα τελευταία δεν συμφιλιώνονται πάντοτε με την Ευκλείδεια δομή του χώρου, ανάγοντάς τα σε άλλους Χώρους όπου τυπικοί μαθηματικοί τελεστές που χρησιμοποιούνται για να αποτυπώσουν την έννοια της απόστασης όπως είναι γνωστή, παύουν να ισχύουν. Με το ενδιαφέρον της διεθνούς κοινότητας του χώρου της Υπολογιστικής Όρασης, αλλά και όχι μόνο, να κεντρίζεται στην αξιοποίηση των πινάκων συνδιακύμανσης, όπως αυτοί μπορούν να σχηματιστούν από την εφαρμογή διαφόρων φίλτρων για την εξαγωγή χαρακτηριστικών και με την επιβολή της συνθήκης ότι τα διανύσματα που τους αποτελούν είναι γραμμικώς ανεξάρτητα³¹, εντοπίζεται ότι τα δεδομένα πλέον αναπαρίστανται με την μορφή σημείων στην επιφάνεια μιας Πολλαπλότητας (*manifold*) που ανήκει στο σύνολο των Συμμετρικών Θετικά Ορισμένων (*SPD*) πινάκων οι οποίοι είναι Riemannian Πολλαπλότητες ή υπο-Πολλαπλότητες (*sub-manifolds*). Με την αξιοποίηση του Riemannian χώρου να οδηγεί τον δρόμο στην αύξηση της περιγραφικότητας των μοντέλων, το παρών κεφάλαιο φωτίζει

²⁸ Το παρών σχόλιο περί του απεικονιστικού περιορισμού που εισάγεται από τις διαστάσεις των δεδομένων, αν και μπορεί να φαντάζει στην αρχή άστοχο, στην πραγματικότητα είναι κάρριο καθώς μια διαισθητική, πόσο μάλλον οπτική, αλληλεπίδραση με αυτά μπορεί να οδηγήσει στην πρόωρη αντιμετώπιση χρονοβόρων κωλυμάτων ή σημείων πνευματικής τεμνάτωσης.

²⁹ Δεδομένης της προαναφερθείσας υπόθεσης και διαχείρισης των δεδομένων.

³⁰ Ο ορισμός της μετρικής θα αναλυθεί παρακάτω.

³¹ Ένας πίνακας συνδιακύμανσης είναι πάντοτε **συμμετρικός** και **θετικός ήμι-ορισμένος**, χωρίς ωστόσο να είναι **θετικά ορισμένος** [50], [51].

το μαθηματικό υπόβαθρο που θα χρειαστεί ο αναγνώστης ώστε να εφοπλιστεί με τα απαραίτητα εφόδια για να κατανοήσει τον χώρο των Riemannian πολλαπλοτήτων. Στο σημείο αυτό διευκρινίζεται ότι μεγάλο κομμάτι των απεικονίσεων έχουν αντληθεί από τις βιβλιογραφικές πηγές που χρησιμοποιήθηκαν για την κατανόηση των εννοιών. Σε όλες τις περιπτώσεις αποδίδονται τα εύσημα και δεν αναγνωρίζεται καμία μορφή ιδιοκτησίας από τον συγγραφέα [52] [53] [54].

3.1 Riemannian Πολλαπλότητες (*Riemannian Manifolds*)

Μέσω της ανάλυσης που θα πραγματοποιηθεί στις επόμενες σελίδες, διαφαίνεται πως το Riemannian Πολύπτυχο επεκτείνει τις αναλυόμενες έννοιες που μπορεί κανείς να εντοπίσει στο [Παράρτημα A](#) και [Παράρτημα B](#) με την εισαγωγή των ομαλά μεταβαλλόμενων θετικά ορισμένων μετρικών τενσόρων που προσδίδουν στην Πολλαπλότητα την έννοια της απόστασης και της γωνίας, επιτρέποντας την γενίκευση των γεωμετρικών και αναλυτικών ιδιοτήτων από τους Ευκλείδειους Χώρους, σε πιο σύνθετους, καμπυλωτούς. Αυτό το «άλμα» θέτει τις βάσεις για την διερεύνηση της πλούσιας γεωμετρίας και φωτίζει νέα προηγμένα αναλυτικά εργαλεία που κανείς μπορεί να αξιοποιήσει φυσικά στην Γεωμετρία του Riemann οδηγώντας την Υπολογιστική Όραση σε μια νέα εποχή καινοτομιών.

3.1.1 Η έννοια της μετρικής και του μετρικού Χώρου

Διασθητικά, ως *μετρικός Χώρος* ορίζεται ένα σύνολο το οποίο εξοπλίζεται με την έννοια της απόστασης ανάμεσα στα στοιχεία του. Η ίδια η απόσταση είναι μια συνάρτηση η οποία ονομάζεται *μετρική* ή *συνάρτηση απόστασης*. Η έννοια της απόστασης που κωδικοποιείται από τα αξιώματα του μετρικού Χώρου εμφανίζει σχετικά λίγες απαιτήσεις, καθώς η γενικότητα του ορισμού της είναι κάτι το οποίο προσδίδει μια ιδιαίτερη και μοναδική ευελιξία καθιστώντας την εύχρηστη σε μια πληθώρα εφαρμογών, μια εξ αυτών να αποτελούν και τα Riemannian Πολύπτυχα [73].

Με την επίσημη σημειογραφία του, μπορεί να γραφεί ότι:

Ορισμός 14: Μετρικός Χώρος (Metric Space)

Ένας μετρικός Χώρος είναι ένα διατεταγμένο ζεύγος (M, d) όπου το M είναι ένα σύνολο και το d είναι μια μετρική στο M , δηλαδή μια συνάρτηση [73]:

$$d: M \times M \rightarrow \mathbb{R} \quad (48)$$

Η σχέση (48) ικανοποιεί τα παρακάτω αξιώματα για όλα τα σημεία $x, y, z \in M$:

1. Η απόσταση από ένα σημείο στον εαυτό του θα πρέπει να είναι μηδενική:

$$d(x, x) = 0$$

2. (Θετικότητα): Η απόσταση ανάμεσα σε δυο ευδιάκριτα σημεία είναι πάντα θετική:

$$\text{για } x \neq y, \text{ τότε } d(x, y) > 0$$

3. (Συμμετρία): Η απόσταση από το x προς το y είναι πάντοτε η ίδια, ακόμη και με την υιοθέτηση αντίστροφης πορείας: $d(x, y) = d(y, x)$
4. Ισχύει η τριγωνομετρική ανισότητα: $d(x, z) \leq d(x, y) + d(y, z)$

Ολοκληρώνοντας, παρατίθεται και ο ορισμός της *μετρικής* χρησιμοποιώντας την σημειογραφία που εμφανίζεται στις *νηματικές δέσμες* (*fiber bundles*) και τις *διανυσματικές δέσμες* (*vector bundles*) ως:

Ορισμός 15: Μετρική

Ένας μετρικός τένσορας είναι μια συνάρτηση [74]:

$$g: TM \times_M TM \rightarrow \mathbb{R} \quad (49)$$

από το νηματικό γινόμενο της εφαπτομενικής δέσμης του M με τον εαυτό της στο \mathbb{R} με τέτοιο τρόπο ώστε ο περιορισμός του g σε κάθε νήμα να είναι μια μη-εκφυλισμένη διγραμμική απεικόνιση (*non-degenerate bilinear mapping*):

$$g_p: T_p M \times T_p M \rightarrow \mathbb{R} \quad (50)$$

Η απεικόνιση που περιγράφεται από τον σχέση (50) απαιτείται να είναι συνεχής, και συχνά συνεχώς διαφορίσιμη ή λεία κάτι το οποίο εξαρτάται από την εφαρμογή που εξετάζεται [74].

3.1.2 Ορισμός της Riemannian Πολλαπλότητας

Για τον ορισμό γεωμετρικών εννοιών, όπως τα μήκη και οι γωνίες σε ένα διανυσματικό χώρο, χρησιμοποιείται το εσωτερικό γινόμενο· ένα είδος συμμετρικών τενσόρων σε έναν διανυσματικό Χώρο. Για τις Πολλαπλότητες, η κατάλληλη δομή είναι μια Riemann μετρική, η οποία είναι ουσιαστικά μια επιλογή του εσωτερικού γινομένου σε κάθε εφαπτόμενο χώρο, που μεταβάλλεται ομαλά από σημείο σε σημείο. Η επιλογή της μετρικής Riemann επιτρέπει τον ορισμό γεωμετρικών εννοιών όπως μήκη, γωνίες και αποστάσεις σε λείες Πολλαπλότητες δίνοντας άμεσα την δυνατότητα χρήσης της Ευκλείδειας Γεωμετρίας [68].

Η έρευνα ξεκινά με την αποτίμηση της έννοιας της Riemannian μετρικής ως:

Ορισμός 16: Riemannian Μετρική

Μια Riemannian μετρική στο M είναι ένα λείο συμμετρικό συμμεταβλητό πεδίο 2-τανυστών (*symmetric covariant 2-tensor field*) στο M που είναι θετικά ορισμένο σε κάθε σημείο [68]. Ενισχύοντας περαιτέρω τον ορισμό με την χρήση της πηγής [60], μια Riemannian μετρική είναι ένα σύνολο εσωτερικών γινομένων

$$T_p M \times T_p M \rightarrow \mathbb{R} : (v, w) \mapsto g_p(v, w)$$

ένα για κάθε σημείο στο Πολύπτυχο, $p \in M$, έτσι ώστε η απεικόνιση:

$$M \rightarrow \mathbb{R} : p \mapsto g_p(X(p), Y(p))$$

να είναι λεία για κάθε ζεύγος διανυσματικών πεδίων $X, Y \in Vect(M)$. Η συνήθης συνοπτική σημειογραφία που την συνοδεύει είναι η $\langle \cdot, \cdot \rangle_p$ [67].

Έχοντας ως έναυσμα τον παραπάνω ορισμό, παρατίθεται ότι:

Ορισμός 17: Riemannian Πολλαπλότητα

Ως Riemannian Πολλαπλότητα ορίζεται το ζεύγος (M, g) , όπου M είναι το λείο Πολύπτυχο και g είναι η Riemannian μετρική στην Εφαπτομενική Δέσμη αυτού.

3.1.3 Γεωδαισιακές

Με την αποτίμηση της φύσης του Χώρου των Riemannian Πολλαπλοτήτων, πλέον παρέχεται η δυνατότητα της αποτύπωσης εννοιών μείζονος σημασίας για την παρούσα εργασία που εμφανίζονται φυσικά στον απείρως διαφορίσιμο Χώρο αυτών. Ίσως μια από τις πιο κρίσιμες εξ αυτών είναι η έννοια του μήκους, όπως αυτή περιγράφεται από τις γεωδαισιακές.

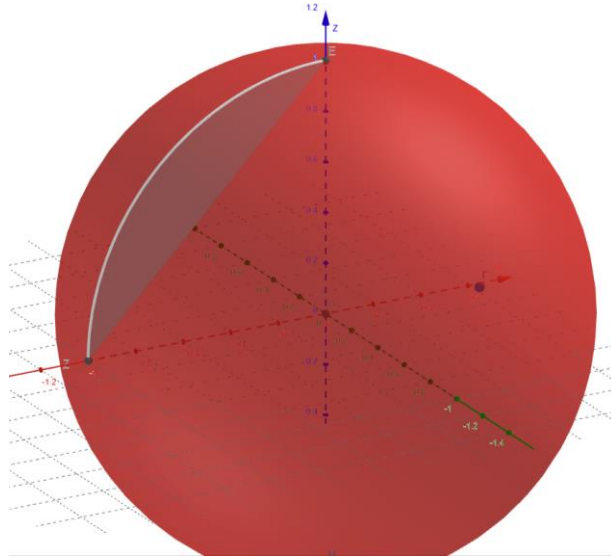
Με την τυπική έννοια της απόστασης στον Ευκλείδειο Χώρο να ορίζεται ως το μήκος του ευθύγραμμου τμήματος εκείνου που ενώνει δυο σημεία του Χώρου, η έννοια της γεωδαισιακής σε ένα Πολύπτυχο αποτελεί μια γενίκευση της ευθείας γραμμής. Προτού αποδοθεί η μαθηματική ορολογία της, αποτελεί ένδειξη καλής πρακτικής να αποδοθεί ο ορισμός του μήκους σε ένα Πολύπτυχο [75] [60].

Ορισμός 18: Μήκος σε μια Πολλαπλότητα

Έστω ότι το $I = [a, b] \subset \mathbb{R}$ είναι ένα συμπαγές διάστημα όπου $a < b$. Συμπληρωματικά έστω ότι $\gamma: I \rightarrow M$ είναι μια λεία καμπύλη στο M . Τότε ως μήκος $L(\gamma)$ σε μια Πολλαπλότητα ορίζεται το [60]:

$$L(\gamma) = \int_a^b |\dot{\gamma}(t)| dt \quad (51)$$

³² Προκειμένου να αποφευχθεί οποιοδήποτε είδους σύγχυσης, η χρήση του t αποτελεί εδώ απλά μια σημειογραφία που χρησιμοποιείται για να υποδηλώσει την παραμετροποίηση της καμπύλης, χωρίς απαραίτητα να εισάγει την έννοια του χρόνου.



Εικόνα 34: Απεικόνιση της έννοιας του μήκους σε μια Πολλαπλότητα.

Αξιοσημείωτη είναι και η **διακύμανση του γ** (*variation of γ*), το οποίο ορίζεται πως είναι μια οικογένεια ομαλών καμπυλών $\gamma_s : I \rightarrow M$, όπου το s λαμβάνει πραγματικές τιμές, έτσι ώστε η απεικόνιση $\mathbb{R} \times I \rightarrow M : (s, t) \mapsto \gamma_s(t)$ να είναι λεία και να ισχύει ότι $\gamma_0 = \gamma$. Το σημαντικό με την ιδιότητα της *διακύμανσης του γ* , $\{\gamma_s\}_{s \in \mathbb{R}}$, είναι πως στην περίπτωση όπου ισχύει ότι $\gamma_s(\mathbf{a}) = \gamma(\mathbf{a})$ και $\gamma_s(\mathbf{b}) = \gamma(\mathbf{b})$ για όλα τα $s \in \mathbb{R}$ τότε εισάγεται φυσικά μια ένα-προς-ένα απεικόνιση των σημείων πάνω στην καμπύλη, εισάγοντας την έννοια των *πάγιων απολήξεων (fixed endpoints)* [60].

Ορισμός 19: Γεωδαισιακή

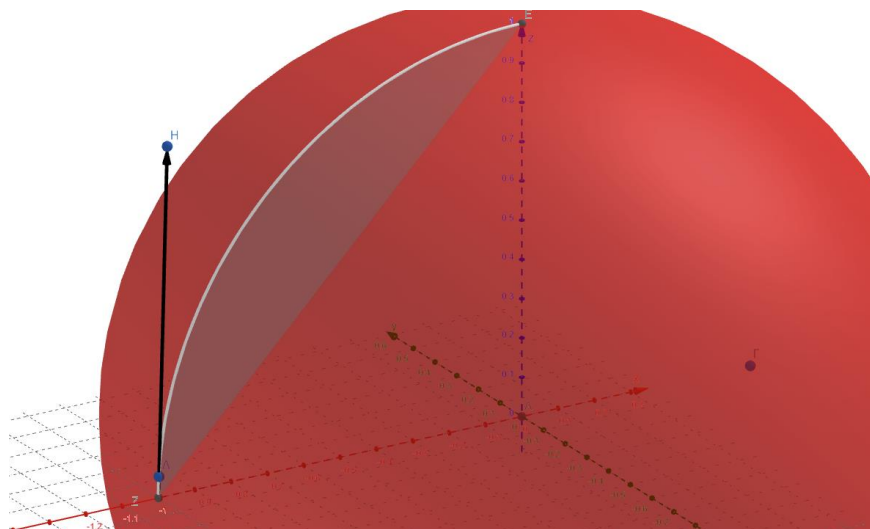
Μια λεία καμπύλη $\gamma : I \rightarrow M$ σε ένα διάστημα I ονομάζεται *γεωδαισιακή* αν οι περιορισμοί της σε κάθε συμπαγές υποδιάστημα ικανοποιούν την συνθήκη [60]:

$$\nabla \dot{\gamma}(t) = 0 \text{ για όλα τα } t \in I \quad (52)$$

Όπου με ∇ ορίζεται πως είναι η *Levi-Civita* σύνδεση. Για λόγους καλής πρακτικής, αλλά χωρίς να γίνει ενδελεχής μελέτη, η σύνδεση *Levi-Civita* είναι μια μοναδική affine γραμμική σύνδεση στην Εφαπτομενική Δέσμη ενός Πολύπτυχου που παρέχει την δυνατότητα διατήρησης της Riemannian μετρικής και είναι απηλλαγμένη από στρέψη (*torsion-free*). Η ύπαρξή της είναι υψίστης σημασίας καθώς επιτρέπει τον ορισμό των γεωδαισιακών, της παράλληλης μεταφοράς και της συμμεταβλητής παραγώγου με τρόπο ο οποίος είναι συμβατός πάντα με την υποκείμενη μετρική. Έτσι το κομμάτι της μετρικής συμβατότητας εξασφαλίζει πως το μήκος των διανυσμάτων καθώς και οι γωνίες μεταξύ αυτών διατηρούνται κατά την παράλληλη μεταφορά ενώ η ιδιότητα της απηλλαγμένης στρέψης εξασφαλίζει την συμμετρική σχέση μεταξύ της σύνδεσης και της δομής του Πολύπτυχου [76], [77].

Πρακτικά, ο *Ορισμός 19* αποτυπώνει πως η συμμεταβλητή παράγωγος (*covariant derivative*) του κάθε σημείου από το οποίο συνίσταται το καμπυλωτό ευθύγραμμο τμήμα που ενώνει δυο σημεία πάνω στην Πολλαπλότητα θα πρέπει να είναι μηδενικό, δηλαδή το διάνυσμα μετακίνησης από το

σημείο Z προς το σημείο E , όπως αποτυπώνονται στην *Εικόνα 35*, θα πρέπει να είναι συνεχώς παράλληλο στον εαυτό του κατά μήκος της κίνησής του πάνω στην καμπύλη.



Εικόνα 35: Αποτύπωση της παραλληλίας του διανύσματος μετακίνησης προς τον εαυτό του.

Αφενός η ύπαρξη των γεωδαισιακών αποτελεί ένα αναλογικό μέσο το οποίο γεφυρώνει την έννοια της ευθείας γραμμής, όπως αυτή είναι γνωστή από την Ευκλείδεια Γεωμετρία, με την εγγενή έννοια της απόστασης στον καμπυλωτό Χώρο των Riemannian Πολλαπλοτήτων, αφετέρου όμως το τελευταίο δεν είναι αρκετό καθώς θα πρέπει να επιστρατευτεί και να αποδοθεί κάποιου είδους σημειογραφία που να αποδεικνύει την ύπαρξη μιας μοναδικής γεωδαισιακής που ενώνει δυο σημεία. Έτσι λοιπόν αποδίδεται το Θεώρημα Μοναδικότητας των γεωδαισιακών ως:

Ορισμός 20: Θεώρημα μοναδικότητας Γεωδαισιακών

Έστω ότι το M είναι ένα Πολύπτυχο με μια γραμμική σύνδεση. Για οποιοδήποτε $p \in M$, οποιοδήποτε $V \in T_p M$ και οποιοδήποτε $t_0 \in \mathbb{R}$ υπάρχει ένα ανοιχτό διάστημα $I \subset \mathbb{R}$ που περιλαμβάνει το t_0 και μια γεωδαισιακή $\gamma : I \rightarrow M$ που ικανοποιεί την συνθήκη $\gamma(t_0) = p$ και $\dot{\gamma}(t_0) = V$. Οποιοσδήποτε δυο τέτοιες γεωδαισιακές συμφωνούν στο κοινό τους τομέα [76].

Προκύπτει λοιπόν μέσω του Θεωρήματος Μοναδικότητας των γεωδαισιακών ότι για οποιοδήποτε $p \in M$ και $V \in T_p M$ υπάρχει μια μοναδική μέγιστη γεωδαισιακή (maximal geodesic), δηλαδή μια γεωδαισιακή που δεν μπορεί να επεκταθεί σε οποιοδήποτε μεγαλύτερο διάστημα τιμών, όπου το $\gamma(0) = p$ και το $\dot{\gamma}(0) = V$, οριζόμενα σε ένα ανοιχτό διάστημα I . Ο συνηθισμένος τρόπος εκπεφρασμού της μέγιστης γεωδαισιακής είναι μέσω της σημειογραφίας γ_V [76].

3.1.4 Εκθετική Απεικόνιση

Μελετήθηκε το πως οι γεωδαισιακές καταφέρνουν να εισάγουν την έννοια της απόστασης στον μη-γραμμικό Χώρο των Riemannian Πολύπτυχων καθώς και το πως η επισφράγιση της μοναδικότητάς τους εγγυάται την ύπαρξη μιας και μοναδικής «πορείας» ανάμεσα σε δυο σημεία πάνω στην Πολλαπλότητα. Διαγράφεται επίσης ότι η επιλογή της υποκείμενης γεωδαισιακής προς επεξεργασία, είναι κάτι που εξαρτάται σε πολύ μεγάλο βαθμό από την επιλογή του σημείου έναρξης. Οπότε εγείρεται η ερώτηση: «Πως μεταβάλλονται οι γεωδαισιακές αν μεταβάλλουμε το αρχικό σημείο»; Η απάντηση σε αυτό το ερώτημα δίνεται από την χρήση της Εκθετικής Απεικόνισης [76].

Πριν την απόθεση του ορισμού της Εκθετικής Απεικόνισης, είναι επιτακτικό να μπορέσει να αποτυπωθεί μια ακόμη κρίσιμη έννοια η οποία σχετίζεται άμεσα με την μέγιστη γεωδαισιακή. Συγκεκριμένα:

Ορισμός 21: Διάστημα μέγιστης ύπαρξης (maximal existence interval)

Για ένα σημείο $p \in M$ και ένα διάνυσμα $u \in T_p M$ ορίζεται το διάστημα:

$$I_{p,u} := \bigcup I \subset \mathbb{R}$$

Όπου I είναι ένα ανοιχτό διάστημα που περιλαμβάνει το 0 και υπάρχει μια γεωδαισιακή $\gamma: I \rightarrow M$ που να ικανοποιεί τις σχέσεις $\gamma(0) = p$ και $\dot{\gamma}(0) = u$. Το $I_{p,u}$ αποκαλείται διάστημα μέγιστης ύπαρξης για την γεωδαιτική που διέρχεται από το p στην κατεύθυνση του u .

Ορισμός 22: Εκθετική Απεικόνιση

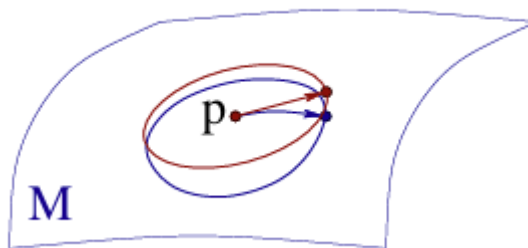
Δοσμένου λοιπόν πάλι του σημείου $p \in M$ ορίζεται το σύνολο $V_p \subset T_p M$ ως

$$V_p := \{u \in T_p M \mid 1 \in I_{p,u}\}$$

το οποίο πλαισιώνει τον ορισμό της Εκθετικής Απεικόνισης ως

$$\exp_p: V_p \rightarrow M \tag{53}$$

Η φυσική σημασία της σχέσης (53) είναι η παροχή της δυνατότητας ανάθεσης σε κάθε εφαπτομενικό διάνυσμα $u \in V_p$ το σημείο $\exp_p(u) := \gamma(1)$, όπου το $\gamma: I_{p,u} \rightarrow M$ είναι η μοναδική γεωδαισιακή που ικανοποιεί τους περιορισμούς $\gamma(0) = p$ και $\dot{\gamma}(0) = u$. Άρα η Εκθετική Απεικόνιση μπορεί να ιδωθεί ως ένας τρόπος προβολής από το Εφαπτομενικό Χώρο πάνω στο Πολύπτυχο όπου με $\gamma(0)$ συμβολίζεται το σημείο έναρξης ενώ με $\gamma(1)$ συμβολίζεται το σημείο λήξης της καμπυλωτής πορείας της γεωδαισιακής, η οποία προφανώς είναι διαφορετική για την επιλογή διαφορετικών σημείων.



Εικόνα 36: Γραφική αποτύπωση της Εκθετικής Απεικόνισης³³.

3.1.5 Οι Πολλαπλότητες των Συμμετρικών Θετικά Ορισμένων Πινάκων

Μια πληθώρα τομέων, όπως η απεικόνιση τανυστή διάχυσης στον τομέα της ιατρικής, η διαχείριση ρίσκου στον τομέα των οικονομικών, η κλιματική μοντελοποίηση στον τομέα των περιβαλλοντικών επιστημών, η επεξεργασία σήματος στον τομέα της μηχανικής όπως και πολλοί ακόμη, έχουν γνωρίσει εξέχουσα άνθηση μέσω της χρήσης των πινάκων συνδιακύμανσης (*covariance matrices*). Τα πλεονεκτήματα που εγείρονται από την χρήση τους εντοπίζονται στον συμπυκνωμένο μα συνάμα πλουραλιστικό τρόπο απόδοσης της εγγενούς μεταβολής του περιγραφόμενου μεγέθους³⁴ συναρτήσει της χαμηλής διαστατικότητας και της εύρωστης δομής τους. Γίνεται αντιληπτό ότι η χρήση ενός τέτοιου μαθηματικού εργαλείου καταφέρνει να αποδώσει ένα καινοτόμο πεδίο περιγραφικότητας σε διαδικασίες που επωφελούνται από την στατιστική φύση των δεδομένων τους, με την Υπολογιστική Όραση να μην εντάσσεται εκτός αυτού του φάσματος.

Η συμμετρική κατανομή των στοιχείων των πινάκων συνδιακύμανσης σε συνδυασμό με το γεγονός ότι, τουλάχιστον για την παρούσα εργασία, οι ιδιοτιμές τους είναι μόνο θετικές τους αποδίδει την συμμετρική και θετικά ορισμένη φύση τους με τους ίδιους να εντοπίζονται ότι “ζουν” στον Πολύπτυχο που συνιστά τους Συμμετρικούς και Θετικά Ορισμένους Πίνακες³⁵. Η παρούσα υποενότητα αναλαμβάνει να μελετήσει αυτόν τον Χώρο, να φέρει στο προσκήνιο κάποια από τα βασικά χαρακτηριστικά του και να υπογραμμίσει τα εργαλεία που βοηθούν στην αξιοποίηση της γεωμετρικής δομής του προς όφελος της ενίσχυσης της περιγραφικότητας των μοντέλων της Υπολογιστικής Όρασης.

³³ Η παρούσα εικόνα προέρχεται από την πηγή [60]

³⁴ Παρέχοντας προσδοφόρο έδαφος για την εισαγωγή στο προσκήνιο της έννοιας της συνδεσιμότητας (*connectivity*).

³⁵ Για λόγους συντομογραφίας, πλέον οι Συμμετρικοί Θετικά Ορισμένοι Πίνακες θα αναγράφονται ως SPD (*Symmetric Positive Definite*).

Εξ' ορισμού ισχύει πως:

Ορισμός 23: Συμμετρικός Θετικά Ορισμένος Πίνακας

Ένας συμμετρικός πίνακας $P \in \mathbb{R}^{n \times n}$ με πραγματικές τιμές είναι **θετικά ορισμένος** όταν έχει αποκλειστικά και μόνο θετικές ιδιοτιμές. Σε πινακική μορφή μπορεί να γραφεί ότι για οποιοδήποτε διάνυσμα $u \in \mathbb{R}^n$, ισχύει [40], [52], [53], [54], [78]:

$$u^T P u > 0 \text{ για καθε } u \in \mathbb{R}^n \setminus \{0\} \quad (54)$$

Ορισμός 24: Το SPD Πολύπτυχο

Έστω ότι M είναι ένα SPD πίνακας, τότε ως SPD Πολύπτυχο ορίζεται το σύνολο όλων των Συμμετρικών Θετικά Ορισμένων Πινάκων και περιγράφεται από τον τύπο [53]:

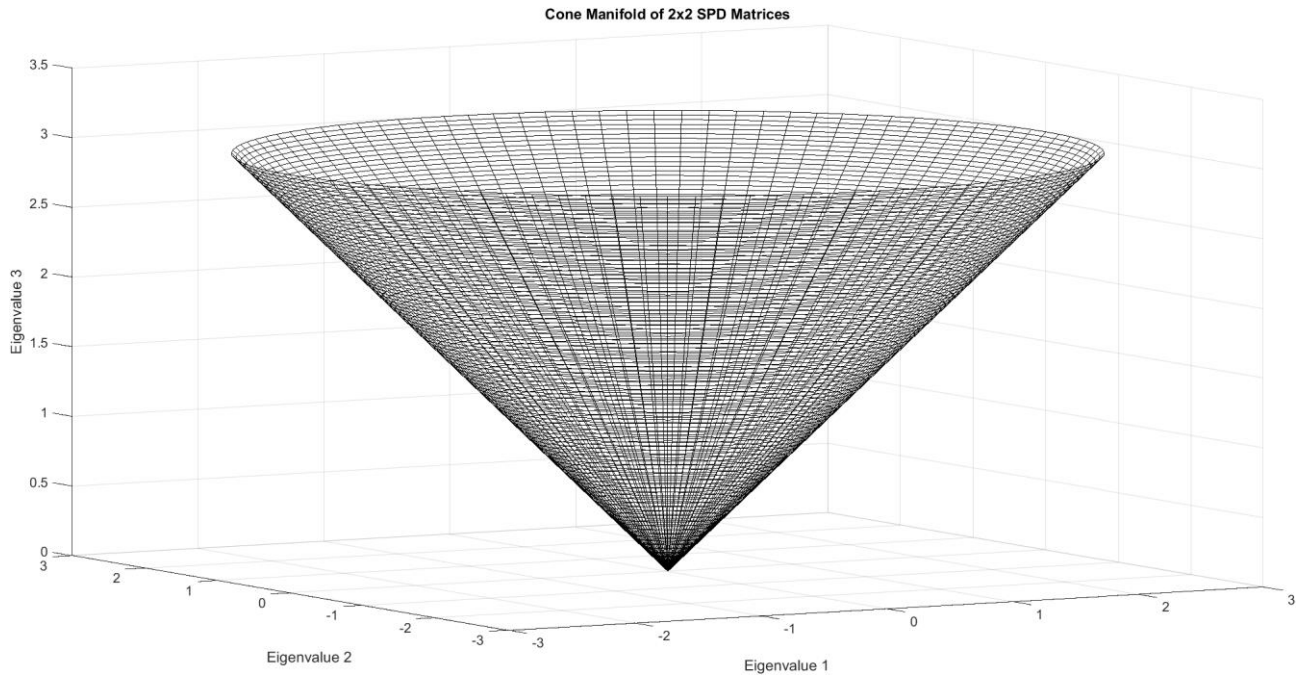
$$\mathcal{S}_d^{++} = \{M \in \mathbb{R}^{d \times d} \mid M = M^T, x^T M x \geq 0, \forall x \in \mathbb{R}^d \setminus \{0\}\} \quad (55)$$

Αυτός ο ορισμός δίνει την δυνατότητα της απεικονιστικής αποτύπωσης του Χώρου των SPD πινάκων ως έναν κυρτό ημι-κώνο στον διανυσματικό Χώρο των $n \times n$ συμμετρικών πινάκων, με τον τελευταίο να συνιστά ένα διαφορίσιμο Riemannian Πολύπτυχο, το οποίο έρχεται εφοδιασμένο με το εσωτερικό γινόμενο:

$$\langle S_1, S_2 \rangle_{\mathcal{T}_P \mathcal{S}_d^{++}} = \left\langle P^{-\frac{1}{2}} S_1 P^{-\frac{1}{2}}, P^{-\frac{1}{2}} S_2 P^{-\frac{1}{2}} \right\rangle \quad (56)$$

Όπου ως

- $\mathcal{T}_P \mathcal{S}_d^{++}$: ορίζεται και πάλι ο Εφαπτομενικός Χώρος των SPD πινάκων αυτή τη φορά στο σημείο $P \in \mathcal{S}_d^{++}$, $S_1, S_2 \in \mathcal{T}_P \mathcal{S}_d^{++}$.
- $\langle \cdot, \cdot \rangle$: ορίζεται ότι είναι ο τυπικός τελεστής του εσωτερικού γινομένου του Ευκλείδειου Χώρου.



Εικόνα 37: Απεικόνιση της ημι-κωνικής δομής του SPD Πολύπτυχου ενός πίνακα συνδιακύμανσης 2×2 .

Ο παραπάνω χαρακτηρισμός του SPD Πολύπτυχου ως ένα διαφορίσιμο Riemannian Πολύπτυχο όμως είναι ημιτελής, καθώς όπως διαφάνεται κι από την υποενότητα *Ορισμός της Riemannian Πολλαπλότητας*, θα πρέπει να τοποθετηθεί στην δομή ένα είδους μετρικής, η οποία με την σειρά της να μπορεί να ερμηνεύσει την έννοια της απόστασης στην Πολλαπλότητα. Έτσι λοιπόν ορίζεται πως:

Ορισμός 25: Η SPD Πολλαπλότητα ως διαφορίσιμο Riemannian Πολύπτυχο

Το SPD Πολύπτυχο, όταν εφοδιάζεται με μια Riemannian μετρική όπως αυτή της Rao-Fisher, σχηματίζει ένα συνεχώς διαφορίσιμο Πολύπτυχο το οποίο είναι ολικά διαφορομορφικό, επιτρέποντας μια αμφίδρομη ένα-προς-ένα απεικόνιση [52], [53].

Με το M να υποδεικνύει ξανά έναν SPD πίνακα, $M \in \mathcal{S}_d^{++}$, και τα $S_1, S_2 \in \mathcal{T}_M \mathcal{S}_d^{++}$, τότε η μετρική Rao-Fisher δίνεται από τον τύπο:

$$\langle S_1, S_2 \rangle_M = \text{tr}(M^{-1} S_1 M^{-1} S_2)^2 \quad ^{36} \quad (57)$$

Με τον ορισμό της SPD Πολλαπλότητας εμπειριστατωμένο και πλήρως δομημένο, το επόμενο βήμα είναι να μπορέσουν να εκφραστούν όλα εκείνα τα εργαλεία τα οποία θα επιτρέψουν την απεικόνιση των γενικευμένων εννοιών που παρουσιάστηκαν σε όλο το μήκος της παρούσας ενότητας στην SPD Πολλαπλότητα συγκεκριμένα με έναν εκ νέου προσδιορισμό.

Με την εισαγωγή της μετρικής Rao-Fisher το Πολύπτυχο εξοπλίζεται με ένα μέσο το οποίο του παρέχει τη δυνατότητα του εκπεφρασμού της έννοιας της απόστασης. Με τη μη-γραμμική και συνάμα καμπυλωτή φύση της Πολλαπλότητας των SPD πινάκων κατά νου, κανείς μπορεί πολύ

³⁶ Με tr συμβολίζεται το ίχνος ενός πίνακα (trace).

εύκολα να συνάγει ότι το μόνο μαθηματικό μέγεθος το οποίο μπορεί να αποτυπώσει την διαδρομή που πρέπει να ακολουθήσει ένα σημείο της Πολλαπλότητας για να φτάσει σε ένα νέο, αλλά συνάμα προκαθορισμένο, σημείο είναι αυτή της γεωδαισιακής. Για την μετρική λοιπόν αυτή ορίζεται ότι:

Ορισμός 26: Γεωδαισιακή στο SPD Πολύπτυχο

Δεδομένης της μετρικής Rao-Fisher, η γεωδαισιακή, δηλαδή η πιο σύντομη διαδρομή, που συνδέει δύο σημεία S_1 και S_2 πάνω στην Πολλαπλότητα δίνεται από την μονοπαραμετρική καμπύλη [52], [53], [54], [78], [79], [80], [81]:

$$\Gamma(t) = S_1^{1/2} (S_1^{-1/2} S_2 S_1^{-1/2})^t S_1^{-1/2}, t \in [0,1] \quad (58)$$

Με τέτοιο τρόπο ώστε το $\Gamma(0) = S_1$ και το $\Gamma(1) = S_2$

Το μήκος τόξου που προκύπτει από την εκάστοτε γεωδαισιακή καμπύλη ορίζει την ακόλουθη Riemannian απόσταση πάνω στο Πολύπτυχο :

$$\begin{aligned} d_R^2(S_1, S_2) &= \|\log(S_2^{-\frac{1}{2}} S_1 S_2^{-\frac{1}{2}})\|_F^2 \\ &\quad \text{ή} \\ &= \sum_{i=1}^n \log^2 \left(\lambda_i \left(S_2^{-\frac{1}{2}} S_1 S_2^{-\frac{1}{2}} \right) \right) \end{aligned} \quad (59)$$

Όπου:

- $S_1, S_2 \in \mathcal{S}_d^{++}$
- $\|\cdot\|_F$: Είναι η γνωστή νόρμα Frobenius.
- $\log(S)$: Είναι η λογαριθμική συνάρτηση πίνακα (*matrix logarithm*).
- $\lambda_i(S)$: Είναι οι *i-οστή* ιδιοτιμή του S.

Αναφέρθηκε προηγουμένως ότι το SPD Πολύπτυχο επιτρέπει μια αμφίδρομη ένα-προς-ένα απεικόνιση, χωρίς όμως να έχει προσδιοριστεί ανάμεσα σε τι. Με μια ανασκόπηση στην έννοια της *Εκθετικής Απεικόνισης*, εντοπίζεται ότι η προαναφερθείσα αποτελεί έναν τρόπο μετάβασης από τον Εφαπτομενικό Χώρο που συντίθεται από την Εφαπτομενική Δέσμη του Πολύπτυχου, στο ίδιο το Πολύπτυχο. Όμως υπάρχει και ο τρόπος με τον οποίο κανείς μπορεί να μεταβεί από το Riemannian Πολύπτυχο «πάνω» στον διανυσματικό Εφαπτομενικό Χώρο ενός σημείου. Αυτό μπορεί να επιτευχθεί με την βοήθεια της *Λογαριθμικής Απεικόνισης*. Με την σημείωση ότι ο Εφαπτομενικός Χώρος $\mathcal{T}_p \mathcal{S}_d^{++}$ ενός σημείου $M \in \mathcal{S}_d^{++}$ είναι ένα σύνολο εφαπτομενικών διανυσμάτων που είναι παράγωγοι των καμπυλών που διέρχονται από το p , τότε μπορεί να γραφεί ότι:

Ορισμός 27: Εκθετική Απεικόνιση στο SPD Πολύπτυχο

Με το $U \in \mathcal{T}_P \mathcal{S}_d^{++}$ να υποδεικνύει ένα εφαπτομενικό διάνυσμα στο σημείο X , ως Εκθετική Απεικόνιση ορίζεται η αποτύπωση μιας μοναδικής γεωδαισιακής από το $X \in \mathcal{S}_d^{++}$ στην κατεύθυνση του U (προβολή) σε ένα σημείο Y στο Πολύπτυχο \mathcal{S}_d^{++} και δίνεται από τον τύπο [52], [53], [54], [78], [79], [80], [81]:

$$Y = \mathbf{Exp}_X(U) = X^{\frac{1}{2}} \mathbf{exp} \left(X^{-\frac{1}{2}} U X^{-\frac{1}{2}} \right) X^{\frac{1}{2}} \in \mathcal{S}_d^{++} \quad (60)$$

Αντίστροφα:

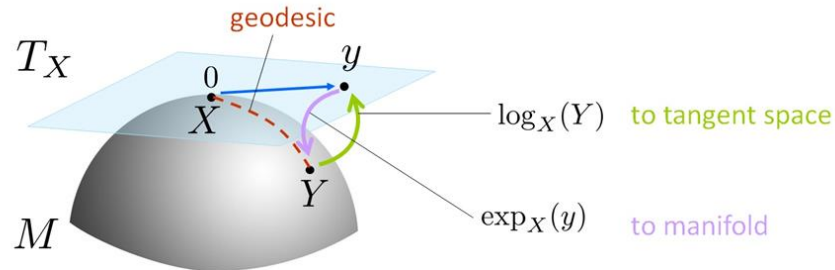
Ορισμός 28: Λογαριθμική Απεικόνιση

Η Λογαριθμική Απεικόνιση, η οποία μπορεί να ιδωθεί ως η αντίστροφη διαδικασία της Εκθετικής, προβάλλει έναν SPD πίνακα³⁷ $Y \in \mathcal{S}_d^{++}$ στο Εφαπτομενικό επίπεδο $\mathcal{T}_P \mathcal{S}_d^{++}$ του $X \in \mathcal{S}_d^{++}$ και δίνεται από τον τύπο [52], [53], [54], [78], [79], [80], [81]:

$$U = \mathbf{log}_X(Y) = X^{\frac{1}{2}} \mathbf{log} \left(X^{-\frac{1}{2}} Y X^{-\frac{1}{2}} \right) X^{\frac{1}{2}} \in \mathcal{T}_P \mathcal{S}_d^{++} \quad (61)$$

Και σε αυτή την περίπτωση, ο λογαριθμικός τελεστής αναφέρεται στην λογαριθμική συνάρτηση πίνακα (*matrix logarithm*). Προς αποφυγή οποιασδήποτε σύγχυσης, τόσο η Εκθετική όσο και η Λογαριθμική Απεικόνιση μπορούν να γραφούν πολύ πιο απλά ως συναρτήσεις που επιτελούν τα εξής:

- $\mathbf{exp}_X : \mathcal{T}_X \mathcal{S}_d^{++} \mapsto \mathcal{S}_d^{++}$
- $\mathbf{log}_X : \mathcal{S}_d^{++} \mapsto \mathcal{T}_X \mathcal{S}_d^{++}$



Εικόνα 38: Απεικόνιση της διαδικασίας της Εκθετικής και Λογαριθμικής Απεικόνισης από ένα σημείο του Εφαπτομενικού Χώρου πίσω στο Πολύπτυχο και το αντίστροφο.³⁸

³⁷ Τονίζεται πως όπως ακριβώς τα διανύσματα μπορούν να θεωρηθούν ως σημεία ενός Χώρου υψηλών διαστάσεων, έτσι και ένας SPD πίνακας μπορεί να θεωρηθεί ως ένα σημείο στο Πολύπτυχο στο οποίο «ζει».

³⁸ Η παρούσα εικόνα έχει ληφθεί από την πηγή [82].

Φωτίζοντας έναν προηγούμενος αδιευκρίνιστο κομμάτι, αξίζει να αποτυπωθεί και το εξής σχόλιο: Αναφέρθηκε πως ο εφαπτομενικός Χώρος που επάγεται από το κάθε σημείο της Πολλαπλότητας των SPD πινάκων είναι ένας τοπικά διανυσματικός Ευκλείδειος Χώρος. Τα διανύσματα αυτά στην πραγματικότητα είναι συμμετρικοί πίνακες οι οποίοι ακριβώς λόγω της ιδιότητας της συμμετρίας τους συμπυκνώνουν³⁹ την ωφέλιμη πληροφορία τους στην κύρια διαγώνιο και στον άνω ή κάτω τριγωνικό πίνακα οδηγώντας στην περιγραφή της υποκείμενης πληροφορίας από $\frac{d(d+1)}{2}$ ανεξάρτητες τιμές, μετασχηματίζοντας την αμιγώς πινακική μορφή τους σε διανυσματική. Αυτός ο διανυσματικός εκπεφρασμός των ανεξάρτητων τιμών του αυθαίρετου συμμετρικού πίνακα υπό εξέταση, τυπικά λαμβάνει μια πιο συνοπτική αναπαράσταση των σημείων στον Εφαπτομενικό Χώρο μέσω της επιβολής⁴⁰ ορισμού ενός συστήματος ορθοκανονικών συντεταγμένων που υλοποιείται από μια διανυσματική διαδικασία. Συγκεκριμένα [82]:

Ορισμός 29: Διανυσματική Αναπαράσταση στον Εφαπτομενικό Χώρο με ορθοκανονικές συντεταγμένες

Οι ορθοκανονικές συντεταγμένες ενός εφαπτομενικού διανύσματος \mathbf{y} στον Εφαπτομενικό Χώρο του σημείου \mathbf{X} δίνεται από τον διανυσματικό τελεστή:

$$\mathbf{vec}_X(\mathbf{y}) = \mathbf{vec}_I \left(\mathbf{X}^{-\frac{1}{2}} \mathbf{y} \mathbf{X}^{-\frac{1}{2}} \right) \quad (62)$$

Όπου με I συμβολίζεται ο ταυτοτικός πίνακας, ενώ ο διανυσματικός τελεστής στον κοινό πόλο που δημιουργείται από τον τελευταίο ορίζεται ως:

$$\mathbf{vec}_I(\mathbf{y}) = [\mathbf{y}_{1,1}, \sqrt{2}\mathbf{y}_{1,2}, \dots, \mathbf{y}_{d,d}]^T \quad (63)$$

Οι παραπάνω μεθοδολογίες μπορούν να θεωρηθούν ως διορθώσεις/κανονικοποιήσεις που υφίστανται οι διανυσματικές αναπαραστάσεις των σημείων των Πολύπτυχων στον Εφαπτομενικό τους Χώρο, ενώ συμπληρωματικά συσχετίζει την Riemannian μετρική στον Εφαπτομενικό Χώρο με την κανονική μετρική που ορίζεται ως:

$$\langle \mathbf{y}, \mathbf{y} \rangle_x = \|\mathbf{vec}_x(\mathbf{y})\|_2^2 \quad (64)$$

Τα παραπάνω δίνουν την δυνατότητα επαναπροσδιορισμού του τύπου (64) που αποτυπώνει την απόσταση σε μια πιο ευανάγνωστη μορφή αξιοποιώντας παράλληλα περαιτέρω τις φυσικές ιδιότητες της Γραμμικής Άλγεβρας ως:

³⁹ Στο παρόν σημείο, ο όρος της συμπύκνωσης χρησιμοποιείται απλά και μόνο για να υποδείξει το φαινόμενο της μείωσης της αρχικής διαστατικότητας και δεν αντικατοπτρίζει την τυπική διαδικασία αυτής, καθώς δεν υπάρχει κάποιο σημείο στο οποίο να χάνεται πληροφορία.

⁴⁰ Αν και η λέξη «επιβολή» φέρει αξιολογική και σημαντική βαρύτητα, η χρήση της εδώ γίνεται σκόπιμα καθώς η μεθοδολογία που περιγράφεται από τον διανυσματικό τελεστή θα συναντηθεί μετέπειτα στα σενάρια των πειραμάτων

$$\begin{aligned} d^2(X, Y) &= \langle \log_X(Y), \log_X(Y) \rangle_X \\ &= \text{trace} \left(\log^2 \left(X^{-\frac{1}{2}} Y X^{-\frac{1}{2}} \right) \right) \end{aligned} \quad (65)$$

Μετασχηματίζοντας τον τύπο (65) έτσι ώστε να είναι απηλλαγμένος από την τετραγωνική μορφή του, μπορεί να γραφεί ότι [81]:

$$d(X, Y) = \left\| \log m \left(X^{-\frac{1}{2}} Y X^{-\frac{1}{2}} \right) \right\|_F \quad (66)$$

Το μαθηματικό μέγεθος που περιγράφεται από τον άνωθεν τύπο ονομάζεται **Affine Invariant Riemannian Metric- AIRM** και αποτελεί τον στυλοβάτη της πλαισίωσης της έννοιας της απόστασης ανάμεσα σε δυο σημεία πάνω στο SPD Πολύπτυχο καθώς παραμένει αναλλοίωτη υπό την επίρεια affine μετασχηματισμών ενώ παράλληλα παρέχει μια φυσική δομή που επιτρέπει την συνεκτική και ουσιώδη σύγκριση ανάμεσα σε SPD πίνακες. Κλείνοντας, αξίζει να αναφερθεί ότι η AIRM δεν είναι η μόνη μετρική που μπορεί να εμφανιστεί στα Πολύπτυχα των SPD πινάκων, με ενδεικτικές εναλλακτικές μετρικές να είναι η Log-Euclidean, η Log-Cholesky, η Bures-Wasserstein και άλλες [53], [54].

4 ΚΕΦΑΛΑΙΟ 4^ο : Αλγόριθμοι στον Ευκλείδειο και Riemannian Χώρο

Μέχρι τώρα, σε όλη την έκταση αυτής της εργασίας περιεγράφηκαν μια πληθώρα από θεωρητικές έννοιες οι οποίες ξεκινούν από το πεδίο της Επεξεργασίας Εικόνας και χρησιμοποιώντας το μαθηματικό υπόβαθρο μετουσιώνονται σε πρακτικές εφαρμογές στο πεδίο της Υπολογιστικής Όρασης με την χρήση στατιστικών εργαλείων της Μηχανικής Μάθησης. Το παρόν κεφάλαιο αποσκοπεί στο να αποτυπώσει το πρακτικό κομμάτι αυτής της διπλωματικής μελετώντας αλγορίθμους τόσο στον Ευκλείδειο όσο και στον Riemannian Χώρο. Η μελέτη πραγματοποιείται με τρόπο αλυσιδωτό και συνεκτικό ξεκινώντας από τον Ευκλείδειο Χώρο με την απόδοση του μοντέλου του Bag of Visual Words (ή αλλιώς Bag of Features), το οποίο αποτέλεσε μια σημαντική καινοτομία της εποχής του 2000 αποδίδοντας την ιστογραμματική αναπαράσταση των εικόνων και την χρήση αυτών σε μοντέλα ταξινόμησης και συνεχίζει με την αποτύπωση της Χωρικής Πυραμίδας Αντιστοιχιών (*Spatial Pyramid Matching*), η οποία ενισχύει το αρχικό μοντέλο του Bag of Visual Words αποτυπώνοντας καλύτερα την χωρική πληροφορία των εικόνων που το πρώτο αδυνατεί να συλλάβει. Αλλάζοντας μεθοδολογία στατιστικής εκπροσώπησης, στο προσκήνιο τοποθετούνται οι πυρήνες Fisher, οι οποίοι μπορούν και αυτοί να ιδωθούν ως μια επέκταση του Bag of Visual Words με την διαφορά ότι σε αυτά αξιοποιούνται τα στατιστικά εργαλεία του μοντέλου των GMM αποτυπώνοντας έμπρακτα την στατιστική υπεροχή των προαναφερόμενων συναρτήσεων του *k-means*,

ενώ τέλος η περιγραφή του Ευκλείδειου Χώρου ολοκληρώνεται με την σκιαγράφηση των Διανυσμάτων των Τοπικά Αθροιζομένων Περιγραφέων (*Vectors of Locally Aggregated Descriptors-VLAD*), τα οποία στηριζόμενα τόσο στο Bag of Visual Words όσο και στους πυρήνες Fisher εισάγουν την έννοια της συσπείρωσης των διαφορών των εικόνων που έχουν βρεθεί ότι είναι πιο κοντά στο υποκείμενο κέντρο υπό εξέταση, όπως το τελευταίο προκύπτει μέσω εφαρμογής του *k-means*. Η διαδικασία ολοκληρώνεται με την αποτύπωση των καινοτόμων προσεγγίσεων στον Riemannian Χώρο και την αποτύπωση της ενίσχυσης της περιγραφικότητας.

4.1 Αλγόριθμοι στον Ευκλείδειο Χώρο

Η παρούσα υποενότητα ασχολείται με την παράθεση αλγορίθμων που επωφελούνται από την τροφοδότηση δεδομένων που εντοπίζονται στον Ευκλείδειο Χώρο για την διαδικασία της ταξινόμησης των εικόνων, ενώ συμπληρωματικά μέσω της παράθεσης του τρόπου λειτουργίας τους, υπογραμμίζεται μια συλλογή από τεχνικές που κωδικοποιούν την πληροφορία με τρόπο ο οποίος μειώνει τον υπολογιστικό φόρτο μα συνάμα διατηρεί την περιγραφικότητα της προσέγγισης. Στους προαναφερόμενους εντάσσονται ο αλγόριθμος Bag of Visual Words, η Πυραμίδα Χωρικής Αντιστοίχισης (*Spatial Pyramid Matching*), οι πυρήνες Fisher και τα Διανύσματα των Τοπικά Συσπειρωμένων Περιγραφέων (*Vectors of Locally Aggregated Descriptors-VLAD*). Τα πειράματα που παρουσιάζονται έχουν ως κύριο θεματικό τους άξονα την προσπάθεια επιτυχούς κατηγοριοποίησης των εικόνων στο σύνολο των ομάδων που ανήκουν, ενώ οι βάσεις δεδομένων των εικόνων προέρχονται από δυο διακεκριμένες πηγές, οι οποίες καθορίζουν και το πρόβλημα προς επίλυση. Συγκεκριμένα:

- Η πρώτη βάση δεδομένων προέρχεται από τα αρχεία που είναι διαθέσιμα από το μάθημα: «*Recognizing and Learning Object Categories*» των Li Fei-Fei, Rob Fergus και Antonio Torralba στην υποκατηγορία: «*Bag of Words models*», όπου παρατίθεται το μοντέλο που είχε συνταχθεί για την παρουσίαση στο [ICCV του 2005](#) για το μοντέλο του BoVW. Η φύση του προβλήματος έγκειται στον διαχωρισμό των εικόνων ανάλογα με το αν είναι σκηνές που αποτυπώνουν παρασκήνιο ή αν είναι σκηνές που αποτυπώνουν πρόσωπα, κατατάσσοντάς το αυτόματα σε ένα δυαδικό πρόβλημα.
- Η δεύτερη βάση δεδομένων προέρχεται από την επίσημη ιστοσελίδα της [Svetlana Lazebnik](#), όπου μπορεί να εντοπιστεί το άρθρο της σχετικά με την Χωρική Πυραμίδα Αντιστοιχίσεων. Η φύση του συγκεκριμένου προβλήματος αφορά την προσπάθεια αναγνώρισης μιας πληθώρας σκηνών με βάση το περιεχόμενό τους (π.χ αν μια σκηνή αποτυπώνει βουνά, ή αν μια σκηνή αποτυπώνει κάποιο κτήριο ή αν μια σκηνή αποτυπώνει αυτοκινητόδρομο) ανάγοντας αυτόματα το πρόβλημα σε αυτό των πολλών κλάσεων (*multi-class problem*).

Οι παραπάνω δυο βάσεις αποτελούν το στήριγμα της περάτωσης των αλγορίθμων που θα μελετηθούν στην παρούσα υποενοότητα με την βασική σημείωση ότι το μοντέλο του BoVW αξιολογείται μόνο στην πρώτη, ενώ πραγματοποιείται μια συγκριτική ανάλυση ανάμεσα στο πως οι υπόλοιποι αλγόριθμοι αποδίδουν στην δεύτερη βάση. Σε όλα τα πειράματα, το μέτρο της αποτελεσματικότητας είναι η ακρίβεια (*accuracy*).

4.1.1 Bag of Features

Η έννοια του μοντέλου «Bag of Words» (BoW), έχει τις ρίζες της στον τομέα της επεξεργασίας φυσικής γλώσσας (NLP) και εμφανίστηκε για πρώτη φορά στις αρχές της δεκαετίας του 1950 στο έργο του Zellig Harris: «*Distributional Structure*» [83], με την καινοτομία να έγκειται στο γεγονός ότι υπήρξε παντελής αδιαφορία για την γραμματική και την σειρά των λέξεων αντιμετωπίζοντας το εκάστοτε έγγραφο ως μια συλλογή (*Bag*) μεμονωμένων λέξεων. Αυτή η απλοϊκή μα συνάμα ισχυρή προσέγγιση, πλαισιωμένη πάνω στην κατανομητική υπόθεση (*distributional hypothesis*)⁴¹, έδωσε την δυνατότητα δημιουργίας ενός διανύσματος χαρακτηριστικών που αντιπροσωπεύει την συχνότητα εμφάνισης των λέξεων οδηγώντας σε μια διαισθητική απόδοση κοινού νοήματος. Βασιζόμενος στο έργο του Harris, ο Luhn ανέλαβε την σκυτάλη επεκτείνοντας περαιτέρω αυτή την ιδέα με την πρόταση μιας στατιστικής προσέγγισης για την αυτόματη περίληψη κειμένου, δίνοντας έμφαση στη σημασία της συχνότητας των λέξεων για την εξαγωγή βασικών πληροφοριών από τα έγγραφα [83], [84], [85], [86].

Με την ανάπτυξη των υπολογιστικών δυνατοτήτων, το συγκεκριμένο μοντέλο έλαβε σημαντική προβολή στις δεκαετίες του 1970 και 1980 με την εμφάνιση πιο εξελιγμένων συστημάτων ανάκτησης πληροφοριών. Η εισαγωγή του μοντέλου διανυσματικού χώρου (VSM) από τους Salton et al. [87] το 1975, βασίστηκε στις αρχές του BoW ενώ έφερε επανάσταση στην αναπαράσταση κειμένου, επιτρέποντας τη σύγκριση εγγράφων με βάση την ομοιότητα συνημίτονου σε έναν χώρο υψηλών διαστάσεων, διευκολύνοντας την ακριβέστερη και αποτελεσματικότερη ανάκτηση σχετικών εγγράφων από μεγάλα σώματα κειμένων.

Όπως όμως όλα τα πράγματα στην ζωή έτσι και αυτό, παρά την απλότητά του εμφανίζει αρκετούς περιορισμούς με τον πιο σημαντικό να είναι η αδυναμία του να συλλάβει τη σειρά των λέξεων και τα συμφραζόμενα, οδηγώντας σε πιθανή απώλεια σημασιολογικής πληροφορίας. Για την αντιμετώπιση αυτών των ζητημάτων, οι ερευνητές έχουν αναπτύξει διάφορες επεκτάσεις και εναλλακτικές λύσεις, όπως τα *n-grams* και το *TF-IDF* (*Term Frequency-Inverse Document Frequency*), τα οποία ενσωματώνουν κάποιες πληροφορίες σχετικά με το πλαίσιο, λαμβάνοντας

⁴¹ Η κατανομητική υπόθεση είναι μια έννοια που εντοπίζεται στην φυσική επεξεργασία γλώσσας (*Natural Language Processing- NLP*) και αναφέρεται στο φαινόμενο κατά το οποίο λέξεις που εμφανίζονται σε κοινό περιεχόμενο τείνουν να έχουν την ίδια έννοια.

υπόψη ζεύγη λέξεων και προσαρμόζοντας τις συχνότητες των λέξεων με βάση τη σπανιότητά τους στα έγγραφα, αντίστοιχα με την επιστράτευση μιας σταθμισμένης στρατηγικής. Η μετάβαση από την ανάλυση κειμένου στην οπτική επεξεργασία δεδομένων σηματοδότησε μια σημαντική εξέλιξη στην εφαρμογή του μοντέλου BoW.

Εμπνευσμένοι από την επιτυχία του στο τομέα του NLP, οι ερευνητές της Υπολογιστικής Όρασης άρχισαν να διερευνούν τις δυνατότητες του μοντέλου για την αναπαράσταση και ταξινόμηση εικόνων. Με οδηγό το έργο: «*Video Google: A Text Retrieval Approach to Object Matching in Videos*» [88] των Sivic και Zisserman, «γεννήθηκε» η έννοια του Bag of Visual Words όπως είναι γνωστή μέχρι και σήμερα με τις βασικές αρχές του αλγορίθμου να είναι παρόμοιες με αυτές του προκατόχου του ενώ το μόνο πράγμα που αλλάζει είναι πως η τροφοδότηση του μοντέλου γίνεται με οπτικά δεδομένα.

Τυπικά η διαδικασία ξεκινά με την εξαγωγή τοπικών χαρακτηριστικών από μια εικόνα χρησιμοποιώντας τεχνικές όπως ο Μετασχηματισμός Αναλλοίωτων Χαρακτηριστικών Κλίμακας (*Scale-Invariant Feature Transform-SIFT* [27]) ή το *Speeded-Up Robust Features-SURF*. Αυτά τα τοπικά χαρακτηριστικά, γνωστά και ως σημεία κλειδιά ή σημεία ενδιαφέροντος (*key-points*), αποτυπώνουν διακριτά μοτίβα, όπως ακμές, γωνίες ή κηλίδες τα οποία είναι μοναδικά σε κάθε εικόνα και εντοπίζονται σε συγκεκριμένες περιοχές αυτών.

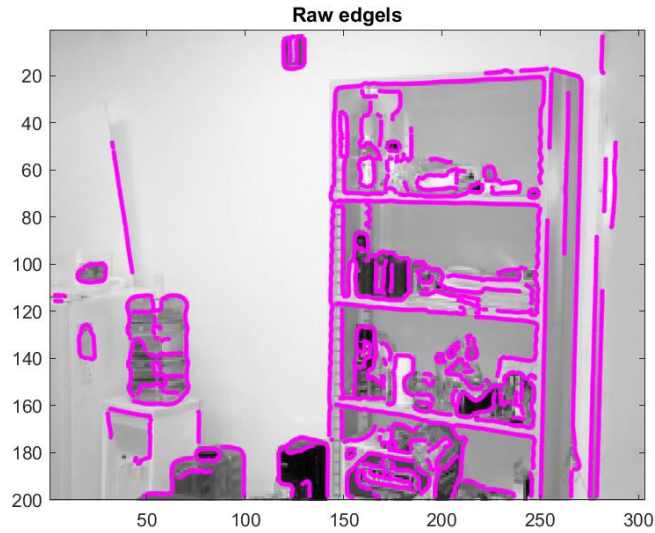
Μόλις ολοκληρωθεί η διαδικασία της εξαγωγής χαρακτηριστικών, σειρά έχει η κβάντισή τους σε ένα διακριτό σύνολο Οπτικών Λέξεων μέσω ενός αλγορίθμου συσταδοποίησης. Με τον *K-means* να είναι ο πιο συχνά αξιοποιήσιμος αλγόριθμος μη-επιβλεπόμενης μάθησης, επιτυγχάνεται τμηματοποίηση του Χώρου χαρακτηριστικών σε *k*-συστάδες, όπου κάθε κέντρο συστάδας αντιπροσωπεύει μια Οπτική Λέξη. Η συλλογή αυτών των Οπτικών Λέξεων σε έναν ενιαίο πίνακα, δομούν αυτό που είναι γνωστό ως *Οπτικό Λεξικό* (*Visual Vocabulary, Dictionary* ή απλά *Vocabulary*) με διαστατικότητα που καθορίζεται από το σύνολο των εικόνων και τον αριθμό των κέντρων/Οπτικών Λέξεων που επιλέγει ο χρήστης.

Ο αλγόριθμος ολοκληρώνεται με την ιστογραμματική αναπαράσταση των εικόνων και την απόδοση της τελικής διανυσματικής τους μορφής με ποσοστιαία φύση. Για να υλοποιηθεί το προκείμενο κομμάτι, υπολογίζονται οι τυπικές Ευκλείδειες αποστάσεις ανά ζεύγη ανάμεσα στον σχηματιζόμενο Λεξικό και τον πίνακα χαρακτηριστικών της εκάστοτε εικόνας εντοπίζοντας εκείνες με την μικρότερη τιμή. Η καταγραφή της συχνότητας εμφάνισης των δεικτών των διακεκριμένων κέντρων που εμφανίζονται σε κάθε εικόνα διαιρούμενες με το συνολικό πλήθος τους αποδίδει την ιστογραμματική φύση εκπεφρασμένη σε διανυσματική μορφή που οριοθετείται από την μονάδα.

Με την παραπάνω χονδροειδή αποτύπωση του αλγορίθμου του BoVW, είναι πλέον στιγμή να περιγραφεί το υποκείμενο πείραμα. Όπως ήδη αναφέρθηκε στην εισαγωγική παράγραφο αυτής της υποενότητας, η βάση δεδομένων που χρησιμοποιείται προέρχεται από τον κώδικα που συντάχθηκε για το ICCV του 2005, όπου το πρόβλημα αφορούσε την δυαδική ταξινόμηση των υποκείμενων εικόνων στο αν αποτυπώνουν παρασκήνιο ή κάποιο πρόσωπο. Κατά την εξέταση των αποκομμάτων κώδικα που είναι διαθέσιμα, εντοπίστηκε από τον συγγραφέα πως οι ερευνητές αφενός παρέμειναν πιστοί στο γενικό διάγραμμα που είθισται να ακολουθήσει κάποιος για να αναπαράγει ξανά τον αλγόριθμο αλλά ενίσχυσαν ένα βασικό κομμάτι αυτού· την εξαγωγή χαρακτηριστικών.

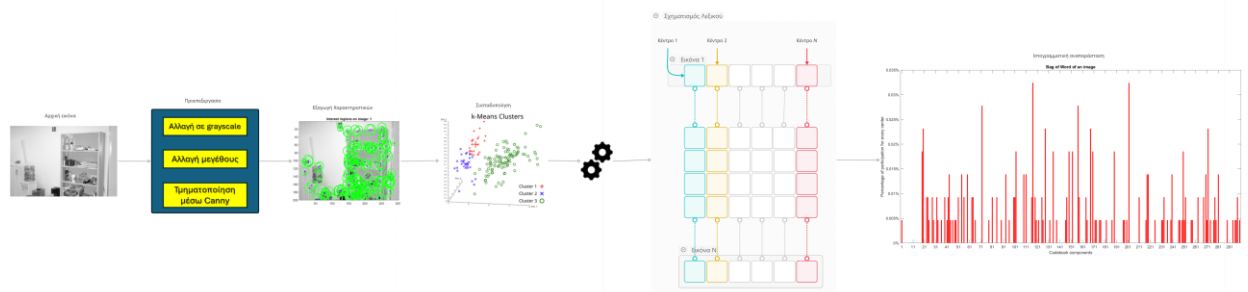
Η ισχυρά ευρετική φύση του SIFT αποτελεί βασική πτυχή του σχεδιασμού του, επιτρέποντάς του να ανιχνεύει και να περιγράφει αποτελεσματικά και αποδοτικά χαρακτηριστικά σε μια μεγάλη ποικιλία εικόνων. Κι ενώ αυτά τα χαρακτηριστικά τον καθιστούν ένα ισχυρό και ευέλικτο εργαλείο, συνάμα εισάγουν κάποιους βασικούς περιορισμούς όπως το γεγονός ότι η απόδοσή του μπορεί να εξαρτάται από συγκεκριμένες ρυθμίσεις των παραμέτρων και μπορεί να απαιτείται εμπειρική ρύθμιση για την επίτευξη των καλύτερων αποτελεσμάτων σε διάφορες εφαρμογές καθιστώντας την συνολική διαδικασία κάτι εν δυνάμει μη-τετριμμένο. Ένας τρόπος για να μπορέσει να αρθεί αυτό το τροχοπέδη είναι να επιστρατευτεί μια επιπρόσθετη μέθοδος επεξεργασίας πριν την εφαρμογή του SIFT, η οποία θα μπορέσει με κάποιον τρόπο να «υποδείξει» τα σημεία ενδιαφέροντος στην κάθε εικόνα, ενισχύοντας έτσι την αποδοτικότητα της εξαγωγής ισχυρών και περιγραφικών χαρακτηριστικών. Μια πολύ συχνή μέθοδος είναι αυτή του εντοπισμού ακμών Canny, ο οποίος ενισχύει την συνολική διαδικασία με τους παρακάτω τρόπους:

1. Ενίσχυση του εντοπισμού των σημείων ενδιαφέροντος: Με τον εντοπισμό ακμών Canny παρέχεται η δυνατότητα επισήμανσης των ακμών σε μια εικόνα, δηλαδή περιοχών με σημαντικές αλλαγές στην ένταση της φωτεινότητας. Αυτές οι περιοχές είναι τυπικά πλούσιες σε χαρακτηριστικά και μπορούν να αποδώσουν σημαντικές πληροφορίες κατά τον σχηματισμό των σημείων-κλειδιών του SIFT.
2. Μείωση θορύβου: Λόγω του γεγονότος ότι ένα από τα στάδια υλοποίησης του Canny περιλαμβάνει την χρήση Γκαουσιανών για την επιβολή θόλωσης της εικόνας, παρέχεται η δυνατότητα να μειωθεί ο επαγόμενος θόρυβος και να απορριφθούν επίπεδες περιοχές ή περιοχές χαμηλής αντίθεσης, εκλαμβάνοντας την εικόνα με μια πιο καθαρή μορφή. Όσο αντιφατικό κι αν φαίνεται, η υψηλή ανάλυση μιας εικόνας σε διαδικασίες που αποσκοπούν στον εντοπισμό αλλαγής φωτεινότητας εισάγει ψευδή τεχνουργήματα (*artifacts*) οι οποίες καταλήγουν να αποτελούν ζημιογόνες για το συνολικό έργο.



Εικόνα 39: Απεικόνιση του εντοπισμού ακμών με την χρήση του αλγορίθμου Canny.

Τέλος εφαρμόζονται και κάποιες ακόμη τυπικές διαδικασίες που εμφανίζονται συχνά στην Επεξεργασία Εικόνας όπως ο μετασχηματισμός των αρχικών έγχρωμων εικόνων σε grayscale καθώς και ο μετασχηματισμός των διαστάσεών τους σε μικρότερες. Οι λόγοι που επιτελούνται τέτοιες διαδικασίες είναι ώστε να μπορεί να μειωθεί η διαστατικότητα οδηγώντας στην αντίστοιχη μείωση του υπολογιστικού χρόνου. Με βάση το τελευταίο σχόλιο, η αποτύπωση του γενικού διαγράμματος σχηματισμού του BoVW, όπως αυτό έχει υλοποιηθεί για την συγκεκριμένη εφαρμογή που εξετάζεται, παρατίθεται στην *Εικόνα 40*:



Εικόνα 40: Απεικόνιση του διαγράμματος για την υλοποίηση του BoVW σχηματικού.

Με το τελευταίο βήμα του σχηματικού να αποδίδει την ιστογραμμική αναπαράσταση των εικόνων, αυτό που θα πρέπει να τονιστεί είναι ότι αφενός η αναπαράσταση αυτή είναι διανυσματικής μορφής αφετέρου όμως από προγραμματιστικής άποψης κανείς θα μπορούσε να τοποθετήσει τα διανύσματα αυτά σε έναν συνολικό πίνακα χωρίς να αλλάζει η πληροφορία που μεταφέρουν. Ο λόγος για τον οποίο γίνεται το τελευταίο βήμα είναι διότι οι περισσότεροι, αν όχι όλοι, οι ταξινομητές δέχονται ως ορίσματά τους εικόνες ή γενικά πίνακες $n \times m$ διαστάσεων, παρέχοντας την δυνατότητα ολιστικής επεξεργασίας των δεδομένων. Δεδομένης λοιπόν της μελέτης που πραγματοποιήθηκε στο κεφάλαιο των αλγορίθμων επιβλεπόμενης μάθησης, δεν αποτελεί έκπληξη πως ο ταξινομητής που επισπεύδεται για να μπορέσει να διαχωρίσει τα δεδομένα μεταξύ τους είναι τα SVM καθώς η αξιοποίηση αυτών για την επίλυση εγγενώς δυαδικών προβλημάτων είναι η τέλεια επιλογή. Το μέτρο

της αποτελεσματικότητας του αλγορίθμου επιλέγεται πως είναι η ακρίβεια (*accuracy*), το οποίο ορίζεται πως είναι το ποιοτικό μέγεθος εκείνο που φανερώνει πόσο συχνά είναι το μοντέλο ακριβές στις προβλέψεις που πραγματοποιεί. Εκπεφρασμένο με ποσοστιαία μορφή, ορίζεται ως [89], [90], [91], [92]:

$$\text{Ακρίβεια} = \frac{\text{Σωστές προβλέψεις}}{\text{Όλες οι προβλέψεις}} \quad (67)$$

Όπου οι σωστές προβλέψεις προκύπτουν από τον τυπικό Πίνακα Σύγχυσης (*Confusion Matrix*) ως το άθροισμα των *True Positive* και *True Negative* μπλοκ:

		Actual Values	
		Positive (1)	Negative (0)
Predicted Values	Positive (1)	TP	FP
	Negative (0)	FN	TN

Εικόνα 41: Γραφική αναπαράσταση ενός τυπικού Πίνακα Σύγχυσης⁴².

Ο ψευδοκώδικας που περιγράφει τον σχηματισμό του υπό μελέτη αλγορίθμου παρατίθεται στην συνέχεια. Προκειμένου να υπάρξει μια όντως διαισθητική εικόνα της αποτελεσματικότητάς του, το συνολικό πείραμα επαναλαμβάνεται 10 φορές ενώ για να διαφυλαχτεί η καλή επαναληψιμότητα αρχικοποιείται μια γεννήτρια παραγωγής τυχαίων αριθμών. Στο σημείο αυτό θα πρέπει να σημειωθεί κάτι πολύ βασικό! Στον τομέα της Μηχανικής Μάθησης ένας από τους πιο συχνούς τρόπους διαχωρισμού των δεδομένων είναι σε τρία βασικά σύνολα: το σύνολο εκπαίδευσης (*training set*), μέσω του οποίου το μοντέλο εκπαιδεύεται στα δεδομένα, το σύνολο της επαλήθευσης (*validation set*) το οποίο χρησιμοποιείται για να αποτρέψει το μοντέλο από την υπερπροσαρμογή (*overfitting*) και το σύνολο του ελέγχου (*testing*), το οποίο εισάγει νέα δεδομένα τα οποία το μοντέλο δεν έχει ξαναδεί ώστε να αξιολογηθεί η δυνατότητα γενίκευσής του. Όταν όμως εφαρμόζεται προσπάθεια βελτιστοποίησης υπερπαραμέτρων των μοντέλων, όπως γίνεται και στα πειράματα που θα παρουσιαστούν, προκειμένου να μπορέσουν να βρεθούν οι καλύτερες δυνατές ρυθμίσεις σε ένα εύλογο διάστημα εμπιστοσύνης και εκτέλεσης της βελτιστοποίησης, τα μοντέλα χρειάζεται να

⁴² Η εικόνα λήφθηκε από τον [σύνδεσμο](#).

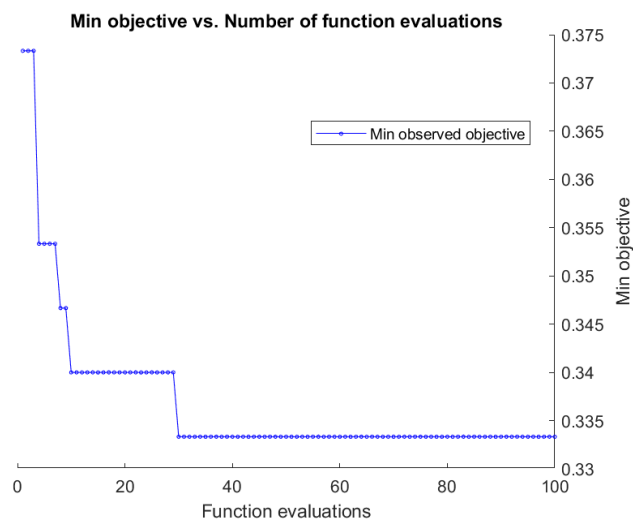
επαληθεύσουν τις εκτιμήσεις τους πάνω στα δεδομένα του προβλήματος. Με γνώμονα αυτήν την γνώση, επιλέγεται να συμπυκνωθούν τα δυο πρώτα σύνολα σε ένα, να χρησιμοποιηθεί η μέθοδος του *k-fold cross validation* και τα μοντέλα να ελέγχονται απευθείας στο σύνολο ελέγχου.

Algorithm 1 Bag of Features Model for Image Classification

```

1: Initialization:
2: Clear workspace, close figures, set random seed
3: Create directories for codebook, features, descriptors
4: Data Loading:
5: Load images from selected directory
6: Initialize imageDatastore
7: Preprocessing:
8: for each image in imageDatastore do
9:     Resize and convert to grayscale
10:    Detect interest points using edge sampling
11: end for
12: Feature Extraction (SIFT):
13: for each preprocessed image do
14:    Extract SIFT descriptors at interest points
15:    Save extracted features and points
16: end for
17: Codebook Formation:
18: Split data into training and testing sets
19: Use k-means clustering on training descriptors to form codebook
20: Save codebook
21: Vector Quantization:
22: for each image in training and testing sets do
23:    Calculate nearest codebook vector for each descriptor
24:    Create and normalize histogram of codebook vector occurrences
25:    Save quantized vector descriptors
26: end for
27: Training Classifier:
28: Train SVM classifier using training histograms
29: Optimize SVM hyperparameters
30: Evaluating Classifier:
31: Predict labels for testing images
32: Calculate accuracy and generate confusion matrix
    
```

Αλγόριθμος 4: Ψευδοκώδικας για τον αλγόριθμο BoVW.



Εικόνα 42: Αποτύπωση της διαδικασίας μείωσης του σφάλματος της αντικειμενικής συνάρτησης προς βελτιστοποίηση συναρτήσει του πλήθους επαληθεύσεών της.

Τα αποτελέσματα των επαναλήψεων εντοπίζονται στον Πίνακα 3:

Πίνακας 3: Αποτύπωση των τιμών της ακρίβειας για τις δέκα διακριτές επαναλήψεις.

Επανάληψη	1	2	3	4	5	6	7	8	9	10
Ακρίβεια (%)	72.00	68.00	72.00	66.00	66.00	70.00	68.00	70.00	74.00	66.00

Με μέση ακρίβεια που ανέρχεται στο 69.20% με το πέρας 10 επαναλήψεων, το μοντέλο του BoVW αποτελεί το εφιαλτήριο για την επέκταση της έρευνας αυτής της εργασίας.

4.1.2 Spatial Pyramid Matching

Η ευελιξία σε συνδυασμό με το χαμηλό υπολογιστικό κόστος και την ευκολία χρήσης με τα οποία συνοφάνεται, καθιστά τη χρήση της τεχνικής του BoVW μια δελεαστική προσέγγιση με αξιόλογα αποτελέσματα ακρίβειας, όπως μπορεί κανείς περίτρανα να εντοπίσει και από τα πειράματα που εντοπίζονται στον Πίνακα 3. Κι όμως, οι σύγχρονες εφαρμογές απαιτούν υψηλότερα επίπεδα αποδοτικότητα με τρόπους που μπορούν να ενισχυθούν από επιπρόσθετη περιγραφικότητα που επάγεται από την ενδογενή «σύλληψη» της δομής των αντικειμένων που αποτυπώνονται. Αυτό είναι και το βασικό σημείο στο οποίο το BoVW χωλαίνει! Η ολιστική προσέγγιση των σκηνών ή των αντικειμένων σε συνδυασμό με την άτακτη συλλογή των Οπτικών Λέξεων του Λεξικού, ενισχύει την αδυναμία του αλγορίθμου να «συλλάβει» την χωρική συσχέτιση ανάμεσα στα χαρακτηριστικά των εικόνων οδηγώντας σε σημαντική απώλεια εννοιολογικών πληροφοριών.

Παραμένοντας πιστοί στην ιστογραμμική αναπαράσταση των εικόνων και εμπνεόμενοι από το έργο των Graumann και Darrell: «*Discriminative classification with sets of image features*» [94] περί της πυραμιδικής αντιστοίχισης για την εύρεση αντιστοιχιών ανάμεσα σε σύνολα εικόνων, η ομάδα της Svetlana Lazebnik παρουσίασε το έργο της: «*Beyond Bag of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories*» [93] το 2006 φέρνοντας στο προσκήνιο μια επέκταση του BoVW. Βασική ιδέα του σχηματικού εκτέλεσης είναι η τμηματοποίηση της εκάστοτε εικόνας σε ένα σύνολο υπο-περιοχών με πιο λεπτομερή ανάλυση, όπου υπολογίζονται τα ιστογράμματα των τοπικών χαρακτηριστικών. Με την κατάτμηση της εικόνας να αποτυπώνεται ως διακριτά ακέραια επίπεδα και έχοντας σχηματίσει ένα σύνολο από ιστογράμματα, εφαρμόζεται η μέθοδος της ιστογραμμικής τομής (*histogram intersection*), όπου καταμετρείται ο αριθμός των αντιστοιχίσεων που έχουν εντοπιστεί σε κάθε επίπεδο. Η μέθοδος ολοκληρώνεται με στάθμιση των χαρακτηριστικών που βασίζεται στο εκάστοτε επίπεδο αναφοράς [93], [94].

Εμβαθύνοντας περαιτέρω για την αποσαφήνιση μετέωρων εννοιών, η κατάτμηση σε επίπεδα δεν είναι τίποτα περισσότερο από την διαδικασία υπέρθεσης ενός πλέγματος πάνω στην εικόνα με

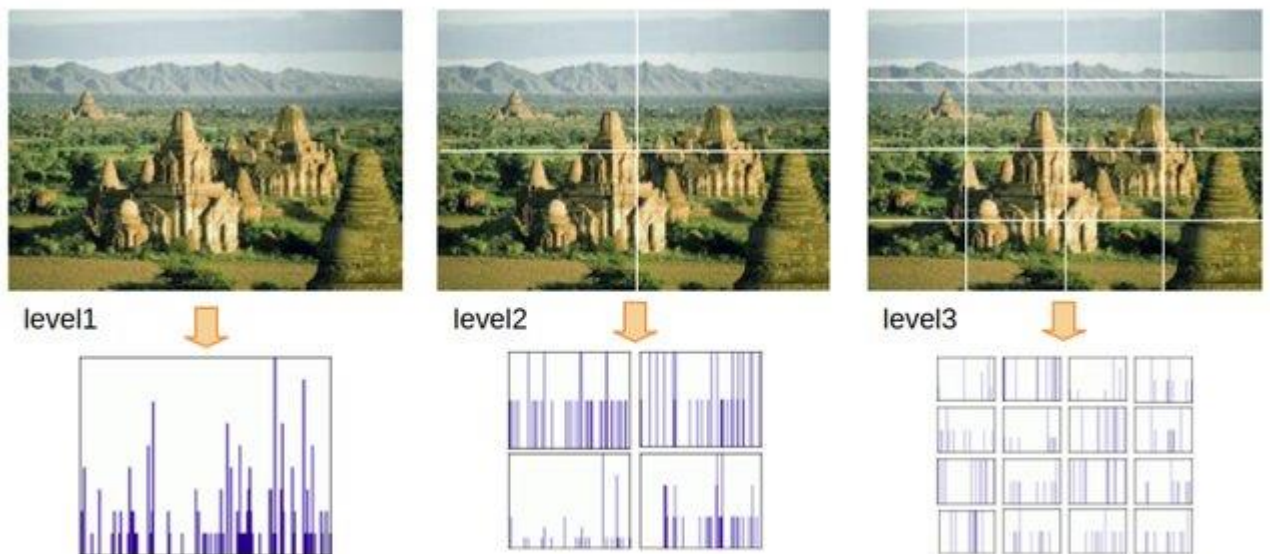
ευδιάκριτες τετραγωνικές περιοχές σταθερού μήκους. Κάθε μια από τις περιοχές αυτές, ονομάζεται *κελί* (*cell*) και το σύνολο αυτών δίνεται από τον τύπο [93]:

$$D = 2^{dl} \quad (68)$$

Όπου:

- d : είναι ο αριθμός των διαστάσεων του χώρου των χαρακτηριστικών (π.χ για τον SIFT τυπικά 128)
- l : το επίπεδο στο οποίο επιθυμείτε να καταμεριστεί η εικόνα.

Τυπικά, για τον διαισθητικό σχηματισμό του πλέγματος, το d συναιρείται μετασχηματίζοντας τον τύπο (68) σε 2^l . Με την αρίθμηση των επιπέδων να ξεκινά από το 0, για $l=0$ το επίπεδο στο οποίο διαχωρίζεται η εικόνα είναι ένα ($2^0 = 1$), αποτυπώνοντας την αρχική εικόνα χωρίς κάποιον διαχωρισμό, στο επίπεδο 1 η εικόνα διαχωρίζεται σε 4 τετράγωνα (ένα πλέγμα διαστάσεων 2×2), ενώ στο επίπεδο 2 η εικόνα διαχωρίζεται σε 8 τετράγωνα (ένα πλέγμα διαστάσεων 4×4). Η Εικόνα 43 απεικονίζει την περιγραφή που ακολούθησε παραπάνω [93]:



Εικόνα 43: Από αριστερά προς τα δεξιά⁴³

- A) Αποτύπωση της κατάτμησης της εικόνας σε 0 επίπεδα (Αρχική εικόνα)
 B) Αποτύπωση της κατάτμησης της εικόνας σε 1 επίπεδο.
 Γ) Αποτύπωση της κατάτμησης της εικόνας σε 2 επίπεδα.

Με την εκάστοτε επιλεγθείσα κατάτμηση της εικόνας να έχει ολοκληρωθεί, σε κάθε ένα από τα μπλοκ που συντελούν το πλέγμα υπολογίζονται τα ιστογράμματα της περιοχής ακολουθώντας την ίδια μεθοδολογία με αυτή του BoVW. Η κορωνίδα και αυτό που ξεχωρίζει το SPM, είναι ο τρόπος διαχείρισης των ιστογραμμάτων. Συγκεκριμένα, αν υποθεθεί ότι με H_X^l και H_Y^l σημειώνονται τα ιστογράμματα δυο d -διάστατων διανυσμάτων X και Y , τότε η *ιστογραμματοική τομή*, δηλαδή η εύρεση

⁴³ Η εικόνα λήφθηκε από τον [σύνδεσμο](#)

του αριθμού των αντιστοιχίσεων ανάμεσα στα ιστογράμματα των διανυσμάτων αυτών σε ένα συγκεκριμένο επίπεδο δίνεται από τον τύπο [93]:

$$I(H_X^l, H_Y^l) = \sum_{i=1}^D \min(H_X^l(i), H_Y^l(i)) \quad (69)$$

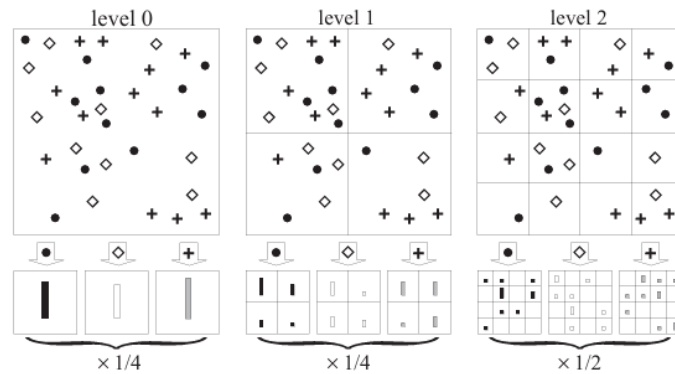
Αυτό όμως που θα πρέπει να ληφθεί πολύ σοβαρά υπόψιν είναι η ιεραρχική δόμηση της περιγραφόμενης διαδικασίας. Με την αλληλένδετη σύνδεση των επιπέδων μεταξύ τους (από τα πιο χονδροειδή προς τα πιο λεπτομερή επίπεδα), συχνά πυκνά εμφανίζονται χαρακτηριστικά που επαναλαμβάνονται καθιστώντας τις αντιστοιχίσεις που έχουν εντοπιστεί στο επίπεδο l ίδιες με εκείνες που έχουν βρεθεί και στα πιο λεπτομερή επίπεδα $l+1$. Αυτό καθιστά επιτακτική την απόδοση ενός «περιορισμού» ότι ο αριθμός των νέων αντιστοιχιών που εντοπίζονται σε ένα επίπεδο l δίνεται από τον τύπο [93]:

$$I_l - I_{l+1} \text{ για κάθε } l = 0, \dots, L - 1 \quad (70)$$

Το τελευταίο κομμάτι του αλγορίθμου βασίζεται σε μια πολύ φυσική πορεία σκέψης η οποία αποτυπώνεται ως εξής: «Δεδομένου ότι η κατάτμηση της εικόνας γίνεται με μετάβαση από τα πιο χονδροειδή προς τα πιο λεπτομερή επίπεδα, πως θα μπορούσε να αποτυπωθεί η έντονη και υψίστης σημασίας που τα τελευταία έχουν στην εύρεση νέων και πιο περιγραφικών χαρακτηριστικών». Η απάντηση σε αυτό προέρχεται από τον σχηματισμό ενός συστήματος στάθμισης (*weighting scheme*) το οποίο θα λαμβάνει υπόψιν του το εκάστοτε επίπεδο εφαρμόζοντας ποινή σε αντιστοιχίσεις που έχουν εντοπιστεί σε μεγαλύτερα κελιά λόγω του γεγονότος ότι περιλαμβάνουν αυξανόμενα ανόμοια χαρακτηριστικά. Το σύστημα αυτό λαμβάνει την μορφή πυρήνα (*πυρήνα Mercer*) και περιγράφεται από τον τύπο [93]:

$$k^L(X, Y) = J^\ell + \sum_{\ell=0}^{L-1} \frac{1}{2^{L-\ell}} (J^\ell - J^{\ell+1}) \quad (71)$$

$$\frac{1}{2^L} J^0 + \sum_{\ell=1}^L \frac{1}{2^{L-\ell+1}} J^\ell$$



Εικόνα 44: Από αριστερά προς τα δεξιά⁴⁴

- A) Αποτύπωση της εύρεσης τριών χαρακτηριστικών και της στάθμισής τους στο επίπεδο 0.
 B) Αποτύπωση της εύρεσης τριών χαρακτηριστικών και της στάθμισής τους στο επίπεδο 1.
 A) Αποτύπωση της εύρεσης τριών χαρακτηριστικών και της στάθμισής τους στο επίπεδο 2.

Με την ολοκληρωμένη παράθεση της λειτουργίας του SPM (*Spatial Pyramid Matching*), ακολουθεί η αποτύπωση των πειραμάτων που διεξήχθησαν. Ακολουθώντας πιστά το άρθρο της Lazebnik, και δεδομένων των διαθέσιμων αποκομμάτων κώδικα που μπορεί κανείς να βρει στην [σελίδα της](#), εφαρμόστηκαν κάποιες προσαρμογές οι οποίες αξιοποιούν την ενδογενή φύση του MATLAB για πράξεις μεταξύ πινάκων και διανυσμάτων με την ταυτόχρονη χρήση των καρτών γραφικών και της παράλληλης επεξεργασίας προκειμένου να εκτελεστούν 10 διακριτά πειράματα.

Στο κομμάτι της εξαγωγής χαρακτηριστικών χρησιμοποιήθηκε για ακόμη μια φορά ο αλγόριθμος SIFT με την εξαίρεση ότι η εφαρμογή του δεν πραγματοποιήθηκε με ολιστικό τρόπο στις εικόνες αλλά σε ένα σταθερό μη επικαλυπτόμενο παράθυρο εικονοστοιχείων 16×16 με απόσταση 8 pixel για την απαγωγή ισχυρών χαρακτηριστικών⁴⁵, ενώ τα πιο ασθενή χαρακτηριστικά εξήχθησαν χρησιμοποιώντας κατευθυνόμενα σημεία ακμών σε δύο κλίμακες και οκτώ κατευθύνσεις.

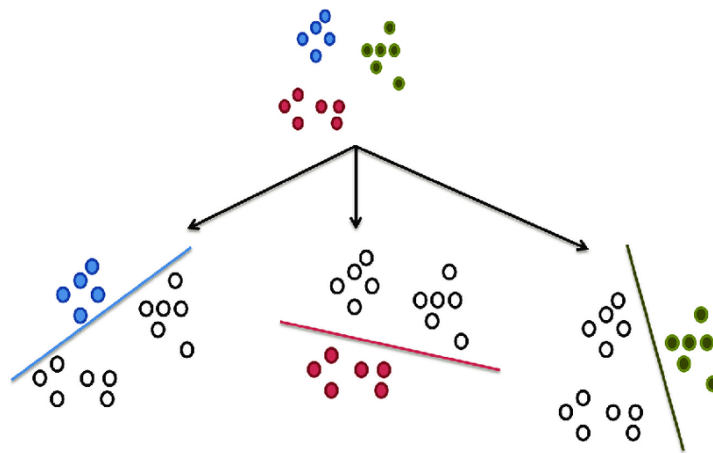
Στη συνέχεια, λόγω της υψηλής διαστατικότητας των χαρακτηριστικών, και λαμβάνοντας υπόψιν την προσπάθεια επιτάχυνσης λήψης αποτελεσμάτων, σχηματίστηκε μια παραλλαγή του *k-means*, η επονομαζόμενη *batch k-means*, καθώς παρατηρήθηκε ότι η χρήση της προεπιλεγόμενης συνάρτησης που παρέχει η MathWorks⁴⁶ συγκλίνει σε χρονικό και υπολογιστικό αδιέξοδο οδηγώντας το πρόγραμμα στο να κρασάρει αρκετά συχνά. Με την υλοποίηση λοιπόν του *batch k-means* παρέχεται η δυνατότητα τμηματικής τροφοδότησης του αλγορίθμου (*lazy loading*) με δεδομένα ώστε να αποφευχθεί υπερφόρτωση του συστήματος που μπορεί να οδηγήσει το πείραμα σε τερματισμό. Τα κέντρα που επιλέχθηκαν να χρησιμοποιηθούν είναι οι διακριτές ακέραιες τιμές 200 και 400.

⁴⁴ Η εικόνα λήφθηκε από την πηγή [93].

⁴⁵ Αυτή η μεθοδολογία συχνά ονομάζεται και πυκνή εξαγωγή χαρακτηριστικών (*dense feature extraction*).

⁴⁶ [k-means clustering - MATLAB kmeans \(mathworks.com\)](#)

Η μορφή πυρήνα της ιστογραμματικής τομής αποτέλεσε τον προπομπό της επιλογής των SVM ως τον αλγόριθμο επιβλεπόμενης μάθησης που θα χρησιμοποιηθεί για την προκειμένη εφαρμογή, καθώς όπως έχει ήδη επισημανθεί στο αντίστοιχο υπο-κεφάλαιο, η χρήση πυρήνων στα SVM ενισχύει την δυνατότητα προσέγγισης μη-γραμμικών προβλημάτων. Δεδομένων των πολλαπλών κλάσεων που είναι εμφανής στο προκειμένο πρόβλημα, γίνεται χρήση μιας τεχνικής που εμφανίζεται στους δυαδικούς ταξινομητές γνωστή και ως *one-vs-all*, σύμφωνα με την οποία, μια κλάση απομονώνεται (συνιστώντας την κλάση-1) και όλες οι υπόλοιπες θεωρούνται πως είναι η ετερόνυμη κλάση (συνιστώντας την κλάση-2). Μέσω λοιπόν μιας επαναληπτικής διαδικασίας που σαρώνει όλους τους δυνατούς συνδυασμούς, το συνολικό πρόβλημα πολλών κλάσεων ανάγεται σε ένα σύνολο πολλαπλών δυαδικών προβλημάτων.



Εικόνα 45: Σχηματική αναπαράσταση του *one-vs-all* σχηματικού⁴⁷

Διατηρώντας το σχηματικό που χρησιμοποιήθηκε για τον διαχωρισμό των δεδομένων στο πείραμα με το BoVW, και δεδομένου ότι γίνεται προσπάθεια βελτιστοποίησης των παραμέτρων των SVM, τα διαθέσιμα σύνολα είναι πάλι αυτά της εκπαίδευσης και του ελέγχου. Συγκεκριμένα ελέγχονται προς βελτιστοποίηση οι υπερπαραμέτροι *BoxConstraint*, *BoxConstraint* & *KernelScale* καθώς και όλοι οι υπερπαραμέτροι μαζί. Ο συνολικός πίνακας που αποτυπώνει τα πορίσματα του πειράματος παρατίθεται στην συνέχεια:

⁴⁷ Η εικόνα λήφθηκε από το [σύνδεσμο](#)

Πίνακας 4: Αποτύπωση του συνολικού πίνακα για τις μετρήσεις που πραγματοποιήθηκαν σε 10 διαφορετικά πειράματα. Με έντονα μαύρη γραφή υποδηλώνεται η καλύτερη τιμή ακρίβειας με τις παραμέτρους προς βελτιστοποίηση και τον αντίστοιχο αριθμό κέντρων.

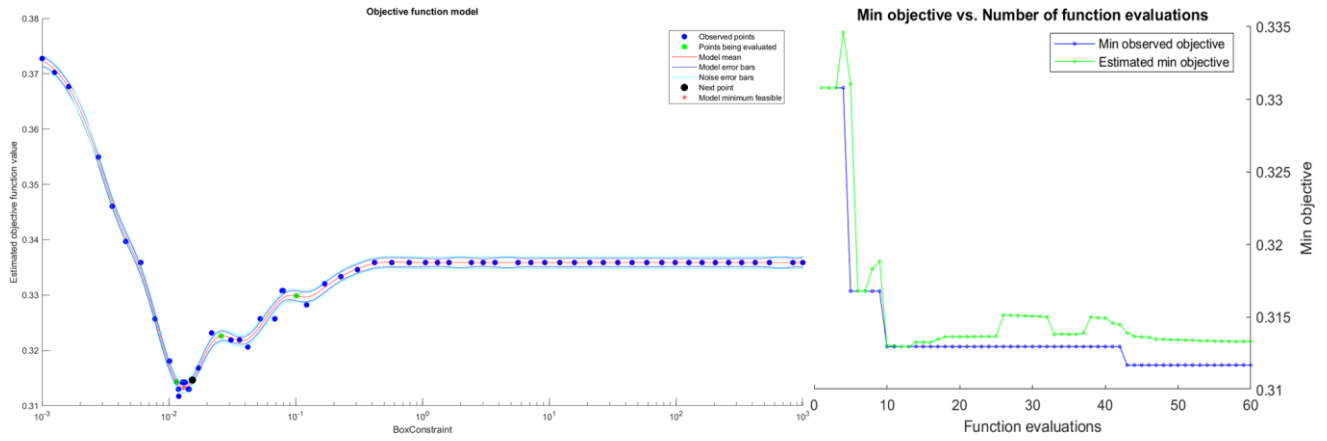
Επίπεδο Πυραμίδας	Αριθμός κέντρων	Παράμετρος προς βελτιστοποίηση	Μέση Ακρίβεια
2	200	BoxConstraint	71.7507
2	200	BoxConstraint & KernelScale	71.3947
2	200	All	71.6320
2	400	BoxConstraint	71.6024
2	400	BoxConstraint & KernelScale	71.8398
2	400	All	70.5935
3	200	BoxConstraint	76.0237
3	200	BoxConstraint & KernelScale	75.1335
3	200	All	76.0534
3	400	BoxConstraint	74.6588
3	400	BoxConstraint & KernelScale	75.9050
3	400	All	76.9139
4	200	BoxConstraint	73.2344
4	200	BoxConstraint & KernelScale	74.5994
4	200	All	75.5786
4	400	BoxConstraint	75.6083
4	400	BoxConstraint & KernelScale	75.0742
4	400	All	74.2433

Εστιάζοντας το ενδιαφέρον στη σύγκριση ανάμεσα στον Πίνακα 3 και Πίνακα 4 είναι εμφανές πως η μέθοδος του SPM όντως αποτύπωσε σταθερά υψηλότερα επίπεδα ακρίβειας σε σχέση με το BoVW αποδεικνύοντας περίτρανα ότι αποτελεί ένας άξιος διάδοχος για την χρήση σε εφαρμογές ταξινόμησης εικόνων με ιστογραμματικά διανυσματική φύση. Στην συνέχεια παρατίθενται ο γενικευμένος ψευδοκώδικας μαζί με τις απεικονίσεις της διαδικασίας βελτιστοποίησης των υπερπαραμέτρων *BoxConstraint* και *BoxConstraint & KernelScale*:

Algorithm 1 Image Classification Experiment

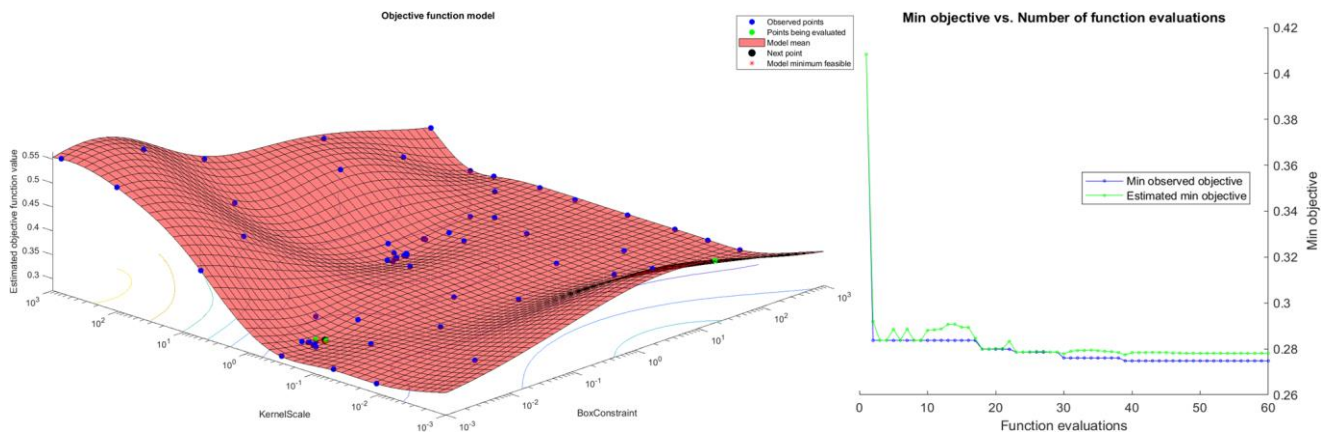
```
1: Clear workspace and delete existing parallel pool
2: Determine and display maximum number of computational threads
3: Create a parallel pool with half of the maximum workers
4: Set random number generator to default for reproducibility
5: Get images directory and form the imageDatastore
6: Select directory containing image data
7: Create imageDatastore with subfolders included and labels from folder
   names
8: Split imageDatastore into training and testing datasets
9: Initialize results table with columns for pyramid levels, number of centers,
   optimization parameter, and mean accuracy
10: Define experiment parameters: pyramid levels, number of centers, and hy-
    perparameters
11: for each pyramid level  $p$  do
12:   for each number of centers  $c$  do
13:     for each hyperparameter setting  $h$  do
14:       Initialize accuracy array
15:       for  $o = 1$  to 10 do
16:         Split data into training and testing sets
17:         Generate SIFT descriptors for training and testing images
18:         Form dictionary of visual words from training features
19:         Represent images as histograms of visual word occurrences
20:         Compile spatial pyramid vectors for training and testing im-
            ages
21:         Compute histogram intersection kernel matrices for training
            and testing
22:         Train SVM model using computed kernel matrices and speci-
            fied hyperparameters
23:         Predict labels for testing set and calculate accuracy
24:       end for
25:       Compute mean accuracy across all iterations
26:       Add results to results table
27:     end for
28:   end for
29: end for
30: Clean up and reformat results table for nicer view
31: Save results table to file
```

Αλγόριθμος 5: Ψευδοκώδικας για την εκτέλεση του πειράματος.



Εικόνα 46: Από αριστερά προς τα δεξιά

- A) Αποτύπωση της πορείας εκτίμησης των τιμών της αντικειμενικής συνάρτησης συναρτήσει της υπερπαραμέτρου προς βελτιστοποίηση ⁴⁸.
 B) Αποτύπωση της εξέλιξης της ελάχιστης τιμής της αντικειμενικής συνάρτησης συναρτήσει των αξιολογήσεών της ⁴⁹.



Εικόνα 47: Από αριστερά προς τα δεξιά

- A) Αποτύπωση του επιπέδου που ορίζεται για την εύρεση των τιμών της αντικειμενικής συνάρτησης προς βελτιστοποίηση εκπεφρασμένο σε λογαριθμική κλίμακα ⁵⁰.
 B) Αποτύπωση της εξέλιξης της ελάχιστης τιμής της αντικειμενικής συνάρτησης συναρτήσει των αξιολογήσεών της.

⁴⁸ Με μπλε κύκλους αποτυπώνονται τα σημεία που έχουν παρατηρηθεί, με πράσινο τα σημεία που έχουν ελεγχθεί, και με μαύρους κύκλους τα επόμενα σημεία προς αξιολόγηση. Με ανοιχτό κόκκινο εντοπίζεται η μέση τιμή του μοντέλου, με μπλε η εξέλιξη του σφάλματος και με ροζ οι γραμμές θορύβου του σφάλματος. Τέλος με αστερίσκο υποδεικνύεται το βέλτιστο δυνατό σημείο που εντοπίστηκε για τις προκειμένες αξιολογήσεις της αντικειμενικής συνάρτησης.

⁴⁹ Με μπλε γραμμή και κύκλο αποτυπώνεται η εξέλιξη του πραγματικού σφάλματος της αντικειμενικής συνάρτησης, ενώ με το αντίστοιχο πράσινο αποτυπώνεται η εξέλιξη του εκτιμώμενου σφάλματος αυτής.

⁵⁰ Με μπλε κύκλους αποτυπώνονται τα σημεία που έχουν παρατηρηθεί, με πράσινο τα σημεία που έχουν ελεγχθεί, και με μαύρους κύκλους τα επόμενα σημεία προς αξιολόγηση. Με κόκκινο παραλληλεπίπεδο ορίζεται η μέση τιμή του μοντέλου ενώ με αστερίσκο υποδεικνύεται το βέλτιστο δυνατό σημείο που εντοπίστηκε για τις προκειμένες αξιολογήσεις της αντικειμενικής συνάρτησης.

4.1.3 Fisher Kernels

Μέχρι πρότινος, η έρευνα που πραγματοποιήθηκε εστίαζε στην κωδικοποίηση των δεδομένων, δηλαδή των χαρακτηριστικών των εικόνων, σε διακριτές και συνάμα αποκλειστικές ομάδες μέσω της διαδικασίας του σχηματισμού του Λεξικού από τον *k-means*. Εγείρεται το ερώτημα αν όντως αυτή η μονοδιάστατη προσέγγιση κατηγοριοποίησης στον Χώρο των χαρακτηριστικών καταφέρνει να οδηγήσει στα καλύτερα δυνατά αποτελέσματα, ειδικά αν κανείς σκεφτεί ότι δυο εικόνες εμφανίζουν όμοια χαρακτηριστικά μεταξύ τους σε περιπτώσεις όπου η δειγματοληψία των τελευταίων γίνεται σε αρκετά χονδροειδές επίπεδο⁵¹. Μια εναλλακτική, είναι η ήπια ανάθεση, η οποία πλαισιώνεται από την χρήση των στατιστικών μεγεθών τα οποία μπορούν να παραχθούν από παραγωγικές μηχανές που συλλαμβάνουν την στατιστική φύση των δεδομένων, όπως είναι τα GMM. Με οδηγό σχηματισμού Λεξικών τα τελευταία και την δουλειά της ομάδας του *Perronnin και Dance*: «*Fisher Kernels on Visual Vocabularies for Image Categorization*» [95] σχετικά με τους πυρήνες Fisher, ανοίγει μια καινούργια και πιο εξελιγμένη προσέγγιση στον τομέα της κατηγοριοποίησης εικόνων.

Εμπνεόμενοι από την απλότητα του τρόπου λειτουργίας του μοντέλου του BoVW αλλά ταυτόχρονα προσπαθώντας να αποφύγουν επιπρόσθετα προβλήματα που εισήχθησαν από εναλλακτικές μεθόδους υλοποίησης αυτών όπως η αλλαγή της μετρικής για τον υπολογισμό της απόστασης των σημείων μεταξύ τους και την αποστασιοποίηση από την χρήση ενός ενιαίου Λεξικού, η ομάδα του Perronnin πρότεινε την εφαρμογή των πυρήνων Fisher ως έναν φυσικό τρόπο ενίσχυσης της διαστατικότητας των Λεξικών με την παράλληλη πιθανοτική κατανομή των χαρακτηριστικών στα κέντρα του Χώρου των χαρακτηριστικών [95].

Εμβαθύνοντας περαιτέρω, η βασική ιδέα ήταν ο χαρακτηρισμός ενός σήματος με ένα διάνυσμα κλίσης που προέρχεται από μια συνάρτηση πυκνότητας πιθανότητας (*probability density function-pdf*), η οποία καταφέρνει να συλλάβει την στατιστική φύση των δεδομένων. Έστω λοιπόν ότι με p συμβολίζεται μια *pdf* και οι παράμετροί της συμβολίζονται με λ , τότε ένα δείγμα, δηλαδή εδώ μια εικόνα, $X = \{x_t, t = 1, \dots, T\}$ εμφανίζει το παρακάτω διάνυσμα κλίσης [95]:

$$\nabla_{\lambda} \log p(X|\lambda) \quad (72)$$

Διασθητικά, η παράγωγος της λογαριθμικής πιθανοφάνειας περιγράφει την κατεύθυνση στην οποία οι παράμετροι θα πρέπει να τροποποιηθούν ώστε να μπορέσουν να προσεγγιστούν καλύτερα τα δεδομένα. Με τον πίνακα πληροφοριών Fisher να δίνεται από τον τύπο [95]:

$$F_{\lambda} = E_X[\nabla_{\lambda} \log p(X|\lambda) \nabla_{\lambda} \log p(X|\lambda)'] \quad (73)$$

⁵¹ Δεν είναι λανθασμένο κανείς να σκεφτεί ότι η εικόνα μιας γάτας και η εικόνα ενός σκύλου εμφανίζουν ίδια χαρακτηριστικά, όπως τα μάτια, η μουσούδα, οι τρίχες στις άκρες αυτής και άλλα.

και το κανονικοποιημένο διάνυσμα της κλίσης να δίνεται [95]:

$$F_{\lambda}^{-1/2} \nabla_{\lambda} \log p(X|\lambda) \quad (74)$$

Η ενημέρωση των παραμέτρων θα εκφράζεται από την κλίση τους υποδεικνύοντας ξανά την κατεύθυνση προς την εύρεση των καλύτερων τιμών. Με τις μερικές παραγώγους να εκφράζονται ως [95]:

$$\frac{\partial \mathcal{L}(X|\lambda)}{\partial w_i} = \sum_{t=1}^T \left[\frac{r_t(i)}{w_i} - \frac{r_t(1)}{w_1} \right] \text{ για } i \geq 2 \quad (75)$$

$$\frac{\partial \mathcal{L}(X|\lambda)}{\partial \mu_i^d} = \sum_{t=1}^T r_t(i) \left[\frac{x_t^d - \mu_i^d}{(\sigma_i^d)^2} \right] \quad (76)$$

$$\frac{\partial \mathcal{L}(X|\lambda)}{\partial \sigma_i^d} = \sum_{t=1}^T r_t(i) \left[\frac{(x_t^d - \mu_i^d)^2}{(\sigma_i^d)^3} - \frac{1}{\sigma_i^d} \right] \quad (77)$$

Κανείς μπορεί πολύ εύκολα να εντοπίσει πως η ποσότητα των βαρών ευθύνης είναι δεσπόζουσα σε όλους τους παραπάνω τύπους προοικονομώντας την άμεση χρήση των GMM για τον υπολογισμό πρώτα των απαραίτητων στατιστικών στοιχείων (*sufficient statistics*) και στην συνέχεια την χρήση των εκάστοτε μερικών παραγώγων για την υπόδειξη της κατεύθυνσης την οποία πρέπει να λάβουν ώστε να οδηγηθούν σε καλύτερα αποτελέσματα. Εν τέλει, η κλίση αποτυπώνεται ως ένα διάνυσμα, τα στοιχεία του οποίου προκύπτουν από την συνένωση των μερικών παραγώγων σε μια ενιαία μορφή [95].

Το επόμενο πρόβλημα που παρουσιάζεται είναι το πως θα μπορέσει να κανονικοποιηθεί το εύρος των διαφορετικών διαστάσεων των επιμέρους διανυσμάτων που προκύπτουν από τις μερικές παραγώγους. Για να επιλυθεί αυτό, υπολογίζεται η διαγώνιος του πίνακα πληροφοριών Fisher, F , μέσω των προσεγγίσεων [95]:

$$f_{w_i} = T \left(\frac{1}{w_i} + \frac{1}{w_1} \right) \quad (78)$$

$$f_{\mu_i^d} = \frac{T w_i}{(\sigma_i^d)^2} \quad (79)$$

$$f_{\sigma_i^d} = \frac{2T w_i}{(\sigma_i^d)^2} \quad (80)$$

Είναι εμφανές πως η εξελιγμένη μορφή των πυρήνων Fisher επιτρέπει την συμπερίληψη στατιστικών τόσο 1^{ης} όσο και 2^{ης} τάξης, όπως είναι η μέση τιμή και η συνδιακύμανση, ενισχύοντας την περιγραφικότητα της προσέγγισης κάτι το οποίο επίσης επιβεβαιώνεται κι από την οπτική του σχηματισμού ενός Λεξικού. Συγκεκριμένα, αν κανείς λάβει υπόψιν του πως μέσω της προσέγγισης

του BoVW ένα τυπικό Λεξικό έχει μέγεθος N , τότε η αντίστοιχη προσέγγιση των πυρήνων Fisher διαμορφώνει ένα Λεξικό μεγέθους $(2 \times D + 1) \times N - 1$, όπου [95]:

- D : είναι η διαστατικότητα των χαρακτηριστικών.
- N : είναι η διαστατικότητα των Γκαουσιανών συνιστωσών, δηλαδή ο αριθμός των κέντρων.

Η μείζον σημασία που εμφανίζει η μέθοδος των πυρήνων Fisher κι αυτό που την καθιστά τόσο ελκυστική είναι ότι επιτρέπει τον χαρακτηρισμό των εικόνων με διανύσματα υψηλής διαστατικότητας, ακόμη κι αν τα Λεξικά που δημιουργούνται συντελούνται από μικρό αριθμό Λέξεων. Με την ολοκλήρωση της παράθεσης λειτουργίας του αλγορίθμου, σειρά έχει η αποτύπωση των πειραμάτων που διεξήχθησαν.

Διατηρώντας την μεθοδολογία που επιτάχθηκε για την περάτωση των πειραμάτων σχετικά με τον αλγόριθμο του SPM άμεμπτη (δες σελ.108), παραμένοντας πιστοί στην πειραματική διάταξη που υλοποίησαν οι συγγραφείς του άρθρου αλλά κάνοντας χρήση της βάσης που παρουσιάστηκε στο πείραμα του BoVW, αποτυπώνεται ένα καινούργιο σχηματικό το οποίο επεκτείνει τον ερευνητικό χαρακτήρα της παρούσας διπλωματικής. Με τον απώτερο σκοπό να συνεχίζει να παραμένει η διερεύνηση τρόπων της μεταβολής της ακρίβειας προς υψηλότερες τιμές για τα μοντέλα κατηγοριοποίησης εικόνων, πλέον το ενδιαφέρον στρέφεται στους ίδιους τους αλγόριθμους μη-επιβλεπόμενης μάθησης που επιτάσσονται για να παράξουν το Οπτικό Λεξικό που θα χρησιμοποιηθεί και την προσπάθεια βελτίωσής τους. Κατά την πορεία της εργασίας, εντοπίστηκε σε πρώτο στάδιο η χρήση της εσωτερικής συνάρτησης του MATLAB, [fitgmdist](#), η οποία αναλαμβάνει να επιτελέσει την διαδικασία της υλοποίησης του EM αλγορίθμου για τον σχηματισμό των GMM παρέχοντας ταυτόχρονα μια πληθώρα από επιπρόσθετες επιλογές που μπορούν να επηρεάσουν την πορεία και την αποτελεσματικότητα του τελευταίου, όπως για παράδειγμα η επιλογή των σημείων έναρξης να γίνεται με τυχαίο τρόπο ή να επιλέγεται αν οι πίνακες συνδιακύμανσης είναι διαγώνιοι ή ακόμη και το επίπεδο ανοχής για την λήξη του αλγορίθμου στην προσπάθεια μείωσης της λογαριθμικής συνάρτησης. Αυτό όμως το οποίο δεν παρέχει η παραπάνω συνάρτηση είναι υποστήριξη της χρήσης GPU [95], [96].

Υπενθυμίζοντας ότι η πολυπλοκότητα των GMM εκφράζεται ως $\mathcal{O}(n \cdot d \cdot k \cdot i)$ όπου:

- n : είναι ο αριθμός των δεδομένων (συνήθως αναφέρεται στις σειρές του πίνακα δεδομένων).
- d : είναι οι διαστάσεις των δεδομένων (συνήθως αναφέρεται στις στήλες του πίνακα δεδομένων).
- k : είναι ο αριθμός των κέντρων που ορίζεται από τον χρήστη.
- i : είναι ο αριθμός των επαναλήψεων.

γίνεται εμφανές ότι το υπολογιστικό κόστος της επιτέλεσης του αλγορίθμου καθίσταται από προεπιλογή μη-τετριμμένο και σχεδόν ασύμφορο αν οι κάρτες γραφικών δεν μπορούν να αξιοποιηθούν. Για τον λόγο αυτό σχηματίστηκαν δυο επιπρόσθετες συναρτήσεις, μία η οποία αντιγράφει τον τρόπο λειτουργίας της *fitgmdist* αλλά για την υλοποίηση μόνο των διαγώνιων πινάκων αξιοποιώντας τις κάρτες γραφικών και τέλος η συνάρτηση που υλοποιεί τον *sEM* κάνοντας κι αυτή χρήση της GPU. Με τις διαδικασίες της Υπολογιστικής Όρασης αλλά και της Μηχανικής Μάθησης γενικότερα να διακατέχονται από συμβιβασμούς (*trade-offs*) που πρέπει να γίνουν ανάμεσα στην ταχύτητα λήψης αποτελεσμάτων και την απόδοση, ο πρώτος στόχος αυτής της παράλληλης διερεύνησης είναι να μελετηθεί ποια από τις τρεις συναρτήσεις καταφέρνει να ολοκληρωθεί πρώτη ενώ ο δεύτερος και ίσως πιο σημαντικός στόχος είναι να ερευνηθεί ποια κατάφερε να προσεγγίσει καλύτερα το πρόβλημα παράγοντας ποιοτικότερα αποτελέσματα.

Έτσι, για μείωση της διαστατικότητας των χαρακτηριστικών με την χρήση PCA τόσο των RGB όσο και των SIFT χαρακτηριστικών σε 50 και με την σταδιακή αύξηση των συνιστωσών των Γκαουσιανών μέχρι το 128, τα χρονικά αποτελέσματα που εξήχθησαν αποτυπώνονται στον Πίνακα 5:

Πίνακας 5: Αποτίπωση των χρόνων για την λήψη των αποτελεσμάτων από τις τρεις συναρτήσεις

Χρόνος για τα RGB δεδομένα (<i>fitgmdist</i>)	Χρόνος για τα SIFT δεδομένα (<i>fitgmdist</i>)	Χρόνος για τα RGB δεδομένα (<i>GMM_NV</i>)	Χρόνος για τα SIFT δεδομένα (<i>GMM_NV</i>)	Χρόνος για τα RGB δεδομένα (<i>sEM</i>)	Χρόνος για τα SIFT δεδομένα (<i>sEM</i>)
890 sec	Αποτυχία	$5.60 \times 10^3 \text{sec}$	$2.60 \times 10^5 \text{sec}$	$2.92 \times 10^5 \text{sec}$	$7.82 \times 10^4 \text{sec}$

```
Number of cluster:26
Error using gmcluster (line 197)
Ill-conditioned covariance created at iteration 89.
Error in gmdistribution.fit (line 102)
    gmcluster(X,k,start, reps, CovType, SharedCov, RegV, options, probtol);
Error in fitgmdist (line 135)
gm = gmdistribution.fit(X,k,varargin{:});
Error in Final_project (line 97)
    fitGMMs = fitgmdist(gather(Training_SIFT_data),i,"CovarianceType","diagonal","Replicates",1, ...
```

Εικόνα 48: Απεικόνιση της αποτυχίας της *fitgmdist* λόγω ill-conditioned πινάκων συνδιακόμανσης.

Αναλύοντας την δεύτερη στήλη του Πίνακα 5 διαφαίνεται πως επιβεβαιώνεται το φαινόμενο κατά το οποίο η *fitgmdist* αδυνατεί να περατώσει την διαδικασία που της ανατέθηκε, τουλάχιστον για τα δεδομένα που προέρχονται από τον SIFT, αλλά επιτυγχάνει να αποδώσει για τα RGB δεδομένα πολύ πιο γρήγορα συγκριτικά από τις υπόλοιπες δυο συναρτήσεις. Αυτό το φαινόμενο το οποίο εντοπίζεται

και στην σύγκριση ανάμεσα στο χρόνο που χρειάζεται η *GMM_NV* και η *sEM* για να επιστρέψουν αποτελέσματα οφείλεται στο υπολογιστικό φόρτο που εισάγεται από την μεταφορά των δεδομένων στις κάρτες γραφικών (*overhead*). Με τον γενικό κανόνα στο προγραμματισμό να επιτάσσει ότι όταν υπάρχουν δυο διαθέσιμοι τρόποι επεξεργασίας δεδομένων, τυπικά CPU και GPU, τα δεδομένα που δεν εμφανίζουν υψηλή διαστατικότητα είναι πιο προσοδοφόρο να αξιοποιούνται οι CPU έναντι των GPU διότι υπεισέρχεται *overhead* στην μεταφορά το οποίο δεν είναι αμελητέο, κάτι το οποίο επιβεβαιώνει και η *fitgmdist*. Στον αντίποδα, ο αναγνώστης καλείται να στρέψει την προσοχή του στις στήλες 4 και 6 του ίδιου πίνακα, όπου παρατηρείται πως τόσο η χρήση της GPU όσο και η χρήση του *sEM* κατάφερε να αποδώσει γρηγορότερα αποτελέσματα μέχρι και μια τάξη μικρότερη από αυτή της *GMM_NV*.

Εν κατακλείδι, λαμβάνοντας υπόψιν το γεγονός ότι η περάτωση τόσο της συγκριτικής ανάλυσης ανάμεσα στις συναρτήσεις όσο και η εξαγωγή των αποτελεσμάτων από τις ίδιες χρειάζεται τυπικά μέρες για να μπορέσει να ολοκληρωθεί, το πείραμα εκτελέστηκε μόνο μια φορά με τον *sEM* να αποδίδει την καλύτερη δυνατή ακρίβεια που εντοπίζεται ότι ανέρχεται στο **71.67%**.

4.1.4 Vectors of Locally Aggregated Descriptors (VLAD)

Η εμφάνιση των πυρήνων Fisher στο προσκήνιο της ταξινόμησης εικόνων ανέδειξε μια καινοτόμο και εξελιγμένη προσέγγιση η οποία καταφέρνει να ενισχύσει την περιγραφικότητα των μοντέλων αξιοποιώντας τα στατιστικά τους στοιχεία ενώ παράλληλα μεγεθύνει την διανυσματική αναπαράσταση των εικόνων συναρτήσει τόσο των διαστάσεων των χαρακτηριστικών όσο και του αριθμού των Γκαουσιανών συντελεστών που επιλέγονται.

Αναγνωρίζοντας την προσφορά της προσέγγισης των πυρήνων Fisher, λίγα χρόνια αργότερα το 2011 η ομάδα του Hervé Jégou στο έργο τους: «*Aggregating local descriptors into a compact image representation*» παρουσίασε μια νέα μέθοδο η οποία στηρίζεται στην ενίσχυση της διανυσματικής αναπαράστασης του Perronnin αλλά αντιμετωπίζει τελείως διαφορετικά την σχέση ανάμεσα στους περιγραφείς (*descriptors*) και στα κέντρα που έχουν δημιουργηθεί [97].

Συγκεκριμένα, η πρόταση της ομάδας ήταν η διανυσματική αναπαράσταση μιας εικόνας που συγχωνεύει του περιγραφείς βασιζόμενη σε ένα κριτήριο τοπικότητας του Χώρου των χαρακτηριστικών. Με την επιστροφή στην χρήση του *k-means* για την αυστηρή ανάθεση σε κέντρα, το πρώτο μέλημα είναι ο σχηματισμός ενός Λεξικού *k*-Οπτικών Λέξεων και η ανάθεση των περιγραφέων στο κοντινότερο κέντρο με βάση την Ευκλείδεια απόσταση. Υποθέτοντας τώρα ότι ένας τοπικός περιγραφέας είναι, όπως πάντα, *d*-διαστάσεων και συμβολίζεται με x , ενώ με u συμβολίζεται το παραγόμενο υπόλοιπο εικόνας (*image residual*). Τότε μπορεί να γραφεί πως, ένα στοιχείο του u λαμβάνεται ως το άθροισμα κατά μήκος όλων των περιγραφέων εικόνων ως [97]:

$$u_{i,j} = \sum_{x \text{ έτσι ώστε } NN(x)=c_i} x_i - c_{i,j} \quad (81)$$

Όπου:

- $u_{i,j}$ για $i = 1, \dots, k$ και $j = 1, \dots, d$: είναι το στοιχείο του υπολοίπου εικόνας που δεικτοδοτεί την i Οπτική Λέξη και το j στοιχείο του τοπικού περιγραφέα.
- $NN(x)$: συντομογραφία που αναφέρεται στον κοντινότερο γείτονα (*Nearest Neighbor*) του x .
- x_j : είναι το j -οστό στοιχείο του περιγραφέα x .
- $c_{i,j}$: είναι το j -οστό στοιχείο της αντίστοιχης Οπτικής Λέξης c_i .

Ως τελευταίο βήμα και προκειμένου να μπορέσουν οι τιμές να είναι όλες στο ίδιο εύρος, το διάνυσμα u κανονικοποιείται σύμφωνα με την $L2$ -νόρμα ως:

$$u := \frac{u}{\|u\|_2}$$

Η παραπάνω διαδικασία ακολουθεί τα χνάρια που έχουν θέσει οι πυρήνες Fisher καθώς καταφέρνει να αποδώσει μια διανυσματική αναπαράσταση διαστάσεων $D = k \times d$, όπου:

- k : είναι ο αριθμός των κέντρων και
- d : είναι η διαστατικότητα των περιγραφέων.

Η αξιοποίηση της διαφοράς μεταξύ των περιγραφέων και των συστάδων που τους αντιστοιχούν αποδίδει μια σημαντικά βελτιωμένη αναπαράσταση εικόνας, καθώς καταγράφει λεπτοφυείς, διακριτές λεπτομέρειες που διαφορετικά θα χάνονταν, διατηρεί τις ανεπαίσθητες παραλλαγές και τα μοναδικά χαρακτηριστικά των τοπικών χαρακτηριστικών προσφέροντας ένα πλουσιότερο και πιο κατατοπιστικό διάνυσμα σε σύγκριση με απλούστερες μεθόδους συνάθροισης και συντηρεί τις βασικές δομικές πληροφορίες βελτιώνοντας την ακρίβεια της ανάκτησης και την ανθεκτικότητα σε μεταβολές όπως οι αλλαγές του φωτισμού και οι μεταβολές της οπτικής γωνίας.

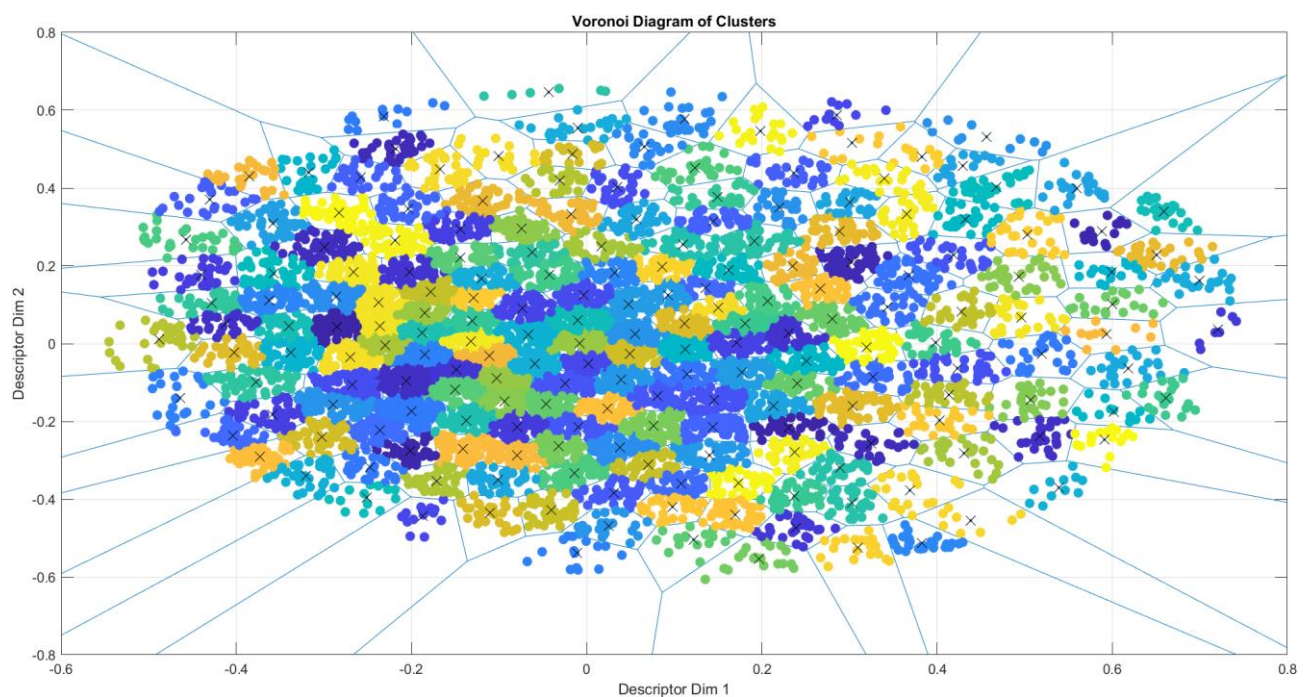
Η πειραματική διάταξη αποκλίνει σε σχέση με αυτή που χρησιμοποιείται στο άρθρο τόσο στο κομμάτι της εξαγωγής χαρακτηριστικών όπου χρησιμοποιείται απλά πυκνή σάρωση των εικόνων με βάση τον αλγόριθμο SIFT όσο και στο κομμάτι της επιλογής των βάσεων, όπου για τελευταία φορά γίνεται χρήση της βάσης των εικόνων που αναδεικνύουν διάφορες σκηνές, όπως παρουσιάστηκε αρχικά στο πείραμα με το SPM, ενώ οι συμπληρωματικές θεωρητικές έννοιες που παρουσιάζονται και αφορούν την δεικτοδότηση εικόνων θεωρούνται περιττές καθώς δεν μπορούν να αξιοποιηθούν στο κομμάτι της κατηγοριοποίησης των εικόνων. Ο Πίνακας 6 αποτυπώνει τα αποτελέσματα της

έρευνας που διεξήχθη για 10 διαφορετικές επαναλήψεις του πειράματος και βελτιστοποιώντας τρεις διαφορετικές υπερπαραμέτρους ανά πείραμα για τα SVM:

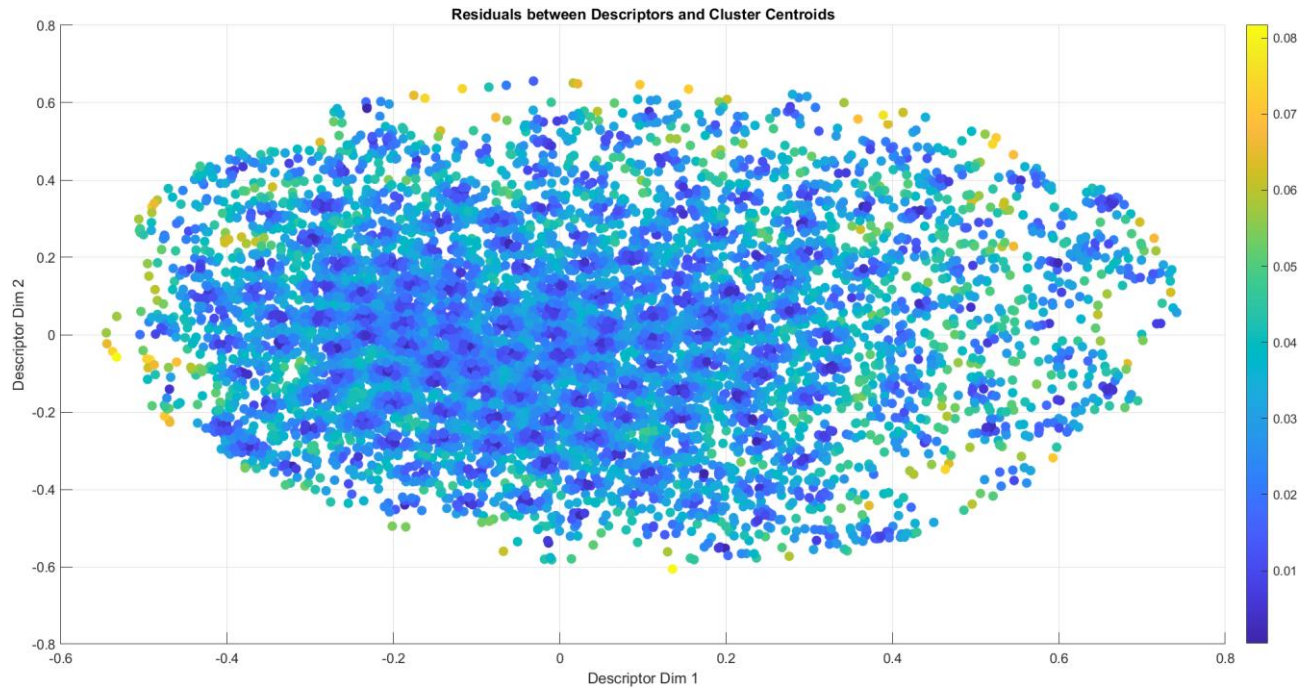
Πίνακας 6: Πίνακας αποτίμησης των μετρήσεων που συλλέχθηκαν από 10 διαφορετικά πειράματα για τα VLAD.
 Με έντονη γραφή σημειώνεται η υψηλότερη μέση ακρίβεια που εντοπίστηκε.

Πείραμα	BoxConstraint	BoxConstraint & KernelScale	All	Baseline
1	81.60	80.71	77.74	80.71
2	75.96	72.70	76.56	75.37
3	76.26	74.48	75.07	75.07
4	74.79	74.48	74.48	74.48
5	78.93	71.51	74.78	78.64
6	75.67	75.67	75.67	75.67
7	72.7	72.70	72.11	72.70
8	71.51	71.81	73.00	71.81
9	74.78	72.40	74.78	74.48
10	72.40	72.70	74.78	72.70
Μέση ακρίβεια (%)	75.46	73.92	74.90	75.16

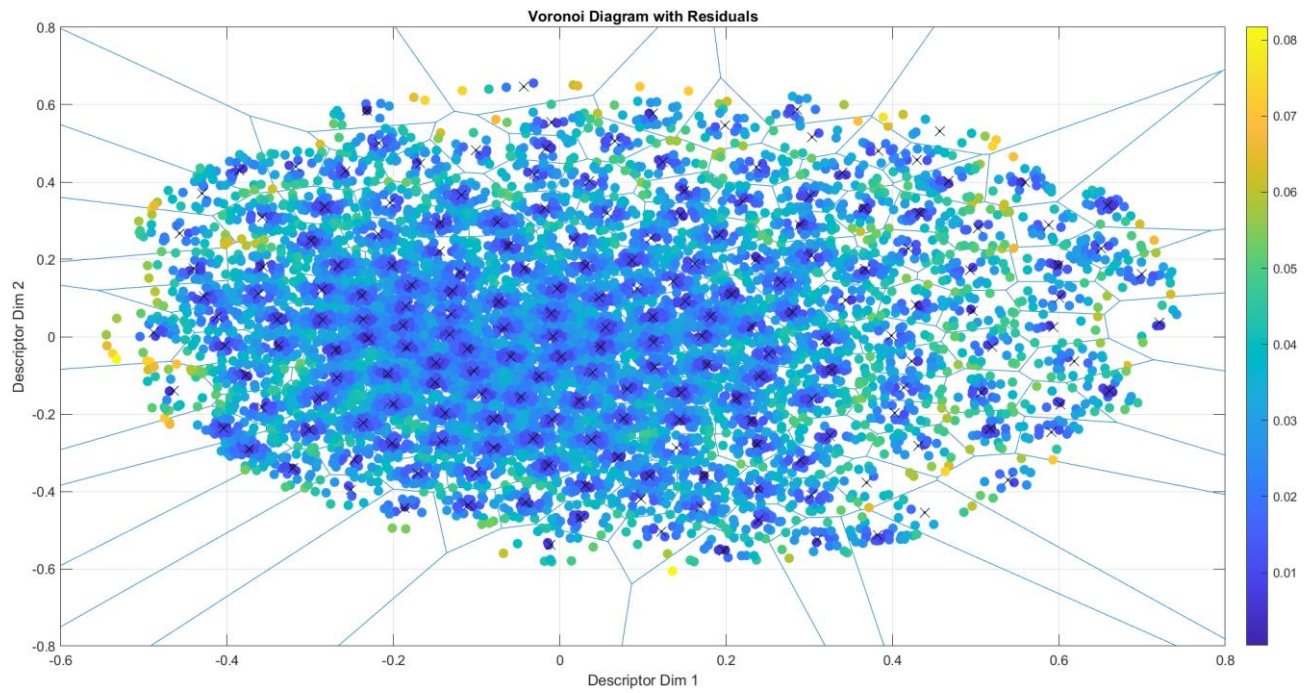
Στην συνέχεια παρατίθεται μια σειρά από απεικονίσεις, ως ένδειξη καλής πρακτική, για την καλύτερη κατανόηση των περιγραφόμενων εννοιών καθώς και ο ψευδοκώδικας υλοποίησης του πειράματος.



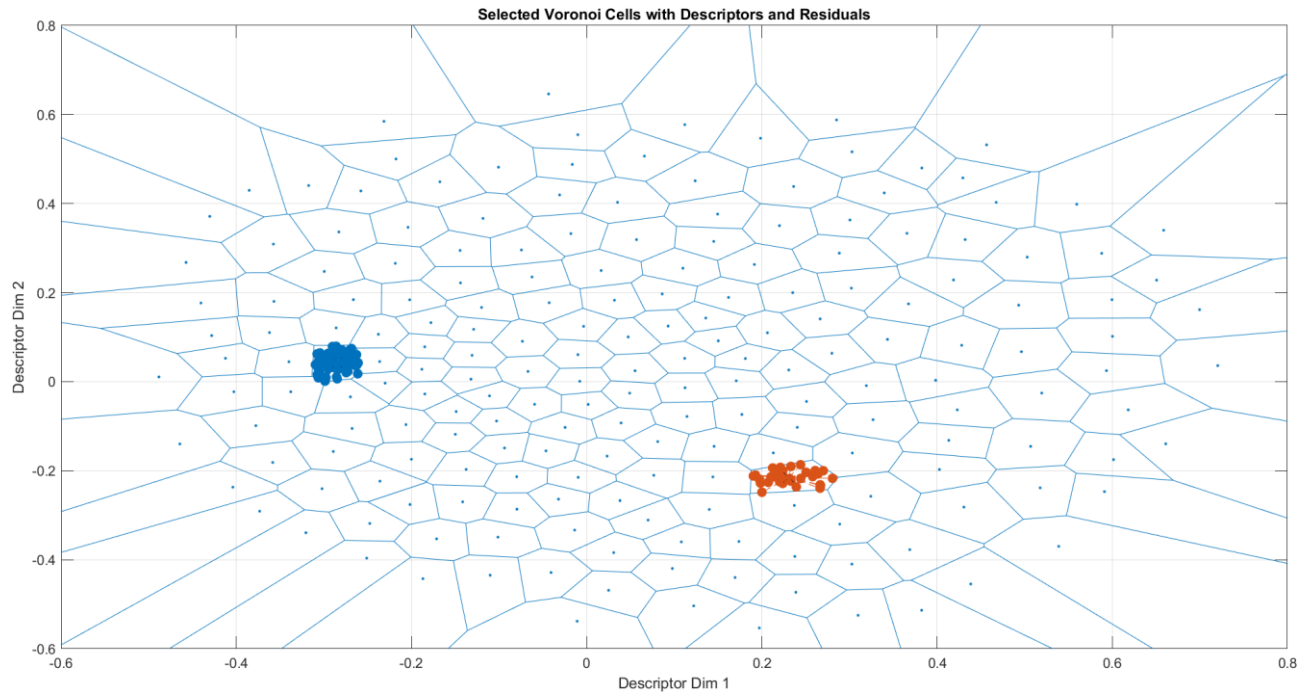
Εικόνα 49: Απεικόνιση του Voronoi διαγράμματος με την υπέρθεση των τοπικών περιγραφέων ως έγχρωμες κουκίδες και των κέντρων με μαύρους σταυρούς.



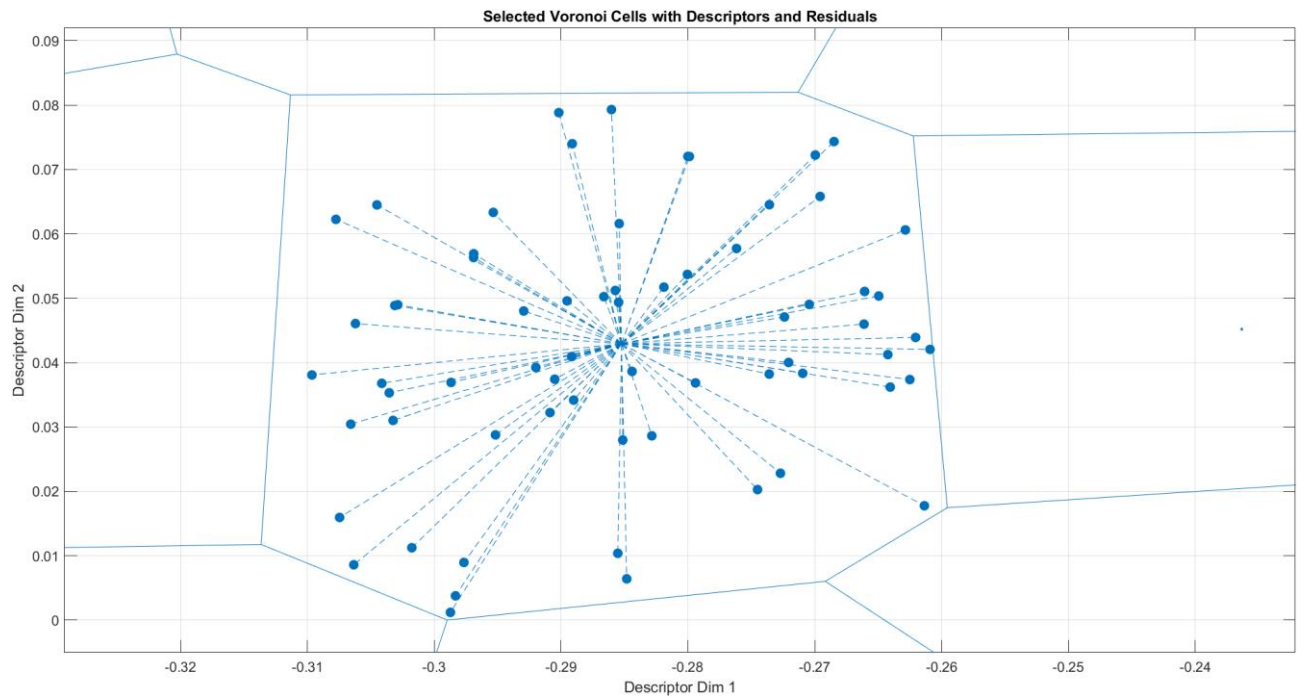
Εικόνα 50: Αποτύπωση των υπολοίπων ανάμεσα στα κέντρα και τους τοπικούς περιγραφείς αυτών.



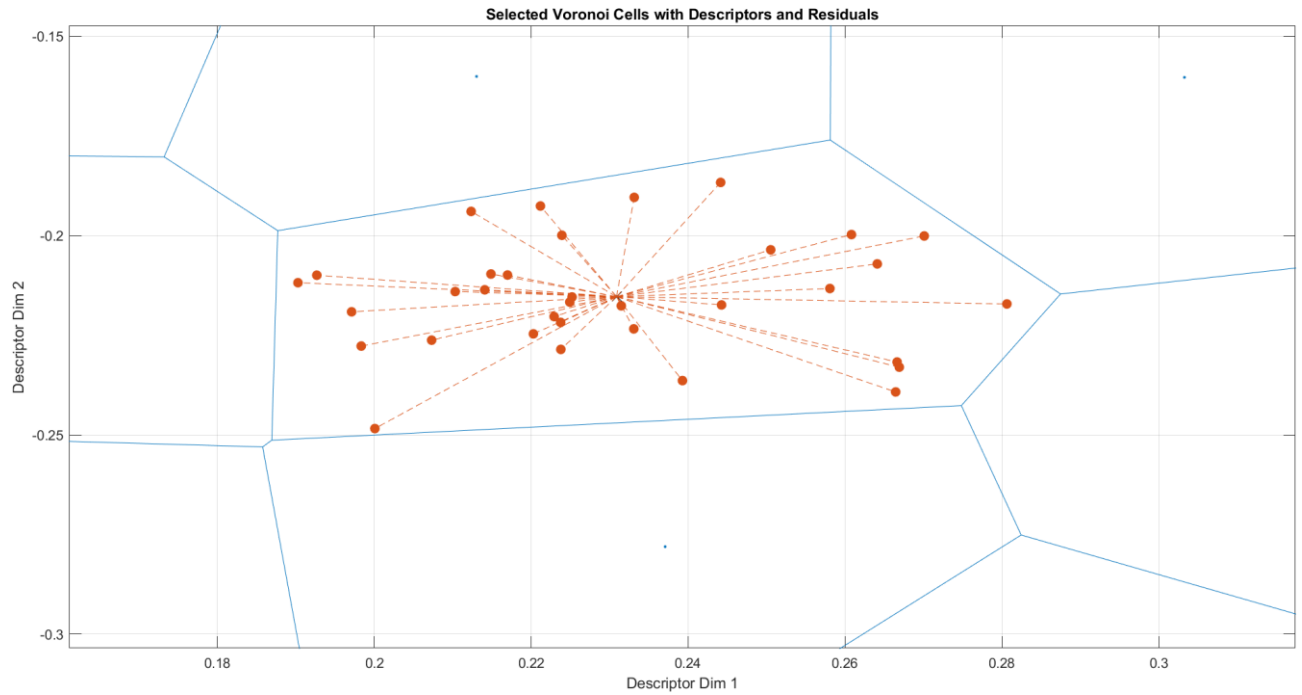
Εικόνα 51: Αποτύπωση των Voronoi κελιών μόνο με τα υπόλοιπα και τα κέντρα.



Εικόνα 52: Απομόνωση δυο μόνο Voronoi κελιών και αποτύπωση απομακρυσμένη οπτική των τοπικών περιγραφέων με τα κέντρα τους.



Εικόνα 53: Μεγεθυμένη οπτική του αριστερά Voronoi κελιού με τους τοπικούς περιγραφείς και το κέντρο του κελιού. Με διακεκομμένες γραμμές αποτυπώνονται τα υπόλοιπα ανάμεσα στους τοπικούς περιγραφείς και το εκάστοτε κέντρο.



Εικόνα 54: Μεγεθυμένη οπική του δεξιά Voronoi κελιού με τους τοπικούς περιγραφείς και το κέντρο του κελιού. Με διακεκομμένες γραμμές αποτυπώνονται τα υπόλοιπα ανάμεσα στους τοπικούς περιγραφείς και το εκάστοτε κέντρο.

Algorithm 1 Optimize SVM Hyperparameters and Calculate Accuracies

- 1: **Initialize** parallel pool and random seed
 - 2: **Get** images directory and form the imageDatastore
 - 3: **Count** the number of labels in the datastore
 - 4: **Set** number of experiments to 10
 - 5: **Initialize** accuracies matrix to store results
 - 6: **for** each experiment *expIdx* from 1 to numExperiments **do**
 - 7: **Split** the datastore into training and testing sets
 - 8: **Generate** SIFT descriptors for training and testing sets
 - 9: **Form** the dictionary using training features
 - 10: **Compute** VLAD descriptors for training and testing sets
 - 11: **Train** SVM model with different hyperparameter optimizations:
 - 12: **Optimize** BoxConstraint
 - 13: **Predict** labels for testing set
 - 14: **Compute** confusion matrix and accuracy
 - 15: **Store** accuracy in accuracies matrix
 - 16: **Optimize** BoxConstraint and KernelScale
 - 17: **Predict** labels for testing set
 - 18: **Compute** confusion matrix and accuracy
 - 19: **Store** accuracy in accuracies matrix
 - 20: **Optimize** all parameters
 - 21: **Predict** labels for testing set
 - 22: **Compute** confusion matrix and accuracy
 - 23: **Store** accuracy in accuracies matrix
 - 24: **Train** baseline SVM model without optimization
 - 25: **Predict** labels for testing set
 - 26: **Compute** confusion matrix and accuracy
 - 27: **Store** accuracy in accuracies matrix
 - 28: **end for**
 - 29: **Calculate** mean accuracies for each optimization
 - 30: **Determine** best accuracy from each experiment
 - 31: **Calculate** mean of the best accuracies
 - 32: **Display** mean accuracies and summary table
-

Αλγόριθμος 6: Ψευδοκώδικας για την εκτέλεση του πειράματος.

4.2 Αλγόριθμοι στην Riemannian Πολλαπλότητα

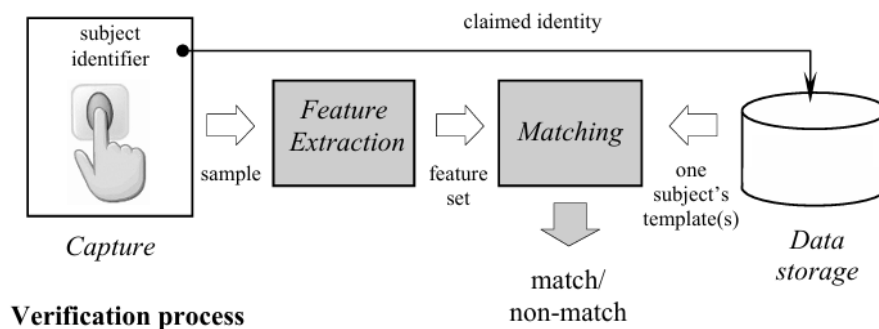
Έχοντας εκπληρώσει τον πρώτο στόχο αυτής της εργασίας, είναι ώρα να μελετηθεί το δεύτερο κομμάτι που είναι η χρήση της Riemannian Γεωμετρίας για την ενίσχυση της περιγραφικότητας των μοντέλων κατηγοριοποίησης εικόνων. Στην παρούσα υποενότητα θα αναλυθεί μια διαφορετική εφαρμογή, αυτή της αυτόματης επαλήθευσης υπογραφών (*Automated Signature Verification - ASV*), η οποία, στηρίζεται στην αξιοποίηση εικόνων χειρόγραφων υπογραφών για την επαλήθευση της ταυτότητας ενός προσώπου και την διεκπεραίωση της γνησιότητας του εγγράφου στο οποίο τοποθετείται η υπογραφή. Είναι κάπως παράξενο ή μπορεί να φαντάζει κάπως ειρωνικό πως από όλες τις υπερσύγχρονες μορφές επαλήθευσης μια υπογραφή, αν και τόσο απλή σαν μέθοδος, αποτελεί ένα από τα πιο σημαντικά μέσα ασφάλειας των προσωπικών δεδομένων ενός ανθρώπου. Αυτή η δυσπιστία άρεται αν αναλογιστεί κανείς πως όπως ακριβώς η λήψη μιας εικόνας δεν μπορεί να είναι ποτέ η ίδια, έτσι ακριβώς και η υπογραφή δυο γραφέων δεν μπορεί ποτέ να μοιράζεται τον ίδιο συνδυασμό χαρακτηριστικών υπογραφής, καθιστώντας την μια αμφιμονοσήμαντη αντιστοίχιση συνυφασμένη μόνο με το πρόσωπο του γνήσιου γραφέα.

Το ερευνητικό έργο της παρούσας υπο-ενότητας ξεκινά με τον εντοπισμό και την παράθεση κάποιων σημαντικών εννοιών που κεντρίζονται τόσο στο κομμάτι που πλαισιώνει τον τομέα της επαλήθευσης των υπογραφών, όσο και στις μετρικές που χρησιμοποιούνται για να αξιολογούνται τα μοντέλα που σχηματίζονται για την επιτέλεση αυτής. Τα *βιομετρικά* λοιπόν, ή αλλιώς *βιομετρική αναγνώριση*, αναφέρονται στην χρήση εξεχόντων και διακριτικών *ανατομικών* (π.χ το αποτύπωμα, το πρόσωπο ή η ίριδα) και *συμπεριφορικών* χαρακτηριστικών (π.χ ο λόγος), τα οποία ονομάζονται *βιομετρικά αναγνωριστικά* (*biometric identifiers*) ή *γνωρίσματα* ή *χαρακτηριστικά*. Τα τελευταία αποτελούν ένα σύνολο από μετρήσεις του σώματος, και όπως ακριβώς τα χαρακτηριστικά των εικόνων, ο σκοπός σχηματισμού τους είναι για την χρήση τους στην αυτόματη αναγνώριση φυσικών προσώπων. Αυτή η άμεση και αμφιμονοσήμαντη σύνδεση των σωματικών δεδομένων ενός ανθρώπου με μόνο ένα φυσικό πρόσωπο δημιουργεί έναν τόσο ισχυρό σύνδεσμο που καθιστά τα βιομετρικά ένα ιδιαίτερα ενδιαφέρον πρόβλημα αναγνώρισης μοτίβων (*pattern recognition*) με την παράλληλη επιστράτευση και χρήση ψηφιακών μέσων να ενισχύει τη διασφάλιση της ασφάλειας των προσωπικών δεδομένων των χρηστών μειώνοντας τις προσπάθειες απάτης, να προτρέπει σε βελτίωση της λογοδοσίας, να διευκολύνει το χρήστη και να αυξάνει την αποδοτικότητα [98].

4.2.1 Βιομετρικά Συστήματα

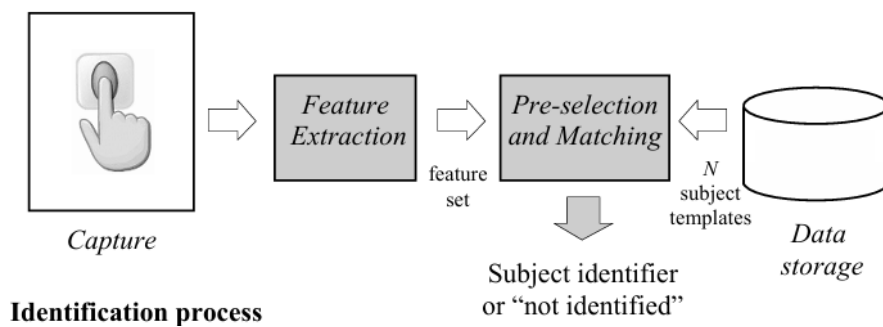
Η ανάλυση των βιομετρικών δεδομένων εφαρμόζεται από εξειδικευμένα βιομετρικά συστήματα τα οποία ανάγονται σε δυο βασικές διαδικασίες: της *επαλήθευσης* (*verification*) και της *αναγνώρισης* (*identification*). Είναι πολύ σημαντικό κανείς να μπορεί να διακρίνει ανάμεσα στις δυο αυτές έννοιες καθώς αντιμετωπίζουν το πρόβλημα με πολύ διαφορετικό τρόπο. Συγκεκριμένα:

- Ένα σύστημα *επαλήθευσης* (*verification system*) πιστοποιεί την ταυτότητα ενός ατόμου με σύγκριση του βιομετρικού χαρακτηριστικού που «συλλαμβάνεται» με ένα προηγουμένως συλλεχθέν (εγγεγραμμένο) βιομετρικό πρότυπο αναφοράς που είναι αποθηκευμένο στο σύστημα. Η διαδικασία αυτή περιλαμβάνει μια σύγκριση ένα προς ένα για να επιβεβαιωθεί εάν ο ισχυρισμός ταυτότητας του ατόμου είναι έγκυρος. Η απόφαση που προκύπτει από ένα τέτοιο σύστημα περιλαμβάνει είτε την αποδοχή είτε την απόρριψη του ισχυρισμού ταυτότητας [98].
- Ένα σύστημα *ταυτοποίησης* (*identification system*) αναγνωρίζει ένα άτομο αναζητώντας σε ολόκληρη τη βάση δεδομένων των εγγεγραμμένων προτύπων μια αντιστοιχία. Αυτή η διαδικασία περιλαμβάνει συγκρίσεις ένα προς πολλά για να προσδιοριστεί αν το άτομο υπάρχει στη βάση δεδομένων, και αν ναι, επιστρέφει το αναγνωριστικό της αντιστοιχισμένης αναφοράς εγγραφής. Σε ένα σύστημα ταυτοποίησης, το σύστημα διαπιστώνει την ταυτότητα του ατόμου (ή προσδιορίζει ότι το άτομο δεν είναι εγγεγραμμένο στη βάση δεδομένων του συστήματος) χωρίς να απαιτεί από το άτομο να διεκδικήσει την ταυτότητα του [98].



Εικόνα 55: Αποτύπωση του διαγράμματος ροής που ακολουθεί ένα σύστημα επαλήθευσης⁵².

⁵² Η εικόνα λήφθηκε από την πηγή [98].



Εικόνα 56: Αποτύπωση του διαγράμματος ροής που ακολουθεί ένα σύστημα ταυτοποίησης⁵³.

Πέρα από τον διαχωρισμό βάσει της εφαρμογής τους, τα βιομετρικά συστήματα κατηγοριοποιούνται και συναρτήσκει του τρόπου «σύλληψης», άρα και κατ' επέκτασιν της επεξεργασίας, της βιομετρικής πληροφορίας σε *on-line* και *off-line*. Η πρώτη κατηγορία απαιτεί η αναγνώριση των βιομετρικών δεικτών ενός προσώπου να εφαρμόζεται γρήγορα, με την εξαγωγή των δεδομένων να στηρίζεται στα δυναμικά χαρακτηριστικά του υποκειμένου (π.χ η τοποθεσία του στυλό, οι γωνίες κλίσης του, καθώς και η πίεση που εφαρμόζεται από το μέσο) και να πραγματοποιείται από μέσα όπως tablet ή έξυπνες συσκευές γενικότερα. Τόσο η διαδικασία όσο και το αποτέλεσμα αυτής καλούνται να είναι πλήρως αυτοματοποιημένες. Αντίστροφα η *off-line* μέθοδος δεν απαιτεί η ταυτοποίηση να πραγματοποιείται άμεσα και βασίζεται στα στατικά χαρακτηριστικά του ατόμου δίνοντας έδαφος σε πιο απλές, και ίσως πιο οικονομικές, μεθόδους λήψης των δεδομένων όπως είναι μια απλή φωτογραφία ή ένα έγγραφο που έχει σκαναριστεί. Τα *off-line* συστήματα τείνουν να είναι ήμι-αυτόματα καθώς μπορεί η απόκτηση των δεδομένων να γίνεται αυτόματα από κάποια ηλεκτρονική διάταξη όπως ένα σκάνερ, αλλά μπορούν να παρεμβάλλονται κι ενδιάμεσες επιβλεπόμενες καταστάσεις (π.χ κάποιος εξειδικευμένος εμπειρογνώμονας να μεριμνεί για την διασφάλιση της ποιότητας της ψηφιακής μορφής των εγγράφων) [98], [99].

4.2.2 Σφάλματα Μονάδων Αντιστοίχισης (*Matching Module Errors*)

Όλες οι διαδικασίες όμως επάγουν κάποιο είδος εισερχόμενου σφάλματος το οποίο πρέπει να αναγνωρίζεται ώστε να μπορεί να σχηματίζεται μια πιο καθαρή εικόνα της αποδοτικότητας ενός μοντέλου στον διαχωρισμό των υποκειμένων με βάση τους βιομετρικούς τους δείκτες. Η παρούσα εργασία εστιάζει και φωτίζει μόνο τα σφάλματα των μονάδων αντιστοίχισης (*matching module errors*) τα οποία επάγουν τους ρυθμούς σφαλμάτων επιβεβαίωσης (*Verification Error Rates*) γνωστοί και ως *Ρυθμός Λανθασμένης Αντιστοίχισης (False Match Rate- FMR)* και *Ρυθμός Μη-Αντιστοίχισης (False Non-Match Rate - FNMR)*. Ανάγοντας τα σφάλματα στον τομέα της αυτόματης επαλήθευσης υπογραφών, η ονομασία τους αλλάζει σύμφωνα με τα γενικά πρότυπα του τομέα σε *Ρυθμό*

⁵³ Η εικόνα λήφθηκε από την πηγή [98].

Λανθασμένης Αποδοχής (*False Acceptance Rate - FAR*), ο οποίος αντικατοπτρίζει τον ρυθμό με τον οποίο οι πλαστογραφημένες υπογραφές έχουν γίνει αποδεκτές από το σύστημα:

$$FAR = \frac{FP}{TN + FP} = \frac{\text{Αριθμός των Λανθασμένων Αποδοχών}}{\text{Συνολικός αριθμός προσπαθειών πλαστοπροσωπίας}} \quad (82)$$

Αντίθετα ο Ρυθμός Λανθασμένης Απόρριψης (*False Rejection Rate - FRR*), αναφέρεται στον ρυθμό με τον οποίο το σύστημα απορρίπτει γνήσιες υπογραφές:

$$FRR = \frac{FN}{FN + TP} = \frac{\text{Αριθμός των Λανθασμένων Απορρίψεων}}{\text{Συνολικός αριθμός Αυθεντικών προσπαθειών}} \quad (83)$$

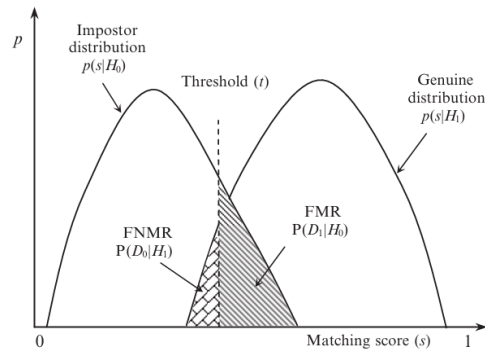
Όπου:

- *False Positive*: είναι η ποσότητα που αναφέρεται στις υπογραφές που λανθασμένα έχουν κατηγοριοποιηθεί ως γνήσιες.
- *True Negative*: είναι η ποσότητα που αναφέρεται στις υπογραφές που σωστά έχουν κατηγοριοποιηθεί ως πλαστογραφημένες.
- *False Negative*: είναι η ποσότητα που αναφέρεται στις υπογραφές που λανθασμένα έχουν κατηγοριοποιηθεί ως πλαστογραφημένες.
- *True Positive*: είναι η ποσότητα που αναφέρεται στις υπογραφές που σωστά έχουν κατηγοριοποιηθεί ως γνήσιες.

Γενικά, οι υψηλές τιμές FAR υποδεικνύουν ότι το σύστημα είναι επιρρεπές σε παραβιάσεις της ασφάλειας, καθώς συχνά κατατάσσει λανθασμένα τις πλαστές υπογραφές ως αυθεντικές. Ένα ισχυρό πλαίσιο αυτοματοποιημένης επαλήθευσης υπογραφών στοχεύει στην ελαχιστοποίησή του, ώστε να μειωθεί ο κίνδυνος μη εξουσιοδοτημένης πρόσβασης και δόλιων δραστηριοτήτων. Αντίστροφα οι υψηλές τιμές FRR υποδηλώνουν ότι το σύστημα είναι υπερβολικά αυστηρό, με αποτέλεσμα να δημιουργούνται προβλήματα στους νόμιμους χρήστες, οι οποίοι ενδέχεται να αντιμετωπίσουν περιττές αρνήσεις. Ως εκ τούτου, ένα αποτελεσματικό σύστημα αυτόματης επαλήθευσης υπογραφών προσπαθεί να εξισορροπήσει την ασφάλεια με την ευκολία των χρηστών, διατηρώντας την σε αποδεκτά χαμηλά επίπεδα.

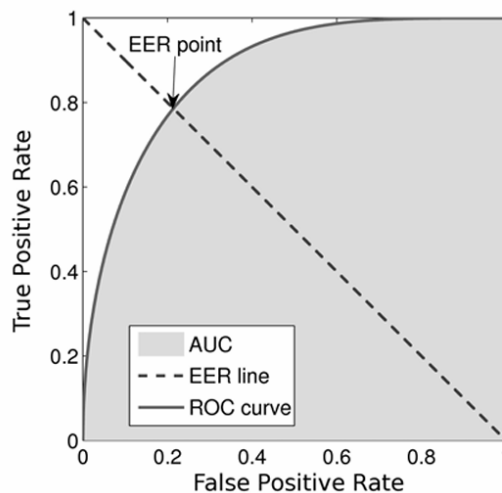
Γίνεται λοιπόν αντιληπτό πως το να μπορέσει να δημιουργηθεί ένα αποτελεσματικό σύστημα επαλήθευσης υπογραφών είναι μια πραγματική πρόκληση καθώς θα πρέπει να βρεθεί ένας τρόπος εξισορρόπησης ανάμεσα στο FAR και το FRR. Ένα στοιχείο που κατευθύνει προς την επίλυση αυτού του εμποδίου είναι η στενή αντιστρόφως ανάλογη σχέση που συνδέει τις δυο αυτές μετρικές, καθώς μείωση του FAR τυπικά ισοδυναμεί με αύξηση του FRR αλλά και το αντίστροφο οδηγώντας στην έγερση ενός συμβιβασμού ο οποίος απαιτεί προσεκτική βαθμονόμηση του κατωφλίου απόφασης του

συστήματος καθώς το τελευταίο καθορίζει το σημείο στο οποίο μια υπογραφή ταξινομείται ως γνήσια ή ως πλαστογραφημένη [98].



Εικόνα 57: Απεικόνιση της κατανομής των αυθεντικών και των πλαστογραφημένων κατανομών, του σημείου στο οποίο τοποθετείται το κατώφλι απόφασης καθώς και των σφαλμάτων FNMR και FMR⁵⁴.

Το Ισοδύναμο Ποσοστό Σφάλματος (Equal Error Rate - EER) χρησιμεύει ως μια ενοποιημένη μετρική που ενσωματώνει το παραπάνω συμβιβασμό, καθώς ορίζεται ότι είναι εκείνο το σημείο στο οποίο το FAR και το FRR είναι ίσα, παρέχοντας μια ενιαία τιμή που αντικατοπτρίζει τη συνολική ακρίβεια του συστήματος. Από μαθηματικής σκοπιάς, προσδιορίζεται από την απεικόνιση της καμπύλης ROC (Receiver Operation Curve- ROC), η οποία είναι ένα δισδιάστατο γράφημα που αποτυπώνει τη μεταβολή του FAR έναντι του FRR για διαφορετικές τιμές κατωφλίου. Το σημείο στο οποίο η καμπύλη τέμνει την γραμμή ισότητας, εκεί δηλαδή που το FAR ισούται με το FRR, λέγεται πως είναι το σημείο του EER.



Εικόνα 58: Αποτύπωση της ROC καμπύλης και του αντίστοιχου EER σημείου που προσδιορίζεται ως το σημείο ισότητας του FAR με το FRR⁵⁵.

⁵⁴ Η εικόνα λήφθηκε από την πηγή [98].

⁵⁵ Η εικόνα λήφθηκε από τον [σύνδεσμο](#).

Υπάρχουν και κάποια ακόμη σημαντικά σημεία του Χώρου της ROC καμπύλης τα οποία αξίζει να σημειωθούν, όπως το αριστερό χαμηλότερο σημείο (0,0) που αντικατοπτρίζει την στρατηγική κατά την οποία δεν αποδίδεται ποτέ θετική ταξινόμηση υπογραμμίζοντας έναν ταξινομητή ο οποίος δεν παρουσιάζει ποτέ False Positive σφάλματα όμως δεν λαμβάνει και True Positives αντίστοιχα ενώ το αντιδιαμετρικό του σημείο, (1,1) υλοποιεί την αντίστροφη στρατηγική. Γενικότερα ένα σημείο στον Χώρο της ROC καμπύλης είναι καλύτερο από ένα άλλο αν εντοπίζεται στο βορειοδυτικό κομμάτι αυτού (εκεί δηλαδή που ο TP ρυθμός είναι υψηλότερος και παράλληλα ο FP ρυθμός είναι χαμηλότερος) με το ιδανικό σημείο της τέλει ταξινόμησης να εμφανίζεται στο σημείο (0,1) [98], [100], [101].

4.2.3 Είδη πλαστογραφίας στις υπογραφές

Η πλαστογράφηση στην επαλήθευση υπογραφών μπορεί να κατηγοριοποιηθεί σε τρεις κύριους τύπους: τυχαίες πλαστογραφίες (*Random Forgeries*), απλές πλαστογραφίες (*Simple Forgeries*) και εξειδικευμένες πλαστογραφίες (*Skilled Forgeries*), όπου η κάθε κατηγορία παρουσιάζει μοναδικές προκλήσεις για τα συστήματα επαλήθευσης, απαιτώντας διαφορετικές προσεγγίσεις για την ακριβή ανίχνευσή τους. Αναλύοντας περαιτέρω τα προαναφερόμενα ορίζεται ότι [102]:

Ορισμός 30: Τυχαίες Πλαστογραφίες (*Random Forgeries*)

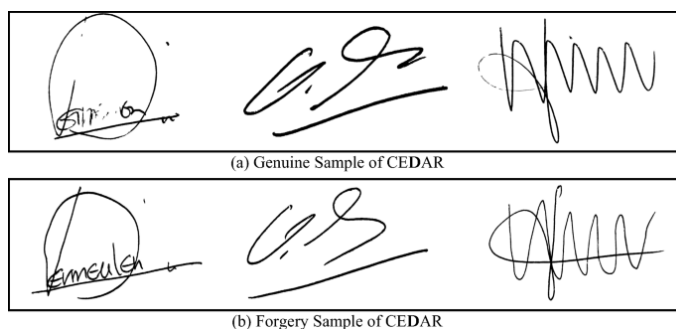
Οι τυχαίες πλαστογραφίες είναι υπογραφές που δημιουργήθηκαν χωρίς καμία γνώση της εμφάνισης της γνήσιας υπογραφής (ούτε καν του ονόματος του γραφέα) καθιστώντας τις τις πιο συχνά εύκολα ανιχνεύσιμες επειδή δεν προσπαθούν να μιμηθούν τα χαρακτηριστικά της γνήσιας υπογραφής. Ωστόσο, σε σενάρια με χαμηλή ευαισθησία στην ασφάλεια, ακόμη και οι τυχαίες πλαστογραφίες μπορούν περιστασιακά να γίνουν αποδεκτές, γεγονός που αναδεικνύει τη σημασία των ισχυρών μηχανισμών επαλήθευσης [102].

Ορισμός 31: Απλές Πλαστογραφίες (*Simple Forgeries*)

Οι απλές πλαστογραφίες είναι προσπάθειες αντιγραφής μιας υπογραφής με βάση μια πρόχειρη ματιά στη γνήσια. Ο πλαστογράφος έχει δει την αυθεντική υπογραφή, αλλά δεν κατανοεί σε βάθος τα ιδιαίτερα χαρακτηριστικά της, όπως η σειρά των γραμμμάτων, η πίεση και η ευχέρεια. Οι απλές πλαστογραφίες είναι πιο δύσκολο να εντοπιστούν από τις τυχαίες πλαστογραφίες, αλλά εξακολουθούν να στερούνται της ακρίβειας και της λεπτομέρειας που συναντάται στις εξειδικευμένες πλαστογραφίες [102].

Ορισμός 32: Εξειδικευμένες Πλαστογραφίες (Skilled Forgeries)

Οι εξειδικευμένες πλαστογραφίες δημιουργούνται από άτομα που έχουν εξασκηθεί στην αντιγραφή μιας γνήσιας υπογραφής πολλές φορές, συχνά με σημαντική γνώση των περίπλοκων λεπτομερειών της. Τέτοιες πλαστογραφίες μπορεί να είναι πολύ κοντά στο πρωτότυπο, γεγονός που τις καθιστά πιο δύσκολο να εντοπιστούν καθώς οι ειδικευμένοι πλαστογράφοι δίνουν προσοχή στις λεπτομέρειες της σειράς των κινήσεων, στις παραλλαγές της πίεσης και στη συνολική ευχέρεια, με αποτέλεσμα συχνά να δημιουργούν υψηλής ποιότητας απομιμήσεις που μπορούν να εξαπατήσουν τόσο τα ανθρώπινα όσο και τα αυτοματοποιημένα συστήματα επαλήθευσης [102].



Εικόνα 59: Παραδείγματα υπογραφών από την βάση δεδομένων CEDAR. Στο άνωθεν τμήμα εντοπίζεται ένα σύνολο από αυθεντικές υπογραφές ενώ στο κάτωθεν τμήμα εντοπίζεται ένα σύνολο από πλαστογραφημένες υπογραφές⁵⁶.

4.2.4 Προσεγγίσεις Ταξινόμησης Υπογραφών

Με την συνηθισμένη πορεία σε ένα αυτοματοποιημένο off-line σύστημα επαλήθευσης υπογραφών να μεταβαίνει από την λήψη των δεδομένων, στην προ-επεξεργασία και τέλος στην λήψη απόφασης, αυτό που δεν έχει διευκρινιστεί μέχρι πρότινος είναι ότι η προσέγγιση του προβλήματος μπορεί να τμηματοποιηθεί σε δυο επιμέρους διακριτές περιπτώσεις: την προσέγγιση που εξαρτάται από τον γραφέα (*Writer Dependent - WD*) ή την προσέγγιση που δεν εξαρτάται από τον γραφέα (*Writer Independent - WI*).

Η περίπτωση που εξαρτάται από τον γραφέα (*WD*), γνωστή και ως εξατομικευμένη επαλήθευση για κάθε χρήστη, περιλαμβάνει τη δημιουργία ενός μοναδικού μοντέλου επαλήθευσης για κάθε μεμονωμένο χρήστη. Αυτό το μοντέλο ταξινομητή εκπαιδεύεται αποκλειστικά στα δείγματα υπογραφών ενός συγκεκριμένου ατόμου, καθιστώντας το ιδιαίτερα προσαρμοσμένο στις συγκεκριμένες συνήθειες και τα χαρακτηριστικά γραφής του συγκεκριμένου χρήστη. Αφενός αυτή η πλήρως εξατομικευμένη προσέγγιση παρέχει την δυνατότητα υψηλής ακρίβειας στον διαχωρισμό επίπλαστων από αυθεντικών υπογραφών του κάθε γραφέα σε συνδυασμό με το γεγονός ότι η αποδοτικότητά της δεν επηρεάζεται από την, πιθανώς, μικρή ποσότητα δειγμάτων αναφοράς,

⁵⁶ Η εικόνα λήφθηκε από την πηγή [103].

αφετέρου όμως συνοδεύεται από κάποια επίσης αξιοπρόσεκτα μειονεκτήματα όπως ότι στην περίπτωση όπου υπάρχει μια μεγάλη συλλογή δειγμάτων τόσο από γραφείς όσο κι από τα δεδομένα τους, ο σχηματισμός επιμέρους μοντέλων για τον κάθε ένα ισοδυναμεί με ένα δίολου ευκαταφρόνητο υπολογιστικό αλλά και χρονικό κόστος [52], [53], [54], [104].

Στον αντίποδα βρίσκεται η προσέγγιση που δεν εξαρτάται από τον γραφέα (WI), γνωστή και ως παγκόσμια ή καθολική επαλήθευση, η οποία συνίσταται από τη δημιουργία ενός γενικού μοντέλου επαλήθευσης που εκπαιδεύεται σε ένα ευρύ σύνολο υπογραφών από πολλούς χρήστες. Αυτό το μοντέλο έχει σχεδιαστεί για να αναγνωρίζει κοινά χαρακτηριστικά και μοτίβα στις υπογραφές που είναι ενδεικτικά της γνησιότητας, ανεξάρτητα από τον εκάστοτε γραφέα, μοντελοποιώντας την πιθανοτική κατανομή ανάμεσα στην ενδοταξική κατηγορία (*intra-class*) των ζευγαριών γνήσιων-γνήσιων υπογραφών και την δια-ταξική σχέση (*inter-class*) μεταξύ των ζευγαριών γνήσιων-πλαστών υπογραφών σχηματίζοντας ένα χώρο απόστασης (μη) ομοιότητας [52], [53], [54], [104].

Ο αναφερόμενος νέος Χώρος που αποτυπώνει την διαφορά ανάμεσα στα δεδομένα που παρέχονται στον ταξινομητή ονομάζεται *Χώρος Διαφορών (Dissimilarity Space)* και ο τρόπος για να μπορέσει κανείς να ανάγει τα δεδομένα σε αυτόν είναι μέσω της τεχνικής του *Μετασχηματισμού της Διχοτόμησης (Dichotomy Transform)*. Σύμφωνα με το έργο του Victor L.F. Souza: «*A white-box analysis on the writer-independent dichotomy transformation applied to offline handwritten signature verification*», η προαναφερόμενη μεθοδολογία επιτρέπει τον μετασχηματισμό ενός προβλήματος αναγνώρισης μοτίβων με πολλές κλάσεις σε ένα δυαδικό πρόβλημα χρησιμοποιώντας ένα μέτρο ανομοιότητας το οποίο καθορίζει αν μια υπογραφή αναφοράς (*reference signature*) και μια αμφισβητούμενη υπογραφή (*questioned signature*) ανήκουν στον ίδιο γραφέα. Συγκεκριμένα, έστω ότι με x_q και x_r συμβολίζονται τα δυο διανύσματα χαρακτηριστικών στον Χώρο των Χαρακτηριστικών. Τότε το διάνυσμα ανομοιότητας που προκύπτει από τον μετασχηματισμό διχοτόμησης υπολογίζεται ως [54], [105]:

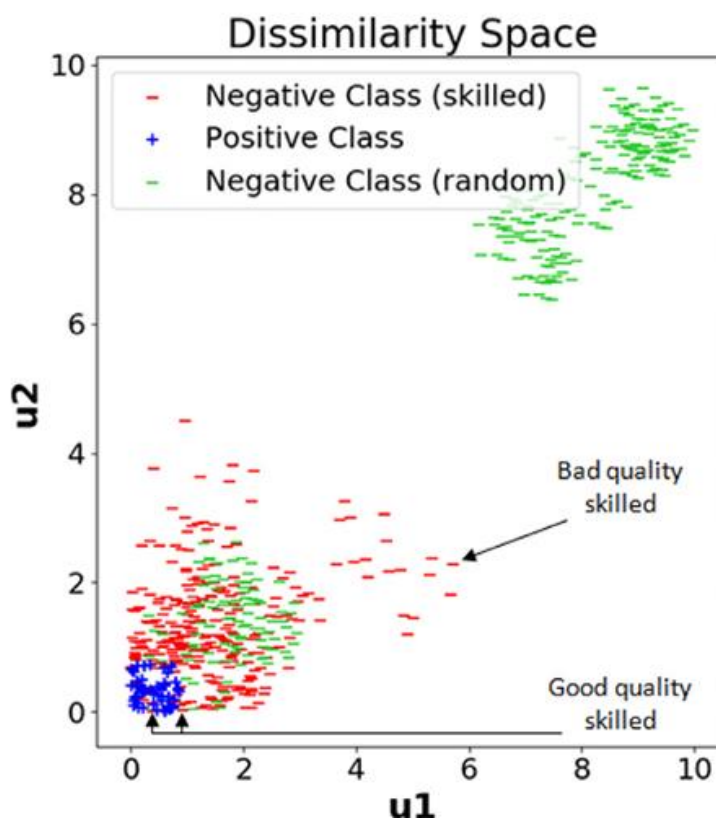
$$u(x_q, x_r) = \begin{bmatrix} |x_{q1} - x_{r1}| \\ |x_{q2} - x_{r2}| \\ \vdots \\ |x_{qn} - x_{rn}| \end{bmatrix} \quad (84)$$

Όπου:

- Με $|\cdot|$ συμβολίζεται ο τελεστής της απόλυτης τιμής της διαφοράς ανάμεσα στο x_q και στο x_r .
- Με x_{qi} και x_{ri} συμβολίζονται τα i -οστά στοιχεία των υπογραφών x_q και x_r αντίστοιχα.

Με αυτόν τον τρόπο, κάθε στοιχείο του διανύσματος u ισοδυναμεί με την απόσταση ανάμεσα στις αντίστοιχες διαστάσεις των διανυσμάτων x_q και x_r σχηματίζοντας έναν ενιαίο Χώρο ίδιας

διαστατικότητας για όλα τα στοιχεία. Ποια είναι όμως η πρακτική σημασία αυτής της μεθόδου; Πέρα από την προαναφερόμενη μείωση της διαστατικότητας του προβλήματος και δεδομένου ότι η παρούσα εργασία εστιάζει κατ' εξοχήν στην εφαρμογή συστημάτων *WI*, μια από τις πιο σημαντικές ιδιότητες που εμφανίζει ο αναφερόμενος μετασχηματισμός είναι ότι καταφέρνει να επάγει μια σημαντική αύξηση στον αριθμό των δειγμάτων του Χώρου ανομοιότητας καθώς συντελείται από κάθε ζεύγος συγκρίσεων υπογραφών. Με αυτόν τον τρόπο, αν ένας M αριθμός γραφών παρέχει ένα R σύνολο υπογραφών αναφοράς ο κάθε ένας, τότε ο τύπος (84) παράγει $\binom{MR}{2}$ διαφορετικά διανύσματα αποστάσεων εκ των οποίων τα $M\binom{R}{2}$ προέρχονται από την θετική κλάση και τα $\binom{M}{2}R^2$ ανήκουν στην αρνητική κλάση παρέχοντας την δυνατότητα παραγωγής πολλών δειγμάτων ακόμη κι αν τα ίδια τα δείγματα αναφοράς είναι λίγα σε ποσότητα [105].



Εικόνα 60: Απεικόνιση του Χώρου Ανομοιότητας που επάγεται από την χρήση του μετασχηματισμού διχοτόμησης⁵⁷.

Έχοντας αποτυπώσει τις βασικές έννοιες που πλαισιώνουν το πρόβλημα της αυτόματης επαλήθευσης χειρόγραφων υπογραφών, τα επαγόμενα σφάλματα που πρέπει να λαμβάνονται υπόψιν, τις προσεγγίσεις του προβλήματος και υπογραμμίζοντας την επιλογή που αναλαμβάνει να αντιμετωπίσει η προκείμενη διπλωματική εργασία, είναι πλέον στιγμή να αποτυπωθούν δυο νέα σενάρια που αξιοποιούν την μη-γραμμική γεωμετρία των πινάκων συνδιακύμανσης.

⁵⁷ Η εικόνα λήφθηκε από την πηγή [105].

4.2.5 Αξιοποιώντας τον Εφαπτομενικό Χώρο

Αναφέρθηκε ότι το κύριο κομμάτι ενασχόλησης της παρούσας διπλωματικής, τουλάχιστον για το κομμάτι της αυτόματης επαλήθευσης χειρόγραφων υπογραφών, εστιάζει στην προσέγγιση του *off-line WI* συστήματος. Αυτό σημαίνει ότι τα ανεπεξέργαστα δεδομένα αποτελούν σκαναρισμένες εικόνες χειρόγραφων υπογραφών οι οποίες πρέπει να υποβληθούν σε μια σειρά από διαδικασίες μέσα από τις οποίες θα εξαχθούν οι περιγραφείς περιοχών (*regional descriptors*). Υιοθετώντας την προσέγγιση που εμφανίζεται στο έργο του *Elias N. Zois*: «*Subscripto multiplex: A Riemannian symmetric positive definite strategy for offline signature verification*» [52] αποτυπώνεται το παρακάτω σχηματικό:

Έστω ότι με $I \in \mathbb{R}^{w \times h}$ συμβολίζεται μια grayscale εικόνα με w στήλες και h σειρές καθώς κι ότι με $F \in \mathbb{R}^{w \times h \times n}$ συμβολίζεται μια στοίβα από αντιστοιχίσεις χαρακτηριστικών με n ανεπεξέργαστα επίπεδα εικόνας που αξιολογούνται με την χρήση ενός αριθμού φίλτρων $F(x, y) = \Phi(I, x, y)$. Τότε δοσμένης μιας τετραγωνικής περιοχής $\mathcal{R} \subset F$, όπου με $f = \{f_i\}_{i=1,2,\dots,S} \in \mathbb{R}^{n \times S}$ ορίζεται πως είναι μια τοπική αντιστοιχία χαρακτηριστικών S εικονοστοιχείων που εντοπίζονται σε αυτή την τετραγωνική περιοχή, η τελευταία μοντελοποιείται από την χρήση του πίνακα συνδιακύμανσης $C_{\mathcal{R}} \in \mathbb{R}^{n \times n}$ των f_i σημείων ως [52]:

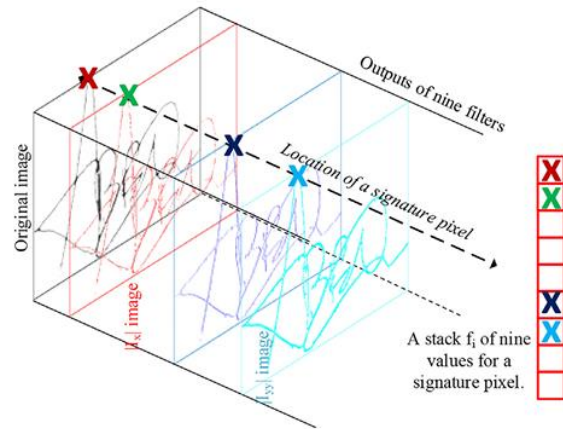
$$C_{\mathcal{R}} = \frac{1}{S-1} \sum_{i=1}^S (f_i - \mu)(f_i - \mu)^T \quad (85)$$

Με την αντιστοίχιση $\Phi(I, x, y)$ που εφαρμόζεται στην εικόνα να δίνεται από την συλλογή εννέα (9) διαφορετικών φίλτρων [52]:

$$\left[I_p, |I_{p,x}|, |I_{p,y}|, |I_{p,xx}|, |I_{p,yy}|, \sqrt{I_{p,x}^2 + I_{p,y}^2}, \tan^{-1}(I_{p,y}/I_{p,x}), x_n, y_n \right]$$

Όπου:

- $I_p \equiv I_p(x, y)$: είναι η grayscale εικόνα της υπογραφής μετά την προ-επεξεργασία της [52].
- $|I_{p,x}|, |I_{p,y}|, |I_{p,xx}|, |I_{p,yy}|$: είναι η πρώτη και δεύτερη παράγωγος της $I_p(x, y)$ και στις δυο κατευθύνσεις [52].
- x_n, y_n : είναι οι συντεταγμένες των εικονοστοιχείων της υπογραφής, κανονικοποιημένες με τον μέγιστο αριθμό των σειρών και των γραμμών του πλαισίου οριοθέτησης της εικόνας [52].
- $\tan^{-1}(I_{p,y}/I_{p,x})$: είναι η κατεύθυνση της κλίσης, κανονικοποιημένη σε radians με τις τιμές να μεταβάλλονται στο εύρος $[-\pi, \pi]$ [52].

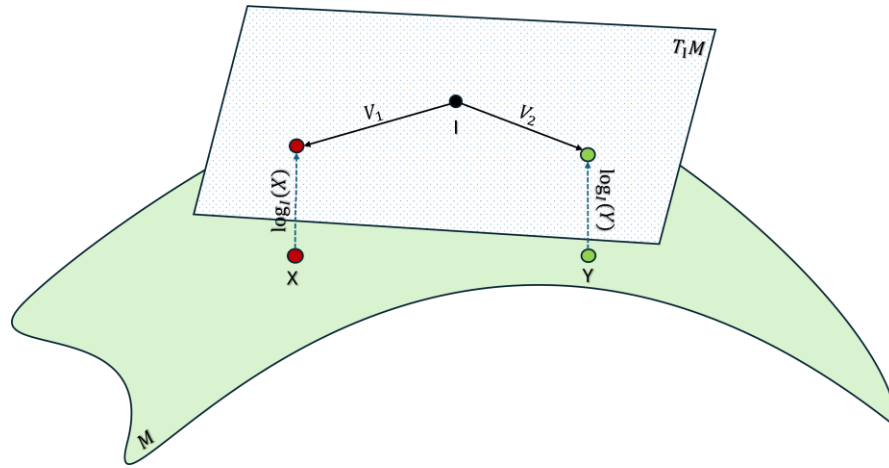


Εικόνα 61: Αποτύπωση του σχηματισμού του περιγραφέα περιοχών ως πίνακα συνδιακύμανσης. Για ένα συγκεκριμένο εικονοστοιχείο που υποδεικνύεται με «x», δημιουργείται μια στοιβάδα από τιμές που προκύπτουν μέσω των 9 φίλτρων που εφαρμόζονται στην αρχική εικόνα⁵⁸.

Με τον περιγραφέα περιοχών να αποτελεί έμπρακτα έναν ισχυρό περιγραφέα αντιστοιχίσεων (*matching descriptor*) απηλλαγμένο από την τοποθεσία των εικονοστοιχείων, επάγοντας έτσι περιστροφική και κλιμακωτή ανεξαρτησία (*scale and rotation invariance*) γύρω από την περιοχή ενδιαφέροντος, ο σχηματισμός του πίνακα συνδιακύμανσης σηματοδοτεί την έμπρακτη αναγωγή των εικόνων σε σημεία του SPD Πολύπτυχου [52].

Με τις αρχικές λοιπόν εικόνες εκπεφραζόμενες μέσω του πίνακα συνδιακύμανσής τους ως σημεία του Πολύπτυχου των Συμμετρικών Θετικά Ορισμένων Πινάκων, το πρόβλημα που εγείρεται είναι το πως θα μπορέσει να εφαρμοστεί ο Μετασχηματισμός της Διχοτόμησης ανάμεσα σε αυτά, καθώς υπενθυμίζεται ότι ο τυπικός τελεστής της αφαίρεσης δεν υφίσταται στο μη-γραμμικό Πολύπτυχο που υπάγονται οι αναπαραστάσεις των εικόνων. Χρησιμοποιώντας τα μαθηματικά εργαλεία που έχουν παρατεθεί στο Κεφάλαιο 3 γίνεται χρήση του Εφαπτομενικού Χώρου, ο οποίος υπενθυμίζεται ότι εισάγει την έννοια της τοπικής Ευκλείδειας Γεωμετρίας μέσω της διανυσματικής αναπαράστασης από την απεικόνιση του σημείου στο Πολύπτυχο μέσω της λογαριθμικής συνάρτησης πίνακα. Αυτό που διευκρινίζεται είναι ότι η διαδικασία αυτή καταφέρνει να ανάγει όλα τα σημεία του Πολύπτυχου σε έναν κοινό πόλο ο οποίος συμβατικά ορίζεται πως είναι αυτός που επάγεται από τον μοναδιαίο πίνακα I .

⁵⁸ Η εικόνα λήφθηκε από την πηγή [52].



Εικόνα 62: Αποτύπωση της αναγωγής των σημείων της Πολλαπλότητας στον κοινό Εφαπτομενικό Χώρο που επάγεται από τον μοναδιαίο πίνακα.
 Με μαύρο, σημειώνεται το σημείο που επάγεται από τον μοναδιαίο πίνακα σχηματίζοντας τον κοινό πόλο.
 Με κόκκινο χρώμα στα αριστερά του κοινού πόλου εντοπίζεται το σημείο της Πολλαπλότητας X και το διάνυσμα που δημιουργείται μέσω της λογαριθμικής αντιστοίχισής του.
 Με πράσινο χρώμα στα δεξιά του κοινού πόλου εντοπίζεται το σημείο της Πολλαπλότητας Y και το διάνυσμα που δημιουργείται μέσω της λογαριθμικής αντιστοίχισής του.

Έχοντας λοιπόν τα παραπάνω ως οδηγούς η προσέγγιση που ακολουθείται, εφόσον έχουν εξαχθεί οι πίνακες συνδιακύμανσης τόσο για τις γνήσιες όσο και για τις πλαστογραφημένες υπογραφές, είναι η «ανύψωση» όλων των σημείων στο Εφαπτομενικό επίπεδο του μοναδιαίου πίνακα I με κοινό πόλο το σημείο που επάγεται από αυτόν. Τι πραγματικά επιτυγχάνεται με αυτό όμως; Πέρα από την προφανή γραμμικοποίηση λόγω της αναγωγής σε έναν διανυσματικό Χώρο, όπως αναφέρθηκε και στο κομμάτι του μετασχηματισμού διχοτόμησης, αναμένεται στοιχεία του πίνακα που τον συντελούν, τα οποία αναφέρονται σε κοντινά σημεία του Πολύπτυχου, να συσπειρώνονται και σε αντίστοιχα κοντινές αποστάσεις το οποίο μεταφράζεται ως: «Συσχετίσεις ανάμεσα σε Γνήσιες-προς-Γνήσιες υπογραφές θα εμφανίζονται πιο κοντά από συσχετίσεις ανάμεσα σε Γνήσιες-προς-Πλαστογραφημένες» σχηματίζοντας αποτελεσματικά τον Χώρο Ανομοιότητας, όπως αυτός περιεγράφηκε και στο έργο του Souza [105].

Με την βάση των δεδομένων των υπογραφών να αποτελεί την CEDAR, μια από τις πιο διαδεδομένες βάσεις στον τομέα της αυτόματης επαλήθευσης υπογραφών, εντοπίζονται 55 διακεκριμένοι γραφείς, όπου για τον καθένα υπάρχουν 24 Γνήσιες και 24 Πλαστογραφημένες υπογραφές. Υλοποιώντας ένα *5-by-2 fold validation* ο διαχωρισμός των δεδομένων γίνεται με 27 γραφείς στο κομμάτι της Μάθησης (Learning) και με τους εναπομείναντες 28 γραφείς στο κομμάτι του Ελέγχου (Testing). Στην συνέχεια για τον κάθε ένα από τους 27 γραφείς, οι 17 από τις 24 υπογραφές δεσμεύονται για την εκπαίδευση (*Training*) και οι υπόλοιπες 7 χρησιμοποιούνται στο σύνολο της επαλήθευσης. Η ίδια ακριβώς διαδικασία με τα ίδια ακριβώς νούμερα εφαρμόζεται και για τις πλαστές υπογραφές. Έχοντας πλέον δεσμεύσει τα απαραίτητα δεδομένα, υλοποιούνται τα σύνολα των Γνήσιων και των Εξειδικευμένων Πλαστογραφιών ως εξής:

- Για το σύνολο των Γνήσιων υπογραφών: Εφαρμόζεται ο Μετασχηματισμός Διχοτόμησης για τον κάθε έναν πιθανό συνδυασμό των 17 γνήσιων υπογραφών. Αυτό σημαίνει ότι θα δημιουργηθούν 136 πιθανά ζευγάρια, τα οποία θα χρησιμοποιηθούν ως δείκτες για να δεικτοδοτήσουν μέσα στα δεδομένα που έχουν δεσμευτεί και να εφαρμόσουν τον Μετασχηματισμό της Διχοτόμησης. Η διαδικασία επαναλαμβάνεται και για το σύνολο της επαλήθευσης (*Validation*) αποδίδοντας έναν πίνακα $136 \times 27 = 3672$ σειρών των 55 διαστάσεων για τα δεδομένα της εκπαίδευσης και $21 \times 27 = 567$ σειρών των 55 διαστάσεων για τα δεδομένα της επαλήθευσης :

Εικόνα 63: Αποτύπωση της δεικτοδότησης των δεδομένων για την εφαρμογή του μετασχηματισμού διχοτόμησης στο σύνολο της εκπαίδευσης για τις γνήσιες υπογραφές.

- Για το σύνολο των Εξειδικευμένων Πλαστογραφήσεων: Για κάθε έναν από τους 27 γραφείς εφαρμόζεται ο μετασχηματισμός διχοτόμησης ανάμεσα στις 17 Γνήσιες και στις 17 Πλαστογραφημένες υπογραφές του. Αυτό έχει ως αποτέλεσμα τον σχηματισμό $17 \times 17 = 289$ διανυσμάτων των 55 διαστάσεων, εκ των οποίων, προκειμένου να μπορέσει να υπάρξει ισορροπία δεδομένων ανάμεσα στο σύνολο των Γνήσιων και των Εξειδικευμένων Πλαστογραφήσεων, δεικτοδοτούνται τυχαία 136 εξ' αυτών σχηματίζοντας ξανά $137 \times 27 = 3672$ διανύσματα των 55 διαστάσεων για το σύνολο της Εκπαίδευσης και αντίστοιχα $21 \times 27 = 567$ διανύσματα των 55 διαστάσεων για το σύνολο της Αξιολόγησης.

Συγκεντρώνοντας τα δεδομένα σε ενιαίους πίνακες, οι τελευταίοι τροφοδοτούνται σε SVM με παράλληλη βελτιστοποίηση των υπερπαραμέτρων τους για την εύρεση των βέλτιστων αποτελεσμάτων. Κλείνοντας προκειμένου το σχηματικό του *5-by-2 fold validation* να είναι έγκυρο, με το πέρας του σχηματισμού του ταξινομητή τα σύνολα αντιστρέφονται παρέχοντας πλέον 28 γραφείς για την διαδικασία της Μάθησης και 27 για την διαδικασία της Επαλήθευσης.

Εν κατακλείδι στο κομμάτι του Ελέγχου, εφαρμόζοντας ξανά το *5-by-2 fold* σχηματικό που χρησιμοποιήθηκε προηγουμένως για κάθε γραφέα, που ανήκει σε αυτό το σύνολο, και για 10 διαφορετικές επαναλήψεις επιλέγονται 10 τυχαίες Γνήσιες υπογραφές ως δεδομένα αναφοράς (*references*) με τις εναπομείναντες 14 να συνδυάζονται με τις 24 Εξειδικευμένες Πλαστογραφήσεις δημιουργώντας 38 γραφείς που λειτουργούν ως τα αμφισβητούμενα δεδομένα (*Questioned Data*). Για κάθε αμφισβητούμενο γραφέα, εφαρμόζεται ο μετασχηματισμός διχοτόμησης ανάμεσα στα διανύσματα του και τα 10 δεδομένα αναφοράς, σχηματίζονται προβλέψεις με βάση το καλύτερο μοντέλο και αποδίδεται το EER του, που προκύπτει από την ROC καμπύλη ανάμεσα στις ετικέτες των αμφισβητούμενων δεδομένων και τα μέγιστα score όπως αυτά έχουν προκύψει από τον ταξινομητή προηγουμένως. Η συνολική διαδικασία παράγει έναν πίνακα EER με διαστάσεις που εξαρτώνται από τον αριθμό των επαναλήψεων καθώς και τον αριθμό των *folds*.

Οι ψευδοκώδικες που υλοποιούν το παραπάνω σχηματικό αποδίδονται στην συνέχεια:

Algorithm 1 Signature Verification using SVM with Hyperparameter Optimization

```
1: Initialization:
2: Clear workspace, set random seed
3: Load covariance matrices from file
4: Separate genuine and forgery covariance matrices
5: Convert covariance matrices to vectors on the tangent plane using
   VecCell.m
6: Learning Stage:
7: for Iter = 1 to numIters do
8:   Randomly sample indices for training data
9:   Split data into learning and testing sets
10:  for Fold = 1 to numFolds do
11:    Partition genuine and forgery data into training and validation sets
      using Train_Val_Split.m
12:    Form  $\omega(+)$  class using OmegaPlusFormation.m
13:    Form  $\omega(-)$  class using OmegaMinusFormation.m
14:    Create labels for training and validation sets
15:    Combine training and validation data into matrices
16:    Perform hyperparameter optimization on SVM using fitcsvm
17:    Store best hyperparameters
18:    Retrain SVM with best hyperparameters
19:    Validate SVM and compute performance metrics (AUC, EER)
20:    Store validation results
21:    Swap learning and testing data for next fold
22:  end for
23: end for
24: Testing Stage:
25: for Iter = 1 to numIters do
26:   Create testing data using saved indices
27:   for Fold = 1 to numFolds do
28:     Retrieve SVM classifier for current fold
29:     Create reference and question data using RefsQuestionNV.m
30:     Store testing results and compute mean EER
31:   end for
32: end for
```

Αλγόριθμος 7: Βασικός αλγόριθμος για το πρώτο Riemannian σενάριο.

Algorithm 1 Train_Val_Split.m

```

1: Input: Data, Options
2: Output: Metadata
3: Initialize:
4: Set default values for Options if not provided
5: Calculate Training_Length as round(size(Data,2) * Options.Train.Portion)
6: Calculate Validation_Length as round(size(Data,2) * (1 - Options.Train.Portion))
7: Form Training Set:
8: for each writer in Data do
9:   Extract TrainingData for the writer
10:  Generate all pairwise combinations of TrainingData
11:  Store pairs of observations in Training_Signature_pairs
12: end for
13: Form Validation Set:
14: for each writer in Data do
15:   Extract ValidationData for the writer
16:   Generate all pairwise combinations of ValidationData
17:   Store pairs of observations in Validation_Signature_pairs
18: end for
19: Return Metadata based on Options:
20: if Options.GetTheCombinations is true then
21:   Set Metadata with Training_Signature_pairs, Validation_Signature_pairs,
   Train_Combinations, Validation_Combinations
22: else if Options.GetTheRefsPerWriter is true then
23:   Set Metadata with Training_Signature_pairs, Validation_Signature_pairs,
   Training_Length, Validation_Length
24: else if Options.GetAllMetadata is true then
25:   Set Metadata with all relevant fields
26: end if

```

Αλγόριθμος 8: Αλγόριθμος για την συνάρτηση Train_Val_Split.m.

Algorithm 1 OmegaPlusFormation.m

```

1: Input: G_Learning_Data, G_Metadata
2: Output: G_Training_set, G_Validation_set
3: Initialize G_Training_set and G_Validation_set as empty
4: Form Training Set:
5: for each writer in G_Learning_Data do
6:   for each combination in G_Metadata.Train_Combinations do
7:     Compute dichotomy transform for genuine pairs
8:     Store result in G_Train_Dichotomy
9:   end for
10: end for
11: Merge all dichotomy transforms into G_Training_set
12: Form Validation Set:
13: for each writer in G_Learning_Data do
14:   for each combination in G_Metadata.Validation_Combinations do
15:     Compute dichotomy transform for genuine pairs
16:     Store result in G_Val_Dichotomy
17:   end for
18: end for
19: Merge all dichotomy transforms into G_Validation_set

```

Αλγόριθμος 9: Αλγόριθμος για την συνάρτηση OmegaPlusFormation.m.

Algorithm 1 OmegaMinusFormation.m

```

1: Input: G_Learning_Data, F_Learning_Data, F_Metadata
2: Output: F_Training_set, F_Val_set
3: Initialize F_Training_set and F_Val_set as empty
4: Training Set Formation:
5: for each writer in G_Learning_Data do
6:   Extract genuine and forgery signatures for training
7:   for each genuine signature do
8:     for each forgery signature do
9:       Compute dichotomy transform and store in structure
10:     end for
11:   end for
12:   Collect all dichotomy transforms into a matrix
13:   Randomly select 136 samples from the matrix
14:   Append selected samples to F_Training_set
15: end for
16: Validation Set Formation:
17: for each writer in G_Learning_Data do
18:   Extract genuine and forgery signatures for validation
19:   for each genuine signature do
20:     for each forgery signature do
21:       Compute dichotomy transform and store in structure
22:     end for
23:   end for
24:   Collect all dichotomy transforms into a matrix
25:   Randomly select 21 samples from the matrix
26:   Append selected samples to F_Val_set
27: end for

```

Αλγόριθμος 10: Αλγόριθμος για την συνάρτηση OmegaMinusFormation.m.

Algorithm 1 RefsQuestionNV.m

```

1: Input: G_Testing_Data, F_Testing_Data, SVM, Options
2: Output: Data, Metrics
3: for each Writer in G_Testing_Data do
4:   for each Testing_Iter from 1 to Options.numOfTestingItrs do
5:     Print current iteration and writer
6:     Generate random indices for Reference and Forgery data
7:     Create remaining indices for Question set
8:     Form Reference and Question data for the current writer
9:     Create Ground Truth labels for Question data
10:    for each Question data point do
11:      for each Reference data point do
12:        Compute Dichotomy Transform
13:      end for
14:      Concatenate Dichotomy Transforms into a matrix
15:      Predict labels and scores using SVM
16:      Store scores and maximum score
17:      Select label with maximum score
18:    end for
19:    Compute ROC curve and AUC
20:    Compute FAR and FRR
21:    Find EER where FAR and FRR are closest
22:    Store EER and AUC for the current iteration
23:    Plot and save ROC curve
24:  end for
25:  Compute mean EER per writer
26: end for

```

Αλγόριθμος 11: Αλγόριθμος για την συνάρτηση RefsQuestionNV.m.

Ο συνολικός πίνακας απόδοσης των EER εκπεφρασμένα σε ποσοστό επί τοις εκατό ανά επανάληψη και ανά *fold* παρατίθεται στην συνέχεια:

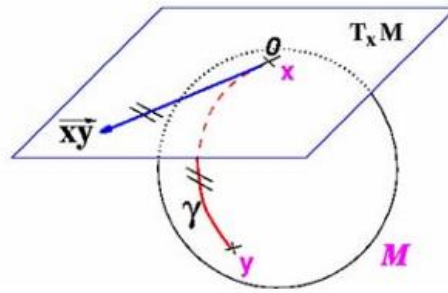
Πίνακας 7: Αποτίμηση του EER ανά επανάληψη κι ανά *fold* καθώς και του συνολικού μέσου EER για το πρώτο Riemannian σενάριο.

Επανάληψη	Fold 1	Fold 2
1	3.735	5.2778
2	79.7024	4.028
3	7.9315	3.6420
4	4.0625	3.2099
5	3.8095	3.3333
Μέσο EER	11.87	

4.2.6 Αξιοποιώντας τη Riemannian Πολλαπλότητα και τον Εφαπτομενικό Χώρο

Σε ένα πολύ διαφορετικό πλέον πλαίσιο και με οδηγούς τα έργα των Pennec: «A Riemannian Framework for Tensor Computing» και Tuzel: «Pedestrian Detection via Classification on Riemannian Manifolds» υιοθετείται μια διαφορετική οπτική στην συσχέτιση των σημείων της Πολλαπλότητας των SPD πινάκων και των επαγόμενων εφαπτομενικών Χώρων που αυτά εισάγουν. Είναι πλέον γνωστό πως ένα διάνυσμα \vec{xy} , το οποίο έχει ως αρχή του το σημείο x , μπορεί επίσης να ιδωθεί κι ως ένα διάνυσμα στον Εφαπτομενικό Χώρο σε εκείνο το σημείο. Αυτό το προκείμενο διάνυσμα μπορεί να γεφυρώσει μια συσχέτιση με ένα άλλο σημείο της Πολλαπλότητας χρησιμοποιώντας την γεωδαισιακή που ξεκινά από το x και έχει το αντίστοιχο διάνυσμα \vec{xy} στον Εφαπτομενικό της Χώρο. Άρα πρακτικά με την χρήση της λογαριθμικής απεικόνισης, δημιουργείται μια άμεση αναλογία της αφαίρεσης δυο σημείων πάνω στην Πολλαπλότητα αφού η βάση του

διανύσματος είναι το x και το πέρας του βέλους του διανύσματος αντιστοιχεί, μέσω της εκθετικής απεικόνισης, σε ένα νέο σημείο στο Πολύπτυχο [81]:



Εικόνα 64: Αποτύπωση του διανύσματος \overline{xy} , της γεωδαισιακής που το ακολουθεί και της διαφοράς των σημείων y και x ⁵⁹.

Προκειμένου αυτή η αφαίρεση να μπορεί να διατηρεί την γεωμετρία του Πολύπτυχου ακέραια ενώ παράλληλα να μπορεί να αποδώσει και μια πιο απλοποιημένη και υπολογιστικά ανεκτή μορφή, εφαρμόζεται η διαδικασία της επιβολής των ορθοκανονικών συντεταγμένων η οποία ορίζεται ως εξής [106]:

Ορισμός 33: Ορθοκανονικές συντεταγμένες

Οι ορθοκανονικές συντεταγμένες ενός εφαπτομενικού διανύσματος y στον εφαπτομενικό χώρο στο σημείο X δίνεται από τον διανυσματικό τελεστή:

$$vec_X(y) = vec_I(X^{1/2}yX^{1/2}) \quad (86)$$

Όπου με I ορίζεται ο μοναδιαίος πίνακας, ενώ ο διανυσματικός τελεστής στον ταυτοτικό πίνακα ορίζεται πως είναι:

$$vec_I(y) = [y_{1,1}, \sqrt{(2)}y_{1,2}, \sqrt{(2)}y_{1,3}, \dots, y_{d,d}]^T \quad (87)$$

Καθώς δεν πρέπει να λησμονείται ότι λόγω της συμμετρικότητας των πινάκων μόνο τα $d(d+1)/2$ στοιχεία του πίνακα φέρουν την σημαντική πληροφορία των ανεξάρτητων συντελεστών (είτε κανείς επιλέγει τον πάνω είτε τον κάτω τριγωνικό πίνακα).

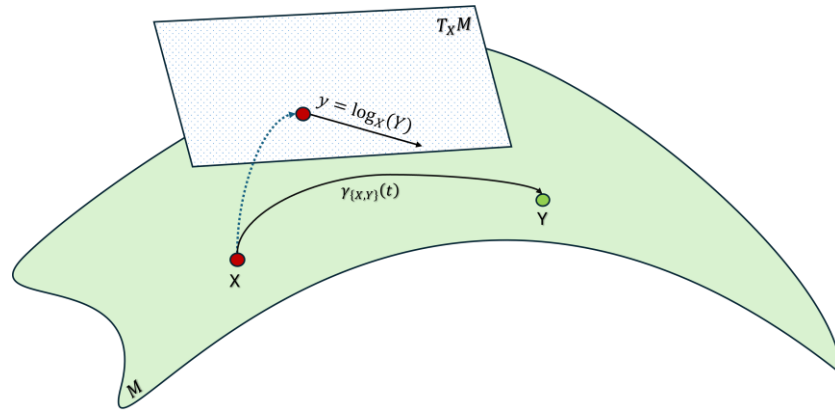
Υιοθετώντας τις παραπάνω θεωρητικές προσεγγίσεις, τα πειράματα μετασχηματίζονται χρησιμοποιώντας την λογαριθμική απεικόνιση ως το μέσο της αποτύπωσης της διαφοράς ανάμεσα στα σημεία της Πολλαπλότητας αντί να χρησιμοποιείται για να τα μεταφέρει στο κοινό εφαπτομενικό επίπεδο όπως εφαρμόστηκε στο προηγούμενο πείραμα. Ως ένδειξη καλής πρακτικής οφείλεται να διευκρινιστεί τι ακριβώς επιτυγχάνεται μέσω της χρήσης του τύπου (85) και πως διαφοροποιείται με την απλή χρήση της λογαριθμικής απεικόνισης.

⁵⁹ Η εικόνα λήφθηκε από την πηγή [81].

Στην προκείμενη περίπτωση υιοθετείται μια πιο εξειδικευμένη προσέγγιση που χρίζει επιμέρους εξήγησης του θεωρητικού υποβάθρου που την πλαισιώνει. Έστω ότι με X και Y υποδηλώνονται, όπως πάντα, δυο σημεία πάνω στην SPD Πολλαπλότητα όπου το πρώτο αναφέρεται στην προκείμενη υπογραφή και το δεύτερο στην αμφισβητούμενη υπογραφή είτε αυτή είναι Γνήσια είτε είναι Πλαστογραφημένη. Προφανώς τα δυο αυτά σημεία, με διαφορετική χωρική τοποθεσία πάνω στο Πολύπτυχο, δεν μπορούν να γεφυρωθούν με κάποιου είδους συσχέτισης καθώς ο μη-γραμμικός Χώρος στον οποίο «ζουν» δεν επιτρέπει άμεσους τρόπους σύγκρισης. Πως θα μπορούσε λοιπόν να αρθεί αυτό το πρόβλημα; Αναλύοντας τον τύπο (61) της λογαριθμικής απεικόνισης:

$$y = \log_X(Y) = X^{\frac{1}{2}} \log \left(X^{-\frac{1}{2}} Y X^{-\frac{1}{2}} \right) X^{\frac{1}{2}}$$

Εντοπίζονται οι ποσότητες του $X^{\frac{1}{2}}$, που αντικατοπτρίζει την τετραγωνική ρίζα του πίνακα X , καθώς και η υψίστης σημασίας ποσότητα $X^{-\frac{1}{2}} Y X^{-\frac{1}{2}}$, η οποία αποτελεί αυτό που ονομάζεται *congruent μετασχηματισμός*. Τέτοιου είδους μετασχηματισμοί παρέχουν την δυνατότητα επανατοποθέτησης ενός σημείου Y σε ένα πλαίσιο αναφοράς που ευθυγραμμίζεται με το σημείο βάσης, εδώ X , αναπαριστώντας έμπρακτα το πρώτο στο σύστημα συντεταγμένων όπου το δεύτερο είναι ο ταυτοτικός πίνακας. Χονδροειδώς, ο μετασχηματισμός ενός SPD πίνακα στο σύστημα συντεταγμένων όπου «ζει» ένας άλλος μπορεί να ιδωθεί ως μια αλλαγή βάσης η οποία διατηρεί τα γεωμετρικά χαρακτηριστικά του μεταφερόμενου πίνακα. Με την αναλύμενη ποσότητα να παράγει έναν χωρικά μετασχηματισμένο SPD πίνακα, μπορεί να εφαρμοστεί ο λογάριθμός του στον λογαριθμικά-Ευκλείδειο Χώρο (*log-Euclidean space*) ενώ τέλος η συμπλήρωση των ποσοτήτων του $X^{\frac{1}{2}}$ εκατέρωθεν αυτού επιτρέπει την σύζευξη/αναγωγή (*conjugation*) του αποτελέσματος πίσω στον Εφαπτομενικό Χώρο του X . Το τελικό αποτέλεσμα της λογαριθμικής απεικόνισης είναι ένα διάνυσμα, το οποίο εντοπίζεται στον διανυσματικό Χώρο των συμμετρικών πινάκων του Εφαπτομενικού Χώρου του X όπου το σημείο έναρξης είναι το τελευταίο και η κατεύθυνσή του υποδεικνύει προς το σημείο Y της SPD Πολλαπλότητας. Με άλλα λόγια, η λογαριθμική απεικόνιση λειτουργεί ως η γραμμική προσέγγιση της γεωδαισιακής απόστασης ανάμεσα στο σημείο X και το σημείο Y , παρέχοντας την πολυπόθητη υπολογιστική ευελιξία που αναζητείται στις εφαρμογές της Υπολογιστικής Όρασης [107].



Εικόνα 65: Απεικόνιση του διανύσματος y που υποδεικνύει προς την κατεύθυνση του σημείου Y ενώ εντοπίζεται στον Εφαπτομενικό Χώρο του X . Παράλληλα απεικονίζεται και η γεωδαισιακή που συνδέει μη-γραμμικά τα δυο σημεία του Πολύπτυχου. Με διακεκομμένη γραμμή σημειώνεται ο μετασχηματισμός της βάσης του Y στο σύστημα συντεταγμένων του Εφαπτομενικού Χώρου του X .

Με το y πλέον να αποτελεί ένα καλά ορισμένο διάνυσμα πάνω στον Εφαπτομενικό Χώρο του X , χρησιμοποιείται και ο τελετής διανύσματος (*vector operator*), όπως αυτός εμφανίζεται στο έργο του *Tuzel*, ώστε να μπορέσει να σχηματίσει την ελάχιστη μα συνάμα περιγραφική και συνεπή (*consistent*) διανυσματική αναπαράσταση των σημείων στον Εφαπτομενικό Χώρο διατηρώντας την γεωμετρική φύση τους άθικτη σε οποιονδήποτε εξωτερικό μετασχηματισμό της αρχικής εικόνας.

Έχοντας κατά νου τις παραπάνω θεωρητικές ιδέες εφαρμόζεται το παρακάτω σχέδιο:

- Για τον σχηματισμό της Ω^+ κλάσης: Ως Ω^+ κλάση ορίζεται η ομάδα εκείνη που ενσωματώνει την διαφορά της εκάστοτε γνήσιας υπογραφής ενός γραφέα με το επαγόμενο σύνολο συνδυασμών των ίδιων του των γνήσιων υπογραφών. Έτσι, η λογαριθμική απεικόνιση χρησιμοποιεί ως βάση την εκάστοτε υπό εξέταση γνήσια υπογραφή του κάθε γραφέα και υψώνει στον Εφαπτομενικό Χώρο αυτής όλες τις αμφισβητούμενες γνήσιες υπογραφές.
- Για τον σχηματισμό της Ω^- κλάσης: Ως Ω^- κλάση ορίζεται η ομάδα εκείνη που ενσωματώνει την διαφορά της εκάστοτε γνήσιας υπογραφής ενός γραφέα με το επαγόμενο σύνολο συνδυασμών των ίδιων του των πλαστογραφημένων υπογραφών. Έτσι, η λογαριθμική απεικόνιση χρησιμοποιεί ως βάση την εκάστοτε υπό εξέταση γνήσια υπογραφή του κάθε γραφέα και υψώνει στον Εφαπτομενικό Χώρο αυτής όλες τις αμφισβητούμενες πλαστογραφημένες υπογραφές του ίδιου ξανά.

Το σενάριο τόσο της Μάθησης όσο και του Ελέγχου παραμένει το ίδιο όπως ακριβώς έχει παρατεθεί και στο πρώτο Riemannian σενάριο. Στην συνέχεια παρατίθενται οι ψευδοκώδικες αλλά και ο πίνακας με τα πορίσματα της έρευνας αυτού του σεναρίου:

Algorithm 1 Second Riemannian Scenario

- 1: Clear workspace and initialize random seed
- 2: Define base and workspace directories
- 3: Create workspace directory if it does not exist
- 4: Delete any existing parallel pool
- 5: Initialize parallel pool with half of the maximum workers
- 6: Load covariance matrices from dataset
- 7: Separate genuine and forgery covariance matrices
- 8: Clear original covariance matrices to save memory
- 9: **for** each iteration from 1 to numIters **do**
- 10: Randomly sample indices for learning data
- 11: Shuffle genuine and forgery learning data using sampled indices
- 12: Create testing data by excluding sampled indices from original data
- 13: **for** each fold from 1 to numFolds **do**
- 14: Partition genuine data into training and validation sets
- 15: Partition forgery data into training and validation sets
- 16: Form $\omega(+)$ class using genuine training data
- 17: Form $\omega(-)$ class using forgery training data
- 18: Create labels for training and validation sets
- 19: Combine training and validation data into matrices
- 20: Perform hyperparameter optimization on SVM
- 21: Store best hyperparameters
- 22: Retrain SVM with best hyperparameters
- 23: Predict labels and scores for validation data
- 24: Compute ROC curve and AUC
- 25: Compute EER from ROC curve
- 26: Store results and indices for current fold
- 27: Swap learning and testing data for next fold
- 28: **end for**
- 29: **end for**
- 30: Save training data and results

Αλγόριθμος 12: Αλγόριθμος για το δεύτερο Riemannian σενάριο.

Algorithm 1 OmegaPlusFormation.m

Require: $G_Learning_Data, G_Metadata$
Ensure: $G_Training_set, G_Validation_set$

- 1: Initialize $G_Training_set$ and $G_Validation_set$ as empty
- 2: **for** each writer in $G_Learning_Data$ **do**
- 3: **for** each combination in $G_Metadata_Train_Combinations$ **do**
- 4: Set X as the basis matrix from $G_Metadata_Training_Signature_pairs$
- 5: Set Y as the target matrix from $G_Metadata_Training_Signature_pairs$
- 6: Compute $X_s = \sqrt{X}$
- 7: Compute the logarithmic mapping G_Train_AIRM
- 8: Compute the vector representation G_Train_Vec
- 9: **end for**
- 10: **end for**
- 11: **for** each writer in G_Train_AIRM **do**
- 12: Merge all pairs' data vertically
- 13: Append merged data to $G_Training_set$
- 14: **end for**
- 15: **for** each writer in $G_Learning_Data$ **do**
- 16: **for** each combination in $G_Metadata_Validation_Combinations$ **do**
- 17: Set X as the basis matrix from $G_Metadata_Validation_Signature_pairs$
- 18: Set Y as the target matrix from $G_Metadata_Validation_Signature_pairs$
- 19: Compute $X_s = \sqrt{X}$
- 20: Compute the logarithmic mapping G_Val_AIRM
- 21: Compute the vector representation G_Val_Vec
- 22: **end for**
- 23: **end for**
- 24: **for** each writer in G_Val_AIRM **do**
- 25: Merge all pairs' data vertically
- 26: Append merged data to $G_Validation_set$
- 27: **end for**
- 28: Return $G_Training_set$ and $G_Validation_set$

Αλγόριθμος 13: Αλγόριθμος για την συνάρτηση OmegaPlusFormation.m

Algorithm 1 OmegaMinusFormation.m

Require: $G_Learning_Data, F_Learning_Data, F_Metadata$
Ensure: $F_Training_set, F_Val_set$

- 1: Initialize $F_Training_set$ and F_Val_set as empty
- 2: Define $Signature_Training_Length$ from $F_Metadata$
- 3: **for** each writer in $G_Learning_Data$ **do**
- 4: Get genuine and forgery signatures for training
- 5: **for** each genuine signature **do**
- 6: **for** each forgery signature **do**
- 7: Compute $X_s = \sqrt{X}$
- 8: Compute $Sign_AIRM$ using logarithmic mapping
- 9: Compute $Sign_Vec$ using vector representation
- 10: **end for**
- 11: **end for**
- 12: Collect data into a single matrix
- 13: Randomly select 136 samples
- 14: Append selected data to $F_Training_set$
- 15: **end for**
- 16: **for** each writer in $G_Learning_Data$ **do**
- 17: Get genuine and forgery signatures for validation
- 18: **for** each genuine signature **do**
- 19: **for** each forgery signature **do**
- 20: Compute $X_s = \sqrt{X}$
- 21: Compute $Sign_Val_AIRM$ using logarithmic mapping
- 22: Compute $Sign_Val_Vec$ using vector representation
- 23: **end for**
- 24: **end for**
- 25: Collect data into a single matrix
- 26: Randomly select 21 samples
- 27: Append selected data to F_Val_set
- 28: **end for**
- 29: Return $F_Training_set$ and F_Val_set

Αλγόριθμος 14: Αλγόριθμος για την συνάρτηση OmegaMinusFormation.m

Algorithm 1 RefsQuestionNV.m

Require: *G_Testing_Data*, *F_Testing_Data*, *SVM*, *Options*

Ensure: *Data*, *Metrics*

```

1: for each Writer in G_Testing_Data do
2:   for each Testing_Iter from 1 to Options.numOfTestingIters do
3:     Print current Testing iteration and Writer
4:     Generate 10 random Reference Writers from G_Testing_Data
5:     Generate 24 random indices for Forgery data from F_Testing_Data
6:     Create remaining indices for Question set from G_Testing_Data
7:     Assign Reference and Question data to Data
8:     Create Ground Truth labels for Question data
9:     for each Question in Data(Writer).Question do
10:      for each Reference in Data(Writer).Refs do
11:        Compute square root of Reference matrix
12:        Compute logarithmic mapping for AIRM
13:        Compute vector representation for Tangent Plane
14:      end for
15:      Concatenate vectors into a matrix
16:      Predict labels and scores using SVM
17:      Store scores and maximum score
18:      Select label with maximum score
19:    end for
20:    Compute ROC curve and AUC
21:    Compute FAR, FRR, and EER
22:    Store EER and AUC in Metrics
23:  end for
24:  Compute mean EER per Writer
25: end for
26: Return Data and Metrics

```

Αλγόριθμος 15: Αλγόριθμος για την συνάρτηση RefsQuestionNV.m

Πίνακας 8: Αποτίπωση του EER ανά επανάληψη κι ανά fold καθώς και του συνολικού μέσου EER για το πρώτο Riemannian σενάριο.

Επανάληψη	Fold 1	Fold 2
1	1.1011	2.2531
2	1.1607	0.0463
3	0.3720	1.4969
4	4.7024	0.1389
5	0.2827	1.8827
Μέσο EER	1.3437	

5 ΚΕΦΑΛΑΙΟ 5^ο : Συμπεράσματα

Η παρούσα εργασία ξεκίνησε με μια υπόσχεση την οποία κατάφερε να εκπληρώσει! Αρχίζοντας από τον Ευκλείδειο Χώρο, όπου μελετήθηκαν οι αλγόριθμοι του BoVW, του SPM, των πυρήνων Fisher καθώς και του VLAD, αποτυπώθηκε με φυσικό και συνδεδετικό τρόπο πώς η διανυσματική αναπαράσταση των εικόνων που υποδείχθηκε για πρώτη φορά από τον αλγόριθμο του BoVW «έστρωσε» τον δρόμο για την ενορχήστρωση μιας νέας σειράς από συνεχόμενες επεκτάσεις που ενισχύουν την ακρίβεια των ταξινομητών υιοθετώντας όλο και πιο εξελιγμένες προσεγγίσεις στον τρόπο διαχείρισης των δεδομένων. Παράλληλα, επιβεβαιώθηκε πως η ενδογενής συνάρτηση του MATLAB, *fitgmdist*, αδυνατεί να παράξει τα απαραίτητα στατιστικά στοιχεία που χρειάζονται για τον σχηματισμό των Λεξικών μέσω των GMM στην μέθοδο των πυρήνων Fisher τόσο από την άποψη του τερματισμού λειτουργίας του προγράμματος λόγω δυσμενώς-οριζόμενων πινάκων συνδιακύμανσης (*ill-conditioned covariances*) όσο κι από την επιβολή χρονικού αδιεξόδου. Στον αντίποδα φωτίστηκε πως τόσο η εξατομικευμένη συνάρτηση *GMM_NV* όσο και η υλοποίηση του

sEM κατάφεραν να φέρουν εις πέρας το πρόβλημα με την παράλληλη επιστράτευση καρτών γραφικών να ενισχύει την επίσπευση ανάκτησης των αποτελεσμάτων. Ενδεικτικό μάλιστα είναι το γεγονός ότι στην περίπτωση που η *fitgmdist* μπορούσε όντως να ανταποκριθεί, η μεθοδολογία που ακολουθήθηκε καταφέρνει να μειώσει τους χρόνους λήψης των αποτελεσμάτων από εβδομάδες, σε μέρες (για την *GMM_NV*) και σε ώρες αντίστοιχα (για τον *sEM*) για την εκτέλεση μιας μόνο επανάληψης σε όλες τις προαναφερόμενες περιπτώσεις. Προκειμένου όμως η συγκριτική διαδικασία να είναι πλήρης και το ερευνητικό έργο να είναι ορθό, με γνώμονα το γεγονός ότι ο αλγόριθμος του BoVW ελέγχθηκε σε μια διαφορετική βάση δεδομένων από αυτή που εξετάστηκαν οι υπόλοιποι αλγόριθμοι, διεξήχθη ένα ακόμη πείραμα το οποίο εκτελέστηκε 10 φορές για όλους τους αλγόριθμους εκτός από τους πυρήνες Fisher, αυτή τη φορά όμως με κοινή βάση δεδομένων. Ο παρακάτω πίνακας συνοψίζει, αποτυπώνει κι επιβεβαιώνει την εξελικτική πορεία που εμφανίζεται ανά τους αλγορίθμους:

Πίνακας 9: Τελικός πίνακας αποτύπωσης της αύξησης της ακρίβειας ανά τους αλγορίθμους.

	<i>BoVW</i>	<i>SPM</i>	<i>Fisher Kernels</i>	<i>VLAD</i>
Μέση ακρίβεια	63.50	73.33	76.91	81.60

Διευκρινίζεται ότι τα αποτελέσματα του παραπάνω πίνακα είναι σχετιζόμενα άμεσα με την εκάστοτε βάση στην οποία εφαρμόζεται. Αυτό σημαίνει ότι υπάρχει η πιθανότητα σε διαφορετικές βάσεις κάποιοι αλγόριθμοι να αποδίδουν πολύ καλύτερα από το σχηματικό του VLAD απλά και μόνο επειδή καταφέρνουν να συλλάβουν καλύτερα με την μεθοδολογία που τους περιγράφει τις υπο-εξέταση εικόνες. Κλείνοντας, αναφερόμενοι στο κομμάτι του Riemannian Χώρου και στρέφοντας την προσοχή στους πίνακες 7 και 8 μπορεί πολύ εύκολα να εξαχθεί το συμπέρασμα πως το αμάλγαμα της χρήσης της εγγενούς φύσης του SPD Πολύπτυχου με την γραμμική προσέγγιση της απόστασης που επάγεται από τον Εφαπτομενικό Χώρο του σημείου της βάσης κατάφερε να ενισχύσει την περιγραφικότητα της εφαρμογής της αυτόματης επαλήθευσης χειρόγραφων υπογραφών σε σχέση με την αποκλειστική χρήση του Εφαπτομενικού Χώρο ρίχνοντας την τιμή του μέσου EER από το 11% στο 1%.

6 Προτάσεις προς επέκταση

Το πέρας της παρούσας διπλωματικής δεν σηματοδοτεί και το τέλος της ανάλυσης. Βασιζόμενοι στο πνεύμα της συνεχούς εξέλιξης, παρακάτω παρατίθεται μια πιθανή λίστα από επεκτάσεις που μπορούν να εφαρμοστούν στο πλαίσιο επέκτασης αυτής της εργασίας:

Για τον Ευκλείδειο Χώρο:

1. Αντικαθιστώντας τα GMM μπορεί να μελετηθεί η χρήση Fuzzy C-means ως μηχανισμός παραγωγής των κέντρων στα οποία ανήκουν οι εικόνες. Με τα GMM να αναθέτουν κάθε σημείο σε μια συγκεκριμένη ομάδα με μια κάποια τιμή πιθανότητας, ο Fuzzy C-means

επιτρέπει μια μερική υπαγωγή, με την έννοια ότι ένα δεδομένο μπορεί να ανήκει σε πολλές ομάδες με έναν μεταβλητό βαθμό συμμετοχής κάτι που προσομοιάζει αρκετά την καθημερινότητα.

Για τον Riemannian Χώρο:

1. Δεδομένης της μη-γραμμικής φύσης του προβλήματος, θα μπορούσε να χρησιμοποιηθεί ένας μη-γραμμικός ταξινομητής, όπως για παράδειγμα ένα νευρωνικό δίκτυο το οποίο έχει εκπαιδευτεί ώστε να μπορεί να «συλλάβει» την Riemannian δομή και φύση των SPD πινάκων.
2. Αν κανείς επιθυμεί να παραμείνει πιστός στην χρήση των SVM για την επίλυση δυαδικών προβλημάτων, μια ενδεχόμενη πρόταση προς ενίσχυση της αποδοτικότητάς τους είναι να σχηματιστεί ένας πυρήνας ο οποίος να μπορεί να «συλλάβει» την φύση του Riemannian Πολύπτυχου των SPD πινάκων.
3. Μια διαφορετική μεθοδολογία είναι να πραγματοποιηθεί μάθησης μετρικών (*metric learning*). Με μια πληθώρα από διαθέσιμες έρευνες στην βιβλιογραφία να υιοθετούν όλο και περισσότερο αυτήν την προσέγγιση, η χρήση μάθησης μετρικών επιτρέπει την αποδοτικότερη απόφαση βασισμένη στην απόσταση των δειγμάτων, καταφέρνει να προσαρμοστεί στην μεταβλητότητα και μειώνει την εξάρτηση από την μεθόδευση προσεκτικών τεχνικών για την εξαγωγή χαρακτηριστικών [108].
4. Μια out of the box ιδέα θα ήταν η χρήση γενετικών αλγορίθμων. Με τους τελευταίους να ξεκινούν από την εύρεση καλύτερων υπερπαραμέτρων στον Χώρο της βελτιστοποίησης αυτών μέχρι και την χρήση τους για την επιλογή χαρακτηριστικών, οι γενετικοί αλγόριθμοι έχουν αρχίσει ήδη να απασχολούν τον τομέα της αυτόματης επαλήθευσης υπογραφών και αποτελεί μια πιθανή πορεία προς εξέταση [109], [110], [111].

Παράρτημα Α: Ο χώρος του Riemann - Πολλαπλότητες

Στην επιστήμη των μαθηματικών ορίζεται πως:

Ορισμός 34: Πολλαπλότητα ή Πολύπτυχο (Manifold)

Μια τοπολογική Πολλαπλότητα (ή Πολύπτυχο) είναι ένας τοπολογικός Χώρος \mathcal{M} που προσομοιάζει τοπικά τον Ευκλείδειο Χώρο \mathbb{R}^n . Συγκεκριμένα, η n -διάστατη Πολλαπλότητα είναι ένας τοπολογικός Χώρος που φέρει την ιδιότητα κατά την οποία για κάθε σημείο μπορεί να βρεθεί μια γειτονιά του που είναι ομοιομορφισμός σε ένα ανοιχτό σύνολο ενός n -διάστατου Ευκλείδειου Χώρου [55], [56], [57], [58], [59], [60].

Για να μπορέσει να αποδοθεί μια πιο πρακτική προσέγγιση και να γίνει καλύτερα κατανοητό, κανείς μπορεί να αναλογιστεί τα εξής:

- Οι γραμμές (*lines*) και οι κύκλοι (*circles*) είναι μονοδιάστατα Πολύπτυχα.
- Η σφαίρα (*sphere*), το επίπεδο (*plane*) και ο τόρος (*torus*) είναι δισδιάστατα Πολύπτυχα ή διαφορετικά επιφάνειες (*surfaces*).

Χώρος και αριθμοί

Με εκτενή και επανειλημμένο τρόπο σε αυτήν την εργασία έχει γίνει χρήση της λέξης «χώρος» χωρίς να έχει εμπεριστατωθεί η ετυμολογία του. Σύμφωνα με τον *Liviu I. Nicolaescu* [61]: «Ο Χώρος είναι το πιο εμφανές φυσικό αντικείμενο, είναι το μέρος όπου όλα τα συμβάντα μπορούν να εκτυλιχθούν». Η απλή όμως περιέλιξη αυθαίρετων δομών αντικειμένων από τον Χώρο δεν επαρκεί καθώς χρειάζονται τρόποι οι οποίοι να μπορούν ανά πάσα χρονική στιγμή και χωρική τοποθεσία να μπορούν να περιγράψουν την ύπαρξή τους μέσα σε αυτόν με τρόπο ουσιώδη, ντετερμινιστικό και εύρωστο. Ο θεμελιωτής αυτής της περιγραφής θεωρείται ο Ευκλείδης, ο οποίος μέσω των αξιωματικών μεθόδων του περιέγραψε το Χώρο ως μια συλλογή σημείων, η χωροταξική θέση των οποίων όταν γίνεται με τρόπο προκαθορισμένο και συγκεκριμένο σχηματίζουν βασικές μορφολογίες που οδηγούν στην έγερση εγγενών συσχετίσεων μεταξύ τους [61].

Αυτή μονάχα η ιδέα αποτέλεσε το εφαλτήριο για την επανάσταση που έφερε η δουλειά του *René Descartes* με την εισαγωγή των συντεταγμένων και την άμεση τοποθέτηση των αριθμών στην μελέτη του Χώρου. Με την εισαγωγή του συστήματος συντεταγμένων (*coordinatization*) να αποτελεί την γεφύρωση ανάμεσα στην Γεωμετρία και την Άλγεβρα⁶⁰, η μελέτη του Χώρου μπορεί πολύ εύκολα να

⁶⁰ Μέσω των ένα-προς-ένα αντιστοιχιών, κάθε γραμμή (*line*) ή επίπεδο (*plane*) μέσω του συντονισμού μπορεί να αναχθεί σε ένα Αλγεβρικό αντικείμενο προς επεξεργασία.

οδηγηθεί στην μελέτη των εξισώσεων αλλά και αντίστροφα. Όμως εδώ εγείρεται ένα πρόβλημα. Η οπτική του Χώρου, άρα και η διαδικασία αρχικοποίησης των συντεταγμένων σε αυτόν, που μπορεί να υιοθετήσει κανείς είναι κάτι τόσο αυθαίρετο, που θέτει το δομημένο αυτό σύστημα υπο αμφισβήτηση [61].

Μια υποτυπώδης λύση θα ήταν να *αλλάξει η περιγραφή των αντικειμένων* ώστε να γίνει ανεξάρτητη των συντεταγμένων τους στον Χώρο. Αυτό το αντικείμενο μπορεί να το προσφέρει η διαφορική γεωμετρία και αυτός είναι ένας από τους λόγους για τον οποίο η μελέτη αυτής είναι μείζονος σημασίας και έχει γνωρίσει ραγδαίο ενδιαφέρον. Η διαφορική δομή των Πολλαπλοτήτων «δημιουργεί» Χώρους που μπορούν να συντεταχθούν τμηματικά (*piecewise coordinatized*), με λεία απεικόνιση στις επικαλύψεις τους. Οι επόμενες υποενότητες κάνουν μια «βουτιά» σε αναγκαίες μαθηματικές έννοιες με τρόπο συνεκτικό και αλυσιδωτό ώστε να σχηματίσουν ένα άρρηκτα δομημένο και εύλογο μαθηματικό πλαίσιο [61].

Ομοιομορφισμός και Διαφορομορφισμός (*Homeomorphism & Diffeomorphism*)

Ανωτέρω σημειώθηκε πως η ανάλυση, τόσο στην απλή όσο και στην διαφορική Γεωμετρία, βασίζεται σε ένα διττό πλαίσιο επεξεργασίας: Αφενός θα πρέπει να βρεθεί ένας τρόπος συσχέτισης των γεωμετρικών δομών των αντικειμένων μέσα στον Χώρο και αφετέρου αυτή η συσχέτιση θα πρέπει να γίνεται με τρόπο ο οποίος είναι ανεξάρτητος των χωρικών συντεταγμένων των αντικειμένων. Γίνεται λοιπόν αντιληπτό ότι θα πρέπει να σχηματιστούν κάποιας μορφής αντιστοιχίες ανάμεσα στα αντικείμενα που μελετώνται. Το σημαντικό αυτό έργο δομείται από τις έννοιες του ομοιομορφισμού και του διαφορομορφισμού, όπως αυτές παρατίθενται παρακάτω.

Ορισμός 35: Ομοιομορφισμός

Μια συνάρτηση $f: X \rightarrow Y$ ανάμεσα σε δυο τοπολογικούς χώρους αποκαλείται ομοιομορφισμός αν επαληθεύει τις παρακάτω ιδιότητες:

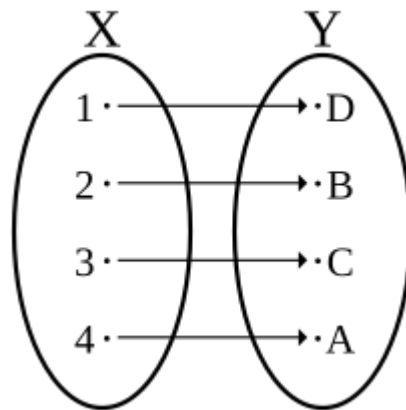
1. Είναι αμφιμονοσήμαντη αντιστοιχία (*bijection*).
2. Είναι συνεχής.
3. Η αντίστροφη συνάρτησή της, f^{-1} , είναι συνεχής.

Η πρακτική σημασία πίσω από τους ομοιομορφισμούς είναι ότι είναι οι *ισομορφισμοί*⁶¹ των τοπολογικών Χώρων, δρουν δηλαδή ως οι απεικονίσεις εκείνες που επιτρέπουν την διατήρηση των

⁶¹ Ο *ισομορφισμός* είναι μια διαδικασία απεικόνισης που επιτρέπει την διατήρηση της δομής ανάμεσα σε δυο επιμέρους δομές του ίδιου τύπου, με την προϋπόθεση ότι μπορούν να αντιστραφούν μέσω της αντίστοιχης αντίστροφης απεικόνισης [62].

τοπολογικών ιδιοτήτων ενός δοσμένου Χώρου. Στο σημείο αυτό, ως ένδειξη καλής πρακτικής, θα πρέπει να γίνει μια σύντομη αναφορά και στην έννοια της *αμφιμονοσήμαντης αντιστοιχίας*, η οποία πρακτικά είναι η γνωστή *συνάρτηση αντιστοιχίας ένα-προς-ένα και επί (one-to-one and onto correspondence function)* ανάμεσα σε δυο σύνολα με τέτοιο τρόπο ώστε κάθε στοιχείο του δεύτερου συνόλου να μπορεί να αντιστοιχηθεί με μόνο ένα στοιχείο του πρώτου συνόλου και αντίστροφα [63],[64].

Στο ακριβώς ίδιο πνεύμα, ο *διαφορομορφισμός* λειτουργεί ως ο *ισομορφισμός* αλλά για τις διαφορίσιμες Πολλαπλότητες (*differentiable Manifolds*), οριζόμενος ως μια αντιστρέψιμη συνάρτηση που δομεί την αντιστοίχιση ανάμεσα σε δυο διαφορίσιμα Πολύπτυχα έτσι ώστε τόσο η ίδια η συνάρτηση όσο και η αντίστροφή της να είναι συνεχώς διαφορίσιμες. Συγκεκριμένα:

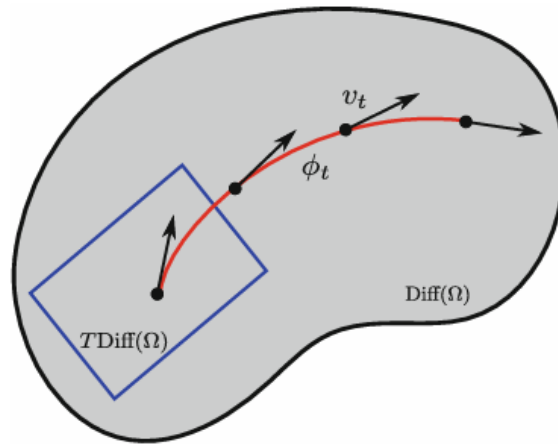


Εικόνα 66: Απεικόνιση της έννοιας της αμφιμονοσήμαντης συνάρτησης (bijection) ⁶².

Ορισμός 36: Διαφορομορφισμός

Δοσμένων δυο διαφορίσιμων Πολύπτυχων M και N , μια διαφορίσιμη αντιστοιχία $f: M \rightarrow N$ είναι διαφορομορφισμός αν είναι αμφιμονοσήμαντη συνάρτηση και ο αντίστροφός της $f^{-1}: N \rightarrow M$ είναι επίσης διαφορίσιμος [60], [61], [65], [66], [67], [68].

⁶² Η εικόνα αντλήθηκε από τον ακόλουθο [σύνδεσμο](#).



Εικόνα 67: Απεικόνιση του χώρου των διαφορομορφισμών⁶³.

Με τους ορισμούς του ομοιομορφισμού και του διαφορομορφισμού να είναι πλήρεις και την έμμεση εισαγωγή των συνόλων στο προσκήνιο, παρέχεται προσοδοφόρο έδαφος για την ανανέωση του Ορισμού 34 από το έργο του *John M. Lee* ως [68]:

Ορισμός 37: Τοπολογική Πολλαπλότητα n -διαστάσεων.

Με την υπόθεση ότι το M είναι ένας τοπολογικός Χώρος, τότε το M είναι ένα τοπολογικό Πολύπτυχο n -διαστάσεων αν ικανοποιεί τις παρακάτω συνθήκες [68]:

- **Είναι ένας Hausdorff Χώρος:** Δηλαδή για κάθε ζευγάρι διακεκριμένων σημείων $p, q \in M$, υπάρχουν ζένα (κενή τομή) ανοικτά υποσύνολα $U, V \subseteq M$ έτσι ώστε $p \in U$ και $q \in V$.
- **Είναι δεύτερος αριθμήσιμος (second countable):** Το ότι είναι δεύτερος αριθμήσιμος αναφέρεται στην ύπαρξη μιας αριθμήσιμης βάσης για την τοπολογία ενός Χώρου. Ανάγοντάς την στο πλαίσιο της τοπολογικής Πολλαπλότητας, αυτό σημαίνει ότι υπάρχει μια αριθμήσιμη συλλογή ανοικτών συνόλων, έτσι ώστε κάθε ανοικτό σύνολο σε αυτή να μπορεί να εκφραστεί ως ένωση από σύνολα της συλλογής αυτής [69], [70], [71].
- **Είναι ένας τοπικά Ευκλείδειος n -διάστατος χώρος:** Κάθε σημείο του M έχει μια γειτονιά που είναι ομοιομορφική σε ένα ανοικτό υποσύνολο του \mathbb{R}^n .

Λείες Πολλαπλότητες (Smooth Manifolds)

Παραπάνω έγινε έντονη αναφορά στην έννοια της διαφορισιμότητας που απαιτεί την έννοια της παραγώγου. Η περιγραφικότητα που προκύπτει από τις τελευταίες για φυσικές δομές όπως είναι οι καμπύλες ή ακόμη και για παραμετρικές συναρτήσεις πραγματικών τιμών, επιτάσσουν τον ορισμό μιας νέας μορφής των Πολλαπλοτήτων: τις *Λείες Πολλαπλότητες*.

⁶³ Η εικόνα αντλήθηκε από την πηγή [66].

Σε ένα διαισθητικό επίπεδο κατανόησης, κανείς θα μπορούσε να αναλογιστεί το εξής παράδειγμα: Έστω ότι ένα αντικείμενο προς διερεύνηση είναι μια τυπική μπάλα και σε αυτήν σημειώνεται με ένα ευδιάκριτο χρώμα, έστω μαύρο στην λευκή της επιφάνεια, ένα σημείο. Συνεχίζοντας, έστω ότι τώρα πάνω σε αυτό το σημείο προσκολλάται ένας χάρακας, δηλαδή ένα επίπεδο παραλληλόγραμμης μορφολογίας και δίνεται μια μικρή ώθηση στο συνολικό σύστημα. Αν λοιπόν τώρα ληφθούν αλληλάλληλα στιγμιότυπα μικρών χρονικών παραθύρων στην διάρκεια της κίνησης του συνολικού συστήματος, κανείς μπορεί να παρατηρήσει ότι η κίνηση του χάρακα μεταβάλλεται συνεχόμενα και ομαλά με την κίνηση της μπάλας ενώ το σημείο στο οποίο είναι προσκολλημένος διαγράφει μια καμπύλη από την θέση έναρξης μέχρι την θέση στην οποία ακινητοποιείται. Ανάγοντας το σημείο στην επιφάνεια της μπάλας με ένα σημείο το οποίο εντοπίζεται πάνω σε μια Πολλαπλότητα και τον χάρακα με την κλίση ή διαφορετικά την παράγωγο που περιγράφει την κίνησή του, μπορεί να οριστεί πως:

Ορισμός 38: Λεία Πολλαπλότητα (Smooth Manifold)

Μια Πολλαπλότητα λέγεται ότι έχει Λεία δομή, όταν αυτή είναι συνεχώς διαφορίσιμη/παραγωγίσιμη. Με την δομή του Πολύπτυχου να περιγράφεται από μια συνάρτηση, αν U και V είναι δυο ανοιχτά υποσύνολα των Ευκλείδειων Χώρων \mathbb{R}^n και \mathbb{R}^m αντίστοιχα, τότε μια συνάρτηση $F: U \rightarrow V$ λέγεται πως είναι **λεία** (ή C^∞ ή απείρως διαφορίσιμη) αν κάθε μία από τις συναρτήσεις που την αποτελούν έχει συνεχείς μερικές παραγώγους όλων των τάξεων [68] [60]:

$$\partial^\alpha f = \frac{\partial^{\alpha_1 + \dots + \alpha_n} f}{\partial x_1^{\alpha_1} \dots \partial x_n^{\alpha_n}}, \quad \alpha = (\alpha_1, \dots, \alpha_n) \in N_{\mathbb{0}}^n$$

Ο Ορισμός 38 αν και φαίνεται πως είναι πλήρης, στην πραγματικότητα δεν είναι. Του λείπει ένα ακόμη δομικό στοιχείο το οποίο απαιτεί την εισαγωγή επιπρόσθετων εννοιών ώστε να μπορέσει να σχηματιστεί το τελικό μαθηματικό πλαίσιο περί της Ομαλότητας των Πολλαπλοτήτων.

Χάρτες Συντεταγμένων (Coordination Charts)

Επιστρέφοντας στην αρχική ορολογία των τοπικών Πολλαπλοτήτων, ορίζεται ότι:

Ορισμός 39: Χάρτες Συντεταγμένων

Έστω ξανά ότι το M είναι μια τοπολογική Πολλαπλότητα n -διαστάσεων. Ένας Χάρτης Συντεταγμένων στο M είναι ένα ζευγάρι (U, φ) , όπου του U είναι ένα ανοιχτό υποσύνολο του M και $\varphi: U \rightarrow \hat{U}$ είναι ένας ομοιομορφισμός από το U προς ένα ανοιχτό υποσύνολο $\hat{U} = \varphi(U) \subseteq \mathbb{R}^n$ [60], [68], [72].

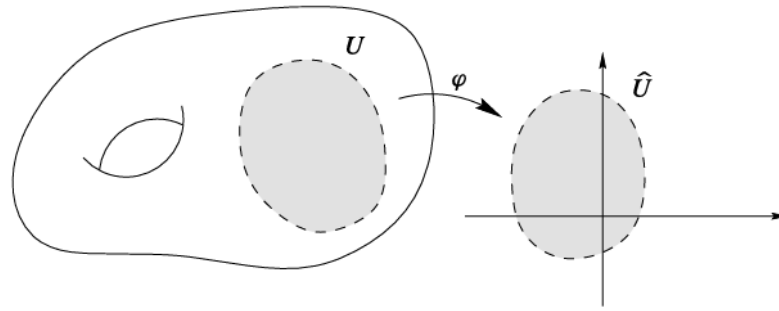


Fig. 1.2 A coordinate chart

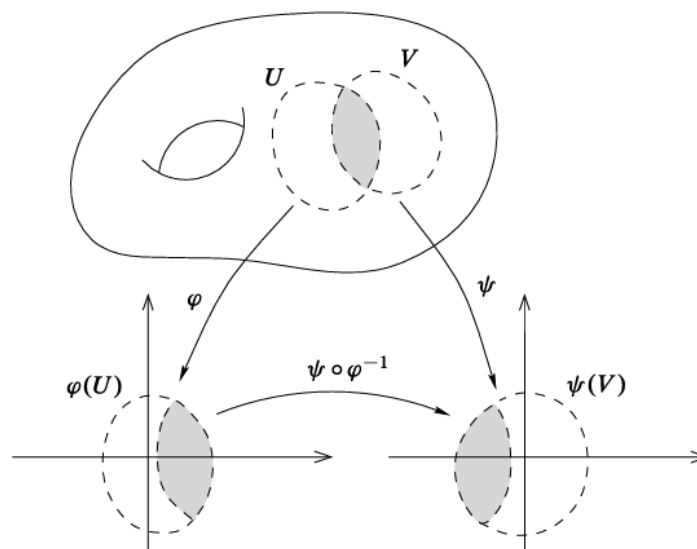
Εικόνα 68: Απεικόνιση ενός Χάρτη Συντεταγμένων⁶⁴.

Μεταβατικές Απεικονίσεις (Transition Maps)

Όπως αναφέρθηκε και παραπάνω, η διερεύνηση της συσχέτισης των Χαρτών Συντεταγμένων επιτρέπει την διευκόλυνση της μελέτης των συσχετίσεων των Πολλαπλοτήτων στις οποίες αντιστοιχούν, καθώς άρει την πολυπλοκότητα που εισάγουν οι θυγατρικές δομές τους. Προκειμένου να μπορέσει να πλαισιωθεί ένα μέτρο σύγκρισης όμως, χρειάζεται να γεφυρωθεί κάπως η μετάβαση από τον ένα Χάρτη στο άλλο. Τον ρόλο αυτό αναλαμβάνουν οι Μεταβατικές Απεικονίσεις. Συγκεκριμένα:

Ορισμός 40: Μεταβατικές Απεικονίσεις (Transition Maps)

Με το M και πάλι να είναι μια τοπολογική Πολλαπλότητα n -διαστάσεων, αν με (U, φ) και (U, ψ) οριστούν δυο Χάρτες έτσι ώστε $U \cap V \neq \emptyset$, τότε η σύνθετη απεικόνιση $\psi \circ \varphi^{-1}: \varphi(U \cap V) \rightarrow \psi(U \cap V)$ ονομάζεται **Μεταβατική Απεικόνιση από το φ στο ψ** [68].



Εικόνα 69: Αποτύπωση των μεταβατικών απεικονίσεων ανάμεσα σε δυο Χάρτες συντεταγμένων⁶⁵

⁶⁴ Η παρούσα εικόνα έχει ληφθεί από την πηγή [68]

⁶⁵ Η παρούσα εικόνα λήφθηκε από την πηγή [68].

Συμπληρώνοντας στον Ορισμό 40:

Ορισμός 41: Λεία συμβατές Μεταβατικές Απεικονίσεις

Δύο Χάρτες (U, φ) και (U, ψ) ορίζεται ότι είναι λεία συμβατά⁶⁶ αν είτε το $U \cap V \neq \emptyset$ είτε η μεταβατική απεικόνιση $\psi \circ \varphi^{-1}$ είναι διαφορομορφισμός. Δεδομένου ότι τόσο το $\varphi(U \cap V)$ όσο και το $\psi(U \cap V)$ είναι ανοιχτά υποσύνολα του \mathbb{R}^n , η ομαλότητα αυτής της απεικόνισης εκλαμβάνεται όπως έχει αποδοθεί και στον Ορισμό 38, δηλαδή ότι υπάρχουν συνεχείς μερικές παράγωγοι όλων των τάξεων [68].

Άτλαντες και Ομαλοί Άτλαντες

Η απόθεση των Χαρτών Συντεταγμένων, επιτρέπει την έγερση ενός σύνθετου στοιχείου το οποίο εγγενώς συντίθεται από αυτά του Άτλαντα.

Ορισμός 42: Άτλαντας

Ένας Άτλαντας του M είναι μια συλλογή από Χάρτες των οποίων τα πεδία (domains) καλύπτουν το M [60], [68]:

$$\mathcal{A} = \{(\varphi_\alpha, U_\alpha)\}_{\alpha \in A} \quad (88)$$

$$\text{Δεδομένου ότι } M = \bigcup_{\alpha \in A} U_\alpha$$

Η αποτύπωση του Άτλαντα ενισχύεται με την εισαγωγή της ιδιότητας της Ομαλότητας σε αυτόν, αποτυπώνοντας ότι:

Ορισμός 43: Λείοι Άτλαντες

Ένας Άτλαντας αποκαλείται λείος αν οποιαδήποτε δυο Χάρτες του είναι λεία συμβατά μεταξύ τους [60], [61], [68].

Η έννοια του Μέγιστου Λείου Άτλαντα

Συμπληρωματικά στην οικογένεια των ορισμών του Άτλαντα έρχεται να προστεθεί και ο τελευταίος και ίσως πιο κρίσιμος ορισμός:

Ορισμός 44: Ολοκληρωμένος Άτλαντας (Maximal/Complete Atlas)

Ένας λείος Άτλαντας \mathcal{A} στο M είναι ολοκληρωμένος αν δεν περιλαμβάνεται σε οποιονδήποτε μεγαλύτερο λείο Άτλαντα. Αυτό οδηγεί στην διατύπωση πως οποιοδήποτε Χάρτης που είναι λεία συμβατός με κάθε Χάρτη στο \mathcal{A} , πολύ απλά ανήκει ήδη στο \mathcal{A} [60], [68].

⁶⁶ Δύο Χάρτες (U, φ) και (U, ψ) λέγεται ότι είναι **συμβατά** αν το $\varphi(U \cap V)$ και το $\psi(U \cap V)$ είναι ανοιχτά σύνολα και η Μεταβατική Απεικόνιση τους είναι διαφορομορφισμός.

Έχοντας αυτά τα μαθηματικά εργαλεία, είναι πλέον δυνατόν να επιτευχθεί ο αρχικός στόχος που τέθηκε, αναδιατυπώνοντας τον Ορισμό 38 ως:

Ορισμός 45: Λεία δομή Πολλαπλότητας (revisited).

Αν M είναι μια τοπολογική Πολλαπλότητα, τότε μια λεία δομή στο M είναι ένας ολοκληρωμένος λείος Άτλαντας. Ένα λοιπόν λείο Πολύπτυχο, είναι ένα ζεύγος (M, \mathcal{A}) , όπου το M είναι το τοπολογικό Πολύπτυχο και το \mathcal{A} είναι μια λεία δομή σε αυτό [60], [68].

Επειδή ακριβώς η Πολλαπλότητα καθορίζεται από το υποβόσκων σύνολο M και μερικούς Άτλαντες, συνήθως στην βιβλιογραφία επιλέγεται να αποσιωπείται η σημειογραφία του τελευταίου και να χρησιμοποιείται μόνο το M .

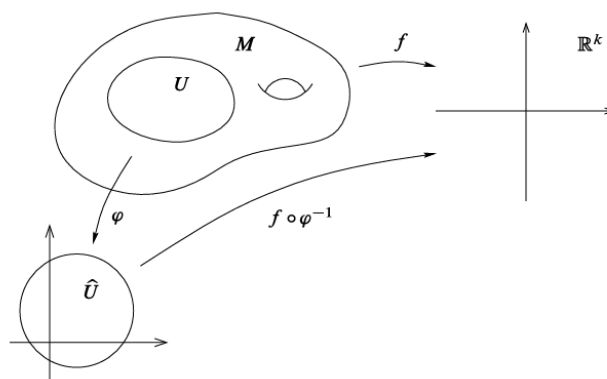
Λείες Συναρτήσεις και Λείες Απεικονίσεις (Smooth Functions and Smooth Maps)

Κλείνοντας την παρούσα υποενότητα, είναι σημαντικό να φωτιστούν οι έννοιες των Λείων Συναρτήσεων που ενισχύουν τις Λείες Απεικονίσεις μεταξύ διαφορετικών Πολλαπλοτήτων.

Ξεκινώντας από τις Λείες Απεικονίσεις αποτυπώνεται ότι:

Ορισμός 46: Λεία Συνάρτηση

Εστω ότι πλέον το M είναι ένα λείο n -διάστατο Πολύπτυχο, ως k ορίζεται ένας μη αρνητικός ακέραιος αριθμός και με $f: M \rightarrow \mathbb{R}^k$ ορίζεται μια συνάρτηση. Η f αποκαλείται **λεία συνάρτηση** αν για κάθε $p \in M$ υπάρχει ένας λείος Χάρτης (U, φ) του M του οποίου το πεδίο περιλαμβάνει το p και η σύνθετη συνάρτηση $f \circ \varphi$ είναι λεία στο ανοιχτό υποσύνολο $\hat{U} = \varphi(U) \subseteq \mathbb{R}^n$ [68].



Εικόνα 70: Απεικόνιση του ορισμού μιας λείας συνάρτησης⁶⁷.

Η πιο σημαντική από όλες τις περιπτώσεις που μπορούν να εμφανιστούν είναι αυτή των λείων συναρτήσεων πραγματικών τιμών $f: M \rightarrow \mathbb{R}$, διότι τα αθροίσματα και τα σταθμίσιμα με σταθερές αυτών των λείων συναρτήσεων είναι επίσης λείες συναρτήσεις, ενώ το σύνολο αυτών των λείων

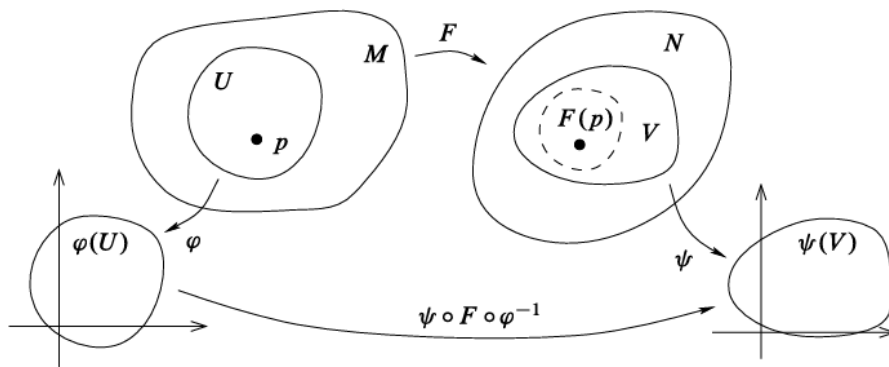
⁶⁷ Η παρούσα εικόνα αντλήθηκε από την πηγή [68].

συναρτήσεων που δηλώνονται με την σημειογραφία $C^\infty(M)$ είναι ένας διανυσματικός χώρος στον \mathbb{R} [68].

Τέλος οι Λείες Απεικονίσεις ορίζονται ως:

Ορισμός 47: Λείες Απεικονίσεις (Smooth Maps).

Εστω ότι με M, N ορίζονται δυο λείες Πολλαπλότητες και με $F: M \rightarrow N$ είναι μια απεικόνιση. Ορίζεται ότι το F είναι μια **λεία απεικόνιση** αν για κάθε $p \in M$ υπάρχουν λείοι Χάρτες (U, φ) που περιέχουν το p και (V, ψ) που περιέχουν το $F(p)$ έτσι ώστε το $F(U) \subseteq V$ και η σύνθετη απεικόνιση $\psi \circ F \circ \varphi^{-1}$ είναι λεία από το $\varphi(U)$ στο $\psi(V)$ [68].



Εικόνα 71: Αποτύπωση του ορισμού των Λείων Απεικονίσεων⁶⁸.

Παράρτημα Β: Βασικοί όροι της Διαφορικής Γεωμετρίας

Βασικοί όροι της Διαφορικής Γεωμετρίας

Μέχρι τώρα στο μαθηματικό υπόβαθρο αναλύθηκαν έννοιες οι οποίες είναι εγγενώς μη-γραμμικές, κάτι το οποίο αφενός είναι εξαιρετικά προσοδοφόρο καθώς δίνει την δυνατότητα ενίσχυσης της περιγραφικότητας των υποκείμενων δομών, αλλά αφετέρου αποτελεί υπολογιστικό τροχοπέδη τόσο από την άποψη φόρτου όσο και από την άποψη χρόνου περάτωσης καθώς οι συνηθισμένοι τελεστές που είθισται να χρησιμοποιούνται καθίστανται παρωχημένοι και η χρήση της Άλγεβρας μπλοκάρει. Ένα πρώτο μέτρο που λήφθηκε ήταν η αξιοποίηση συγκεκριμένων μαθηματικών ιδιοτήτων, οι οποίες δίνουν την δυνατότητα της μεταφοράς σε τοπικούς χώρους που άρουν την σύνθετη πολυπλοκότητα των Πολύπτυχων χωρίς όμως το αρχικό πρόβλημα να μπορεί να επιλυθεί.

Με την Άλγεβρα να αποτελεί διαισθητικά ένα μέσο για την επίτευξη γραμμικής προσέγγισης άρα και την παραγωγή γραμμικών μοντέλων τα οποία μπορούν να υπολογιστούν σαφώς πιο εύκολα από τις υπολογιστικές διατάξεις, η παρούσα υποενότητα αποσκοπεί στην αποτύπωση βασικών εννοιών της

⁶⁸ Η παρούσα εικόνα αντλήθηκε από την πηγή [68].

Διαφορικής Γεωμετρίας που γεφυρώνουν ένα εννοιολογικό πλαίσιο της χρήσης αυτής στις Πολλαπλότητες αντιμετωπίζοντας αποτελεσματικά το αρχικό πρόβλημα.

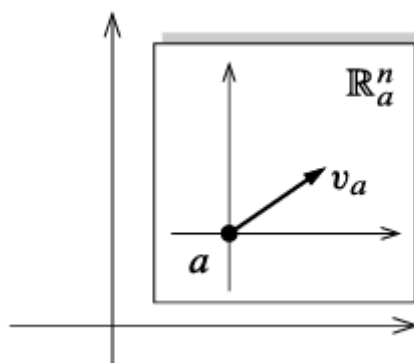
Γεωμετρικά Εφαπτομενικά διανύσματα και Εφαπτομενικοί Χώροι

Η ανάλυση ξεκινά με την υπενθύμιση ότι οι Πολλαπλότητες θεωρούνται τοπολογικοί Χώροι οι οποίοι τοπικά προσομοιάζουν τον Ευκλείδειο Χώρο. Όμως η αντίληψη των στοιχείων που περικλείει ο \mathbb{R}^n είναι διττός: Από την μια πλευρά μπορούν να θεωρηθούν ως *σημεία*, με τις συντεταγμένες να περιγράφουν το χωρικό τους «αποτύπωμα», κι από την άλλη μπορούν να θεωρηθούν ως *διανύσματα*. Λαμβάνοντας υπόψιν ότι τα τελευταία είναι μεγέθη που αποτυπώνουν ποιοτικά την κατεύθυνση αλλά και το μέτρο ενός φυσικού μεγέθους με το σχήμα ενός βέλους η αρχή του οποίου μπορεί να βρίσκεται σε οποιοδήποτε σημείο του \mathbb{R}^n , αυτό τα καθιστά αυτομάτως απηλλαγμένα από το χωρικό σημείο έναρξής τους επιβεβαιώνοντας την ανάγκη της εύρεσης ενός τρόπου αποτύπωσης των δεδομένων που να είναι ανεξάρτητος των χωρικών συντεταγμένων, όπως έχει αναφερθεί και στην υποενότητα *Χώρος και Αριθμοί*. Είναι εμφανές πως η υιοθέτηση των διανυσμάτων ως τα εργαλεία της ανάλυσης που θα ακολουθήσει αποτελεί μονόδρομο [68]!

Με τα παραπάνω υπόψιν, ακολουθεί ο ορισμός του *Εφαπτομενικού Διανύσματος* στον Ευκλείδειο Χώρο ως:

Ορισμός 48: Εφαπτομενικό Διάνυσμα

Δοσμένου ενός σημείου $a \in \mathbb{R}^n$, ως **Εφαπτομενικό Διάνυσμα** του \mathbb{R}^n στο σημείο a , θεωρείται ένα σύνολο $\{a\} \times \mathbb{R}^n = \{(a, u) : u \in \mathbb{R}^n\}$ και συμβολίζεται ως \mathbb{R}_a^n . Στην ίδια φιλοσοφία ένα **γεωμετρικό εφαπτομενικό διάνυσμα** στον \mathbb{R}^n είναι ένα στοιχείο του \mathbb{R}_a^n για κάποιο $a \in \mathbb{R}^n$ [68].



Εικόνα 72: Αποτύπωση του γεωμετρικού εφαπτομενικού διανύσματος⁶⁹.

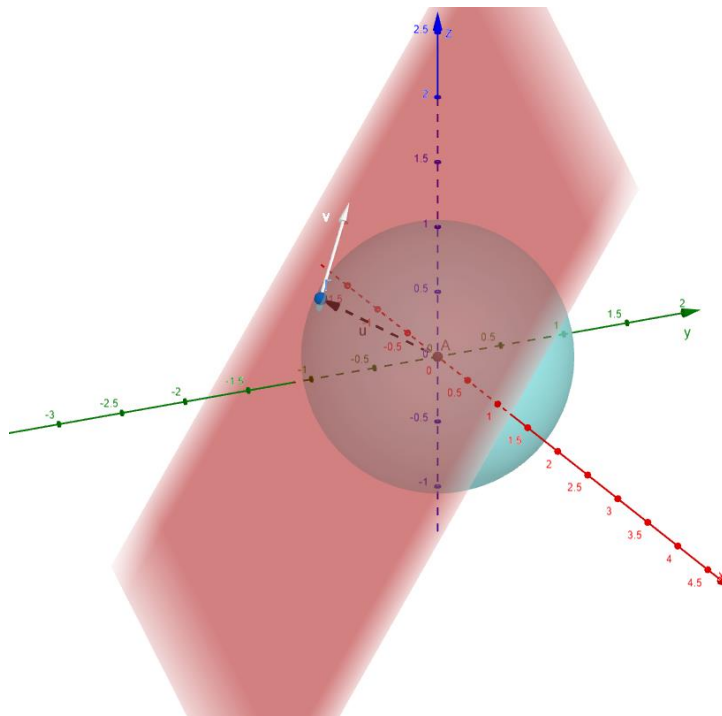
Γίνεται λοιπόν αντιληπτό ότι το σύνολο \mathbb{R}_a^n είναι ένας διανυσματικός χώρος πραγματικών αριθμών, ο οποίος είναι ακριβώς ο ίδιος όπως ο \mathbb{R}^n υπό τις φυσικές πράξεις:

⁶⁹ Η παρούσα εικόνα αντλήθηκε από την πηγή [68].

- $u_a + w_a = (u + w)_a$
- $c(u_a) = (cu)_a$

Χρησιμοποιώντας λοιπόν το Πολύπτυχο που συνιστά μια μοναδιαία σφαίρα n -διαστάσεων, S^{n-1} , κανείς θα μπορούσε να θεωρήσει ότι ο εφαπτομενικός χώρος που εισάγεται στο S από το σημείο a το οποίο εντοπίζεται πάνω σε αυτόν, $a \in S^{n-1}$ ως έναν συγκεκριμένο υποχώρο του \mathbb{R}_a^n , ο οποίος είναι ίδιος με τον \mathbb{R}^n και περιλαμβάνει διανύσματα που είναι ορθοκανονικά στο μοναδιαίο διάνυσμα που εκτείνεται από το κέντρο της σφαίρας και κινείται ακτινικά σε αυτήν, όπως μπορεί να εντοπιστεί στην *Εικόνα 73*. Ο παραπάνω ορισμός όμως δεν είναι ενιαίος καθώς δεν μπορεί να εξυπηρετήσει περιπτώσεις όπου η μορφολογία του ομαλού Πολύπτυχου είναι άγνωστη άρα και ο υποκείμενος περιβάλλοντας Ευκλείδειος Χώρος είναι *a priori* άγνωστος [68].

Η λύση σε αυτό το πρόβλημα δίνεται από τον ορισμό της παραγώγισης στο a . Με απόδειξη η οποία αφήνεται στον αναγνώστη για να επεξεργαστεί, αποτυπώνεται πως [68]:



Εικόνα 73: Απεικόνιση του μοναδιαίου ακτινικού διανύσματος, u (μαύρες διακεκομμένες γραμμές), το οποίο είναι ορθοκανονικό σε ένα τυχαίο διάνυσμα, v (λευκό χρώμα), που έχει ως αρχή το σημείο a και αυθαίρετη κατεύθυνση.

Ορισμός 49: Παραγώγιση σημείου σε Λείο Πολύπτυχο (Derivation at point a of a Smooth Manifold)

Με την υπόθεση ότι ως a ορίζεται ένα σημείο στον \mathbb{R}^n , μια απεικόνιση $w: C^\infty(\mathbb{R}^n) \rightarrow \mathbb{R}$ αποκαλείται παραγώγιση στο a αν είναι γραμμική στον \mathbb{R} και ικανοποιεί τον παρακάτω κανόνα αλυσίδας [68]:

$$w(fg) = f(a)wg + g(a)wf \quad \text{για όλα τα } f, g \in C^\infty(M) \quad (89)$$

Η έννοια της τοπικής παραγώγισης ή όπως εμφανίστηκε παραπάνω της παραγώγισης σημείου, αποτελεί έναν γεφυρωτικό ορισμό για τους *Εφαπτομενικούς Χώρους*. Συγκεκριμένα:

Ορισμός 50: Εφαπτομενικός Χώρος ενός Πολύπτυχου (Tangent Space of a Manifold)

Το σύνολο όλων των παραγώγων του $C^\infty(M)$ στο σημείο p ενός Πολύπτυχου, που αποτυπώνεται με την σημειογραφία T_pM , είναι ένας διανυσματικός Χώρος και αποκαλείται **Εφαπτομενικός Χώρος του M στο p** . Τα διανύσματα που πλαισιώνονται σε αυτόν, ονομάζονται **Εφαπτομενικά Διανύσματα** [68].

Η Εφαπτομενική Δέσμη (The Tangent Bundle)

Το τμηματικά διακριτό αντίγραφο του Χώρου \mathbb{R}^n που υπεισέρχεται από τον Εφαπτομενικό Χώρο σε κάθε σημείο που εξετάζεται στο Πολύπτυχο, αποτελεί ένα εξαιρετικά χρήσιμο εργαλείο καθώς επιτρέπει πλέον την χρήση των Ευκλείδειων τελεστών για την επιτέλεση όλων των τυπικών πράξεων συσχέτισης διανυσμάτων μεταξύ τους. Η έννοια της επαγόμενης τοπικότητας όμως είναι ιδιαίτερα περιοριστική καθώς εγκλωβίζει τα σημεία σε Χώρους οι οποίοι δεν επικοινωνούν, οδηγώντας έτσι σε αδυναμία πλαισίωσης σύγκρισης των εφαπτομενικών διανυσμάτων, άρα κατ' επέκτασιν και των σημείων πάνω στα διαφορετικά σημεία της εκάστοτε Πολλαπλότητας. Το πρόβλημα αυτό έρχεται να επιλύσει η *Εφαπτομενική Δέσμη*, η οποία καταφέρνει να συλλέξει όλους τους Εφαπτομενικούς Χώρους σε μια ενιαία γεωμετρική δομή.

Ορισμός 51: Εφαπτομενική Δέσμη (Tangent Bundle)

Δοσμένου ενός Λείου Πολύπτυχου M , ορίζεται πως η **Εφαπτομενική Δέσμη** είναι η ξένη ένωση (*disjoint union*) των εφαπτόμενων Χώρων σε όλα τα σημεία του M . Συμβολιζόμενη ως TM , αποτυπώνεται ως [68]:

$$TM = \coprod_{p \in M} T_pM \quad (90)$$

Μια ακόμη πολύ σημαντική ιδιότητα που προσδίδει η Εφαπτομενική Δέσμη είναι ότι είναι εφοδιασμένη με μια φυσική προβολική απεικόνιση (*projection map*) [68]:

$$\pi: TP \rightarrow M \quad (91)$$

Αυτή η προβολική απεικόνιση δίνει την δυνατότητα μεταφοράς κάθε διανύσματος που βρίσκεται στον *Εφαπτομενικό Χώρο* στο σημείο p της Πολλαπλότητας στο οποίο εφάπτεται, δηλαδή [68]:

$$\pi(\pi, u) = \pi \quad (92)$$

Ορισμός 52: Η Εφαπτομενική Δέσμη ως Λεία Πολύπτυχο

Για κάθε λεία n -διάστατη Πολλαπλότητα M , η Εφαπτομενική Δέσμη TM έχει μια φυσική τοπολογία και μια λεία δομή που την καθιστούν μια $2n$ -διάστατη Λεία Πολλαπλότητα. Σε σχέση με αυτή τη δομή, η προβολή $\pi: TM \rightarrow M$ είναι λεία [68].

Βιβλιογραφία – Αναφορές - Διαδικτυακές Πηγές

- [1] M. N. Oude Lohuis *et al.*, “Multisensory task demands temporally extend the causal requirement for visual cortex in perception,” *Nat. Commun.*, vol. 13, no. 1, Art. no. 1, May 2022, doi: 10.1038/s41467-022-30600-4.
- [2] M. A. McDaniel and G. O. Einstein, “Bizarre imagery as an effective memory aid: The importance of distinctiveness,” *J. Exp. Psychol. Learn. Mem. Cogn.*, vol. 12, no. 1, pp. 54–65, 1986, doi: 10.1037/0278-7393.12.1.54.
- [3] W. W. Patton, “Opening students’ eyes: Visual learning theory in the Socratic classroom,” *Law Psychol. Rev.*, vol. 15, pp. 1–18, 1991.
- [4] M. P. Verdi, J. T. Johnson, W. A. Stock, R. W. Kulhavy, and P. Whitman-Ahern, “Organized Spatial Displays and Texts,” *J. Exp. Educ.*, vol. 65, no. 4, pp. 303–317, Jul. 1997, doi: 10.1080/00220973.1997.10806606.
- [5] “Digital Image Processing.” Accessed: May 26, 2024. [Online]. Available: <https://www.imageprocessingplace.com/>
- [6] K. Ding, K. Ma, S. Wang, and E. P. Simoncelli, “Comparison of Full-Reference Image Quality Models for Optimization of Image Processing Systems,” *Int. J. Comput. Vis.*, vol. 129, no. 4, pp. 1258–1281, Apr. 2021, doi: 10.1007/s11263-020-01419-7.
- [7] M. Boccia, V. Sulpizio, F. Bencivenga, C. Guariglia, and G. Galati, “Neural representations underlying mental imagery as unveiled by representation similarity analysis,” *Brain Struct. Funct.*, vol. 226, no. 5, pp. 1511–1531, Jun. 2021, doi: 10.1007/s00429-021-02266-z.
- [8] C. J. Erkelens, “Multiple Photographs of a Perspective Scene Reveal the Principles of Picture Perception,” *Vision*, vol. 2, no. 3, Art. no. 3, Sep. 2018, doi: 10.3390/vision2030026.
- [9] J. Pearson, “The human imagination: the cognitive neuroscience of visual mental imagery,” *Nat. Rev. Neurosci.*, vol. 20, no. 10, pp. 624–634, Oct. 2019, doi: 10.1038/s41583-019-0202-9.
- [10] S. Fan, B. L. Koenig, Q. Zhao, and M. S. Kankanhalli, “A Deeper Look at Human Visual Perception of Images,” *SN Comput. Sci.*, vol. 1, no. 1, p. 58, Jan. 2020, doi: 10.1007/s42979-019-0061-5.
- [11] “Camera Sensors: What Are They and How Do They Work? | FUJIFILM Exposure Center – USA.” Accessed: May 27, 2024. [Online]. Available: <https://fujifilm-x.com/en-us/series/fundamentals-of-photography/camera-sensors-what-are-they-and-how-do-they-work/>
- [12] “Science Of Selfies: What Do Pixels Do?,” Teledyne Photometrics. Accessed: May 27, 2024. [Online]. Available: <https://www.photometrics.com/learn/imaging-topics/what-happens-when-light-hits-a-pixel>
- [13] “6.1: Images as Two-Dimensional Signals,” Engineering LibreTexts. Accessed: May 27, 2024. [Online]. Available: [https://eng.libretexts.org/Bookshelves/Mechanical_Engineering/Introduction_to_Autonomous_Robots_\(Correll\)/06%3A_Vision/6.01%3A_Images_as_Two-Dimensional_Signals](https://eng.libretexts.org/Bookshelves/Mechanical_Engineering/Introduction_to_Autonomous_Robots_(Correll)/06%3A_Vision/6.01%3A_Images_as_Two-Dimensional_Signals)
- [14] “Nyquist–Shannon sampling theorem,” *Wikipedia*. May 21, 2024. Accessed: May 25, 2024. [Online]. Available: https://en.wikipedia.org/w/index.php?title=Nyquist%E2%80%93Shannon_sampling_theorem&oldid=1224971221
- [15] “Σημειώσεις από το μάθημα ‘Ψηφιακή Επεξεργασία Σήματος.’”
- [16] “Matrix (mathematics),” *Wikipedia*. May 26, 2024. Accessed: May 27, 2024. [Online]. Available: [https://en.wikipedia.org/w/index.php?title=Matrix_\(mathematics\)&oldid=1225817652](https://en.wikipedia.org/w/index.php?title=Matrix_(mathematics)&oldid=1225817652)
- [17] “A Quick intro to Indexing in Python - Samarth Bhargav,” HackerEarth. Accessed: May 27, 2024. [Online]. Available: <https://www.hackerearth.com/practice/notes/samarthbhargav/a-quick-intro-to-indexing-in-python/>
- [18] “Matrix Indexing in MATLAB.” Accessed: May 27, 2024. [Online]. Available: <https://www.mathworks.com/company/technical-articles/matrix-indexing-in-matlab.html>
- [19] “Raster scan,” *Wikipedia*. Nov. 19, 2023. Accessed: May 27, 2024. [Online]. Available: https://en.wikipedia.org/w/index.php?title=Raster_scan&oldid=1185873890
- [20] “Row-Major and Column-Major Array Layouts - MATLAB & Simulink.” Accessed: May 27, 2024. [Online]. Available: <https://www.mathworks.com/help/coder/ug/what-are-column-major-and-row-major-representation-1.html>

- [21] "Body relative direction," *Wikipedia*. Apr. 17, 2024. Accessed: May 29, 2024. [Online]. Available: https://en.wikipedia.org/w/index.php?title=Body_relative_direction&oldid=1219402077
- [22] samir khanal, "Relationships between pixels(Neighbours and Connectivity)," Medium. Accessed: May 31, 2024. [Online]. Available: <https://samirkhanal35.medium.com/relationships-between-pixels-neighbours-and-connectivity-d38e473cd994>
- [23] saqib, "Basic Relationships Between Pixels | Digital Image Processing," Cuitutorial. Accessed: May 31, 2024. [Online]. Available: <https://cuitutorial.com/basic-relationships-between-pixels/>
- [24] Y.-J. Zhang, "Pixel Spatial Relationship," in *Handbook of Image Engineering*, Y.-J. Zhang, Ed., Singapore: Springer, 2021, pp. 353–384. doi: 10.1007/978-981-15-5873-3_9.
- [25] "Affine transformation - Wikipedia." Accessed: Jun. 02, 2024. [Online]. Available: https://en.wikipedia.org/wiki/Affine_transformation
- [26] E. W. Weisstein, "Affine Transformation." Accessed: Jun. 02, 2024. [Online]. Available: <https://mathworld.wolfram.com/>
- [27] D. G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, Nov. 2004, doi: 10.1023/B:VISI.0000029664.99615.94.
- [28] "Supervised learning," *Wikipedia*. Apr. 25, 2024. Accessed: Jun. 06, 2024. [Online]. Available: https://en.wikipedia.org/w/index.php?title=Supervised_learning&oldid=1220694111
- [29] H. Wickham, "Tidy Data," *J. Stat. Softw.*, vol. 59, pp. 1–23, Sep. 2014, doi: 10.18637/jss.v059.i10.
- [30] "Unsupervised learning," *Wikipedia*. May 25, 2024. Accessed: Jun. 06, 2024. [Online]. Available: https://en.wikipedia.org/w/index.php?title=Unsupervised_learning&oldid=1225667081
- [31] "Supervised vs Unsupervised Learning - Javatpoint." Accessed: Jun. 06, 2024. [Online]. Available: <https://www.javatpoint.com/difference-between-supervised-and-unsupervised-learning>
- [32] "Supervised vs. unsupervised learning," Google Cloud. Accessed: Jun. 06, 2024. [Online]. Available: <https://cloud.google.com/discover/supervised-vs-unsupervised-learning>
- [33] "What Is Supervised Learning? | IBM." Accessed: Jun. 06, 2024. [Online]. Available: <https://www.ibm.com/topics/supervised-learning>
- [34] B. E. Boser, I. M. Guyon, and V. N. Vapnik, "A training algorithm for optimal margin classifiers," in *Proceedings of the fifth annual workshop on Computational learning theory*, in COLT '92. New York, NY, USA: Association for Computing Machinery, Apr. 1992, pp. 144–152. doi: 10.1145/130385.130401.
- [35] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, Sep. 1995, doi: 10.1007/BF00994018.
- [36] "Mathematics for Machine Learning," Mathematics for Machine Learning. Accessed: Jun. 07, 2024. [Online]. Available: <https://mml-book.com/>
- [37] A. Ng, "CS229 Lecture Notes." Stanford University, Jun. 11, 2023. [Online]. Available: https://cs229.stanford.edu/main_notes.pdf
- [38] "Support vector machine," *Wikipedia*. May 24, 2024. Accessed: Jun. 11, 2024. [Online]. Available: https://en.wikipedia.org/w/index.php?title=Support_vector_machine&oldid=1225371653#Kernel_trick
- [39] "Kernel method," *Wikipedia*. Mar. 31, 2024. Accessed: Jun. 11, 2024. [Online]. Available: https://en.wikipedia.org/w/index.php?title=Kernel_method&oldid=1216509663
- [40] "Definite matrix," *Wikipedia*. Jun. 09, 2024. Accessed: Jun. 11, 2024. [Online]. Available: https://en.wikipedia.org/w/index.php?title=Definite_matrix&oldid=1228081917
- [41] *Pattern Recognition and Machine Learning*. Accessed: Jun. 15, 2024. [Online]. Available: <https://link.springer.com/book/9780387310732>
- [42] "k-means clustering - MATLAB kmeans." Accessed: Nov. 24, 2023. [Online]. Available: <https://www.mathworks.com/help/stats/kmeans.html>
- [43] X. Jin and J. Han, "K-Means Clustering," in *Encyclopedia of Machine Learning*, C. Sammut and G. I. Webb, Eds., Boston, MA: Springer US, 2010, pp. 563–564. doi: 10.1007/978-0-387-30164-8_425.
- [44] "k-means clustering," *Wikipedia*. Jun. 01, 2024. Accessed: Jun. 15, 2024. [Online]. Available: https://en.wikipedia.org/w/index.php?title=K-means_clustering&oldid=1226686614
- [45] J. Macqueen, "SOME METHODS FOR CLASSIFICATION AND ANALYSIS OF MULTIVARIATE OBSERVATIONS," *Multivar. Obs.*.
- [46] "Normal distribution," *Wikipedia*. Jun. 10, 2024. Accessed: Jun. 17, 2024. [Online]. Available: https://en.wikipedia.org/w/index.php?title=Normal_distribution&oldid=1228320060

- [47] “Convex combination - Wikiversity.” Accessed: Jun. 19, 2024. [Online]. Available: https://en.wikiversity.org/wiki/Convex_combination
- [48] L. MIRANDA, “Understanding softmax and the negative log-likelihood,” Lj Miranda. Accessed: Jun. 19, 2024. [Online]. Available: <https://lvmiranda921.github.io/notebook/2017/08/13/softmax-and-the-negative-log-likelihood/>
- [49] P. Liang and D. Klein, “Online EM for unsupervised models,” in *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics on - NAACL '09*, Boulder, Colorado: Association for Computational Linguistics, 2009, p. 611. doi: 10.3115/1620754.1620843.
- [50] giominas, “Answer to ‘Is a sample covariance matrix always symmetric and positive definite?’” Cross Validated. Accessed: Jun. 29, 2024. [Online]. Available: <https://stats.stackexchange.com/a/366908>
- [51] T. Rot, “Answer to ‘Why does the space of SPD matrices form a differentiable manifold?’” Mathematics Stack Exchange. Accessed: Jun. 29, 2024. [Online]. Available: <https://math.stackexchange.com/a/2773444>
- [52] E. N. Zois, S. Said, D. Tsourounis, and A. Alexandridis, “Subscripto multiplex: A Riemannian symmetric positive definite strategy for offline signature verification,” *Pattern Recognit. Lett.*, vol. 167, pp. 67–74, Mar. 2023, doi: 10.1016/j.patrec.2023.02.002.
- [53] Α. Γιαζιτζής, “Meta-heuristic optimization methods on SPD Manifolds and applications to computer vision,” *Μέθοδοι υπολογιστικής όρασης με χρήση μετα-ευρετικών αλγορίθμων βελτιστοποίησης στον χώρο πινάκων συνδιακύμανσης*, Jul. 2023, doi: 10.26265/polynoe-4664.
- [54] Α. Καλαϊτζάκης, “Τεχνικές εύρωστης εκμάθησης μετρικών σε χώρους πολλαπλότητας πινάκων συνδιακύμανσης για χρήση σε εφαρμογές ταξινόμησης προτύπων και μηχανικής μάθησης,” *Robust distance measure for similarity-based classification on the SPD manifold*, Mar. 2024, doi: 10.26265/polynoe-5951.
- [55] “Manifold,” *Wikipedia*. Jun. 28, 2024. Accessed: Jun. 30, 2024. [Online]. Available: <https://en.wikipedia.org/w/index.php?title=Manifold&oldid=1231519786>
- [56] J. Wilson, “The Definition of a Manifold and First Examples,” 2012.
- [57] GeometryForPhysicists, *What is a manifold?*, (Sep. 11, 2015). Accessed: Jun. 30, 2024. [Online Video]. Available: <https://www.youtube.com/watch?v=zljBARHTPZ4>
- [58] M. Spivak, *A Comprehensive Introduction to Differential Geometry VOLUME TWO, Second Edition*, 2nd edition. Berkeley: Publish or Perish, Incorporated, 1979.
- [59] E. W. Weisstein, “Manifold.” Accessed: Jun. 30, 2024. [Online]. Available: <https://mathworld.wolfram.com/>
- [60] J. W. Robbin and D. A. Salamon, “INTRODUCTION TO DIFFERENTIAL GEOMETRY”.
- [61] L. I. Nicolaescu, “Lectures on the Geometry of Manifolds”.
- [62] “Isomorphism,” *Wikipedia*. Jun. 20, 2024. Accessed: Jul. 01, 2024. [Online]. Available: <https://en.wikipedia.org/w/index.php?title=Isomorphism&oldid=1230049983>
- [63] “Homeomorphism,” *Wikipedia*. Apr. 11, 2024. Accessed: Jul. 01, 2024. [Online]. Available: <https://en.wikipedia.org/w/index.php?title=Homeomorphism&oldid=1218409201>
- [64] “Bijection,” *Wikipedia*. May 24, 2024. Accessed: Jul. 01, 2024. [Online]. Available: <https://en.wikipedia.org/w/index.php?title=Bijection&oldid=1225416520>
- [65] “Diffeomorphism,” *Wikipedia*. Feb. 23, 2024. Accessed: Jul. 01, 2024. [Online]. Available: <https://en.wikipedia.org/w/index.php?title=Diffeomorphism&oldid=1209782586>
- [66] H. Q. Minh and V. Murino, Eds., *Algorithmic Advances in Riemannian Geometry and Applications: For Machine Learning, Computer Vision, Statistics, and Optimization*. in *Advances in Computer Vision and Pattern Recognition*. Cham: Springer International Publishing, 2016. doi: 10.1007/978-3-319-45026-1.
- [67] N. Boumal, *An Introduction to Optimization on Smooth Manifolds*, 1st ed. Cambridge University Press, 2023. doi: 10.1017/9781009166164.
- [68] J. M. Lee, *Introduction to Smooth Manifolds*, vol. 218. in *Graduate Texts in Mathematics*, vol. 218. New York, NY: Springer, 2012. doi: 10.1007/978-1-4419-9982-5.
- [69] “Paracompact space,” *Wikipedia*. Jun. 16, 2024. Accessed: Jul. 02, 2024. [Online]. Available: https://en.wikipedia.org/w/index.php?title=Paracompact_space&oldid=1229420605
- [70] “Second-countable space,” *Wikipedia*. Nov. 24, 2023. Accessed: Jul. 02, 2024. [Online]. Available: https://en.wikipedia.org/w/index.php?title=Second-countable_space&oldid=1186675989

- [71] “Geometry of Manifolds | Mathematics,” MIT OpenCourseWare. Accessed: Jul. 02, 2024. [Online]. Available: <https://ocw.mit.edu/courses/18-965-geometry-of-manifolds-fall-2004/>
- [72] E. W. Weisstein, “Coordinate Chart.” Accessed: Jul. 03, 2024. [Online]. Available: <https://mathworld.wolfram.com/>
- [73] “Metric space,” *Wikipedia*. May 11, 2024. Accessed: Jul. 09, 2024. [Online]. Available: https://en.wikipedia.org/w/index.php?title=Metric_space&oldid=1223326328
- [74] “Metric tensor,” *Wikipedia*. Jul. 08, 2024. Accessed: Jul. 09, 2024. [Online]. Available: https://en.wikipedia.org/w/index.php?title=Metric_tensor&oldid=1233253786#Metric_in_a_vector_bundle
- [75] “Euclidean distance,” *Wikipedia*. Mar. 17, 2024. Accessed: Jul. 14, 2024. [Online]. Available: https://en.wikipedia.org/w/index.php?title=Euclidean_distance&oldid=1214273566
- [76] “Riemannian Manifolds: An Introduction to Curvature | SpringerLink.” Accessed: Jul. 15, 2024. [Online]. Available: <https://link.springer.com/book/10.1007/b98852>
- [77] “Levi-Civita connection,” *Wikipedia*. Jul. 03, 2024. Accessed: Jul. 15, 2024. [Online]. Available: https://en.wikipedia.org/w/index.php?title=Levi-Civita_connection&oldid=1232377520
- [78] “Parallel Transport on the Cone Manifold of SPD Matrices for Domain Adaptation | IEEE Journals & Magazine | IEEE Xplore.” Accessed: Jul. 17, 2024. [Online]. Available: <https://ieeexplore.ieee.org/document/8624413>
- [79] “Re-visiting Riemannian geometry of symmetric positive definite matrices for the analysis of functional connectivity - ScienceDirect.” Accessed: Jul. 18, 2024. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1053811920309496>
- [80] M. Moakher, “A Differential Geometric Approach to the Geometric Mean of Symmetric Positive-Definite Matrices,” *SIAM J. Matrix Anal. Appl.*, vol. 26, no. 3, pp. 735–747, Jan. 2005, doi: 10.1137/S0895479803436937.
- [81] X. Pennec, P. Fillard, and N. Ayache, “A Riemannian Framework for Tensor Computing,” *Int. J. Comput. Vis.*, vol. 66, no. 1, pp. 41–66, Jan. 2006, doi: 10.1007/s11263-005-3222-z.
- [82] F. Porikli, O. Tuzel, and P. Meer, “Designing a Boosted Classifier on Riemannian Manifolds,” in *Riemannian Computing in Computer Vision*, P. K. Turaga and A. Srivastava, Eds., Cham: Springer International Publishing, 2016, pp. 281–301. doi: 10.1007/978-3-319-22957-7_13.
- [83] Z. S. Harris, “Distributional Structure,” *WORD*, vol. 10, no. 2–3, pp. 146–162, Aug. 1954, doi: 10.1080/00437956.1954.11659520.
- [84] “Bag-of-words model,” *Wikipedia*. Jul. 10, 2024. Accessed: Jul. 19, 2024. [Online]. Available: https://en.wikipedia.org/w/index.php?title=Bag-of-words_model&oldid=1233710007
- [85] “The Distributional Hypothesis in NLP: Foundations, Applications, and Computational Methods,” GeeksforGeeks. Accessed: Jul. 19, 2024. [Online]. Available: <https://www.geeksforgeeks.org/the-distributional-hypothesis-in-nlp-foundations-applications-and-computational-methods/>
- [86] H. P. Luhn, “A Statistical Approach to Mechanized Encoding and Searching of Literary Information,” *IBM J. Res. Dev.*, vol. 1, no. 4, pp. 309–317, Jul. 1957, doi: 10.1147/rd.14.0309.
- [87] G. Salton, A. Wong, and C. S. Yang, “A vector space model for automatic indexing,” *Commun ACM*, vol. 18, no. 11, pp. 613–620, Aug. 1975, doi: 10.1145/361219.361220.
- [88] Sivic and Zisserman, “Video Google: a text retrieval approach to object matching in videos,” in *Proceedings Ninth IEEE International Conference on Computer Vision*, Jul. 2003, pp. 1470–1477 vol.2. doi: 10.1109/ICCV.2003.1238663.
- [89] D. K. McGrath, “Accuracy and Explainability in Artificial Intelligence: Unpacking the Terms,” 2021.
- [90] Y. Rimal, N. Sharma, and A. Alsadoon, “The accuracy of machine learning models relies on hyperparameter tuning: student result classification using random forest, randomized search, grid search, bayesian, genetic, and optuna algorithms,” *Multimed. Tools Appl.*, Feb. 2024, doi: 10.1007/s11042-024-18426-2.
- [91] A. Botchkarev, “Performance Metrics (Error Measures) in Machine Learning Regression, Forecasting and Prognostics: Properties and Typology,” *Interdiscip. J. Inf. Knowl. Manag.*, vol. 14, pp. 045–076, 2019, doi: 10.28945/4184.
- [92] G. Naidu, T. Zuva, and E. M. Sibanda, “A Review of Evaluation Metrics in Machine Learning Algorithms,” in *Artificial Intelligence Application in Networks and Systems*, R. Silhavy and P. Silhavy, Eds., Cham: Springer International Publishing, 2023, pp. 15–25. doi: 10.1007/978-3-031-35314-7_2.

- [93] S. Lazebnik, C. Schmid, and J. Ponce, "Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories," in *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Volume 2 (CVPR'06)*, New York, NY, USA: IEEE, 2006, pp. 2169–2178. doi: 10.1109/CVPR.2006.68.
- [94] K. Grauman and T. Darrell, "The pyramid match kernel: discriminative classification with sets of image features," in *Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1*, Jul. 2005, pp. 1458–1465 Vol. 2. doi: 10.1109/ICCV.2005.239.
- [95] F. Perronnin and C. Dance, "Fisher Kernels on Visual Vocabularies for Image Categorization," in *2007 IEEE Conference on Computer Vision and Pattern Recognition*, Minneapolis, MN, USA: IEEE, Jun. 2007, pp. 1–8. doi: 10.1109/CVPR.2007.383266.
- [96] "Fit Gaussian mixture model to data - MATLAB fitgmdist." Accessed: Jul. 25, 2024. [Online]. Available: <https://www.mathworks.com/help/stats/fitgmdist.html>
- [97] H. Jegou, M. Douze, C. Schmid, and P. Perez, "Aggregating local descriptors into a compact image representation," in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, San Francisco, CA, USA: IEEE, Jun. 2010, pp. 3304–3311. doi: 10.1109/CVPR.2010.5540039.
- [98] D. Maltoni, D. Maio, A. K. Jain, and S. Prabhakar, *Handbook of Fingerprint Recognition*. London: Springer, 2009. doi: 10.1007/978-1-84882-254-2.
- [99] M. Okawa, "Offline Signature Verification Based on Bag-of-VisualWords Model Using KAZE Features and Weighting Schemes," in *2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Las Vegas, NV, USA: IEEE, Jun. 2016, pp. 252–258. doi: 10.1109/CVPRW.2016.38.
- [100] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognit. Lett.*, vol. 27, no. 8, pp. 861–874, Jun. 2006, doi: 10.1016/j.patrec.2005.10.010.
- [101] M. Diaz, M. A. Ferrer, D. Impedovo, M. I. Malik, G. Pirlo, and R. Plamondon, "A Perspective Analysis of Handwritten Signature Technology," *ACM Comput. Surv.*, vol. 51, no. 6, pp. 1–39, Nov. 2019, doi: 10.1145/3274658.
- [102] S. Pal, S. Chanda, U. Pal, K. Franke, and M. Blumenstein, *Off-line signature verification using G-SURF*. 2012, p. 591. doi: 10.1109/ISDA.2012.6416603.
- [103] M. Houtinezhad and H. R. Ghaffari, "Off-line signature verification system using features linear mapping in the candidate points," *Multimed. Tools Appl.*, vol. 81, no. 17, pp. 24815–24847, Jul. 2022, doi: 10.1007/s11042-022-12499-7.
- [104] G. S. Eskander, R. Sabourin, and E. Granger, "Hybrid writer-independent–writer-dependent offline signature verification system," *IET Biom.*, vol. 2, no. 4, pp. 169–181, 2013, doi: 10.1049/iet-bmt.2013.0024.
- [105] V. L. F. Souza, A. L. I. Oliveira, R. M. O. Cruz, and R. Sabourin, "A white-box analysis on the writer-independent dichotomy transformation applied to offline handwritten signature verification," *Expert Syst. Appl.*, vol. 154, p. 113397, Sep. 2020, doi: 10.1016/j.eswa.2020.113397.
- [106] "Pedestrian Detection via Classification on Riemannian Manifolds | IEEE Journals & Magazine | IEEE Xplore." Accessed: Jul. 30, 2024. [Online]. Available: <https://ieeexplore.ieee.org/document/4479482>
- [107] W. Gao, Z. Ma, C. Xiong, and T. Gao, "Dimensionality reduction of SPD data based on Riemannian manifold tangent spaces and local affinity," *Appl. Intell.*, vol. 53, no. 2, pp. 1887–1911, Jan. 2023, doi: 10.1007/s10489-022-03177-0.
- [108] A. Giazitzis and E. N. Zois, "SigmML: Metric Meta-Learning for Writer Independent Offline Signature Verification in the Space of SPD Matrices," presented at the Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2024, pp. 6312–6322. Accessed: Jul. 30, 2024. [Online]. Available: https://openaccess.thecvf.com/content/WACV2024/html/Giazitzis_SigmML_Metric_Meta-Learning_for_Writer_Independent_Offline_Signature_Verification_in_WACV_2024_paper.html
- [109] A. A. Abdulhussien, M. F. Nasrudin, S. M. Darwish, and Z. A. Alyasseri, "Improving arabic signature authentication with quantum inspired evolutionary feature selection," *Multimed. Tools Appl.*, vol. 83, no. 28, pp. 71495–71524, Aug. 2024, doi: 10.1007/s11042-024-18198-9.
- [110] F. Luna, J. C. M. Romo, M. Mora-González, E. Martínez-Cano, and V. L. Rivas, "Handwritten Signature Verification Using Multi Objective Optimization with Genetic Algorithms in a Forensic Architecture," in *Logistics Management and Optimization through Hybrid Artificial Intelligence Systems*, IGI Global, 2012, pp. 141–180. doi: 10.4018/978-1-4666-0297-7.ch006.

- [111] A. A. Abdulhussien, M. F. Nasrudin, S. M. Darwish, and Z. A. A. Alyasseri, "A Genetic Algorithm Based One Class Support Vector Machine Model for Arabic Skilled Forgery Signature Verification," *J. Imaging*, vol. 9, no. 4, Art. no. 4, Apr. 2023, doi: 10.3390/jimaging9040079.