ΕΠΑΝΑΤΑΥΤΟΠΟΙΗΣΗ ΑΤΟΜΩΝ ΑΠΟ ΕΙΚΟΝΕΣ ΚΑΙ ΒΙΝΤΕΟ

Η
ΜΕΤΑΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ ΕΞΕΙΔΙΚΕΥΣΗΣ

Υποβάλλεται στην

ορισθείσα από την Γενική Συνέλευση Ειδικής Σύνθεσης
του Τμήματος Μηχανικών Η/Υ και Πληροφορικής
Εξεταστική Επιτροπή

από τον

Μπακάλη Βασίλειο

ως μέρος των Υποχρεώσεων

για τη λήψη

του

ΜΕΤΑΠΤΥΧΙΑΚΟΥ ΔΙΠΛΩΜΑΤΟΣ ΣΤΗΝ ΠΛΗΡΟΦΟΡΙΚΗ

ΜΕ ΕΞΕΙΔΙΚΕΥΣΗ ΣΤΙΣ ΤΕΧΝΟΛΟΓΙΕΣ-ΕΦΑΡΜΟΓΕΣ

Φεβρουάριος 2018

ΑΦΙΕΡΩΣΗ

Στην οικογένεια μου

ΠΕΡΙΕΧΟΜΕΝΑ

ΕΥΡΕΤΗΡΙΟ ΠΙΝΑΚΩΝ

ΕΥΡΕΤΗΡΙΟ ΣΧΗΜΑΤΩΝ

ΠΕΡΙΛΗΨΗ

Μπακάλης Βασίλειος, MSc στη Πληροφορική.
Τμήμα Μηχανικών Η/Υ & Πληροφορικής, Πανεπιστήμιο Ιωαννίνων
Φεβρουάριος 2018
Επαναταυτοποίηση ατόμων από εικόνες και βίντεο.
Επιβλέποντας: Χ. Νίκου, Αναπληρωτής Καθηγητής.

Η επαναταυτοποίηση ατόμου (Person Re-identification) αποτελεί μια από τις βασικές διεργασίες ενός συστήματος αυτοματοποιημένης βιντεο-επιτήρησης (video surveillance) και έχει καταστεί αντικείμενο έντονης έρευνας τα τελευταία χρόνια. Έχοντας ως δεδομένο μια εικόνα/βίντεο ενός ατόμου που έχει ληφθεί από μια κάμερα παρακολούθησης, η επαναταυτοποίηση είναι η διαδικασία αναγνώρισης του ίδιου ατόμου από εικόνες/βίντεο που έχουν ληφθεί από διαφορετικές κάμερες ή οπτικές. Στις πρώτες μέρες αλγόριθμοι μικρής κλίμακας αναπτύχθηκαν και αξιολογήθηκαν κατά κύριο λόγο. Τα τελευταία χρόνια ωστόσο διαπιστώθηκε η εμφάνιση σειρών δεδομένων μεγάλης κλίμακας και συστημάτων βαθιάς μάθησης (deep learning) που χρησιμοποιούν μεγάλους όγκους δεδομένων και εξάγουν καλύτερα αποτελέσματα.
Στα πλαίσια αυτής της εργασίας παρουσιάζεται πολύπλευρα το πρόβλημα της επαναταυτοποίησης ατόμου (Person Re-identification) και γίνεται μια προσπάθεια ανάδειξης των προβλημάτων, των περιορισμών και των προκλήσεων που εμφανίζονται κατά τη διαδικασία αυτή. Τα ανοικτά ζητήματα και οι προκλήσεις του προβλήματος τονίζονται με μια συζήτηση σχετικά με τις δυνατότητες που υπάρχουν, καθώς και παρουσιάζονται μέσα από τη σύγκριση των κύριων αλγορίθμων που υπάρχουν.
Αυτό το έγγραφο: 1) εισάγει τον αναγνώστη στο ιστορικό του person re-identification και τη σχέση του με την ταξινόμηση της εικόνας και την ανάκτηση των περιπτώσεων. 2) παρουσιάζει έρευνες οι οποίες έχουν γίνει για την επιλογή των συστημάτων και των μεθόδων μεγάλης κλίμακας τόσο σε αναπαράσταση ταυτότητας εικόνας όσο και βίντεο. 3) συγκρίνει βασικούς αλγορίθμους οι οποίοι χρησιμοποιούνται τόσο για την ανάλυση δεδομένων από εικόνες (στατικές) αλλά και βίντεο 4) επισημαίνει ορισμένα σημαντικά αλλά ανεπαρκώς ανεπτυγμένα ζητήματα τα οποία τίθενται προς συζήτηση για το θέμα αυτό.

EXTENDED ABSTRACT IN ENGLISH

Mpakalis Vasileios, MSc in Computer Science.
Department of Computer Science & Engineering, University of Ioannina.
February 2018
Image -Video based Person Re-identification
Advisor: C. Nikou, Associate Professor.

Person Re-identification is one of the main processes of an automated video surveillance system and has been the subject of intense research in recent years. Given a person's video / video taken from a surveillance camera, re-authentication is the process of identifying the same person from images / videos taken from different cameras or optics. In the early days, small-scale algorithms were developed and evaluated primarily. In recent years, however, the emergence of large-scale datasets and deep learning systems that use large volumes of data and produce better results have been identified.

In this work, the Person Re-identification problem is presented on a multifaceted basis and an effort is made to highlight the problems, constraints and challenges that arise during this process. The open issues and the challenges of the problem are highlighted by a discussion of the possibilities that exist, as well as presented through a comparison of the main algorithms that exist.

This document: 1) introduces the reader to the historical person's re-identification and its relation to the image classification and the recovery of the cases. 2) presents research that has been made to select large-scale systems and methods in both image and video representation. 3) compares key algorithms that are used both to analyze data from pictures (static) but also video 4) points to some important but underdeveloped issues that are being raised for discussion on this issue.

ΚΕΦΑΛΑΙΟ 1. ΕΙΣΑΓΩΓΗ

1.1. Στόχοι

Η αναγνώριση και η επαναταυτοποίηση του ατόμου (person re-identification) αποτελεί βασικό καθήκον της αυτοματοποιημένης βιντεοεπιτήρησης (video surveillance) και υπήρξε ένας τομέας έντονης έρευνας τα τελευταία χρόνια. Δεδομένης μιας εικόνας / βίντεο ενός ατόμου που έχει ληφθεί από μια κάμερα, η επαναταυτοποίηση είναι η διαδικασία αναγνώρισης του ατόμου από εικόνες / βίντεο που έχουν ληφθεί από διαφορετική κάμερα και είναι απαραίτητη για την καθιέρωση συνεπούς επισήμανσης για την αποκατάσταση των αποσυνδεδεμένων ή χαμένων κομματιών.

Εκτός από την παρακολούθηση το φαινόμενο αυτό έχει εφαρμογές στη ρομποτική, τα πολυμέσα, την ιατροδικαστική κ.α. Επιπλέον, η επαναταυτοποίηση είναι ένα δύσκολο πρόβλημα λόγω της οπτικής ασάφειας και της χωρο-χρονικής αβεβαιότητας στην εμφάνιση ενός ατόμου σε διάφορες κάμερες. Αυτές οι δυσκολίες συχνά συνδυάζονται με εικόνες χαμηλής ανάλυσης ή κακής ποιότητας βίντεο που δεν βοηθούν στην επαναταυτοποίηση. Οι χωρικές ή χρονικές συνθήκες για περιορισμό του προβλήματος είναι δύσκολο να καταγραφούν, ωστόσο, το πρόβλημα έχει λάβει σημαντική προσοχή από την ερευνητική κοινότητα λόγω της ευρείας εφαρμογής και χρησιμότητάς του φαινομένου.

Σε αυτή την εργασία, εξετάζεται το πρόβλημα της επαναταυτοποίησης του ατόμου και συζητούνται οι τρέχουσες λύσεις - αλγόριθμοι οι οποίοι βρίσκονται στη βιβλιογραφία. Τα ανοικτά ζητήματα και οι προκλήσεις του προβλήματος τονίζονται και γίνεται σύγκριση των αλγορίθμων προκειμένου να παρουσιαστεί ένα αποτέλεσμα που να σχετίζεται με την εύρεση του πιο αποδοτικού αλγορίθμου. Στην ερευνητική κοινότητα καταγράφονται κυρίως αλγόριθμοι με τη χρήση περιγραφέων, οι οποίοι χρησιμοποιούν την εκπαίδευση του συστήματος σαν όπλο για την επίτευξη των στόχων του φαινομένου. Τα τελευταία χρόνια παρατηρείται η εμφάνιση σειρών δεδομένων μεγάλης κλίμακας και συστημάτων βαθιάς μάθησης που χρησιμοποιούν μεγάλους όγκους δεδομένων.

Η επαναταυτοποίηση του ατόμου αποτελεί πρόβλημα εύρεσης ενός ατόμου από μια συλλογή δεδομένων εικόνων. Αυτό είναι ένα δύσκολο πρόβλημα λόγω των πολλαπλών χαρακτηριστικών τα οποία υπάρχουν κατά την επεξεργασία εικόνας, όπως είναι οι παραλλαγές φωτισμού, θέσης ή θέασης και ο θόρυβος. Δύο βασικά προβλήματα είναι κρίσιμα για την ταυτοποίηση του ατόμου, η ταύτιση των χαρακτηριστικών και η μετρική μάθησης. Μια αποτελεσματική απεικόνιση χαρακτηριστικών θα πρέπει να είναι ισχυρή προκειμένου να έχει σωστά αποτελέσματα και να μην επηρεάζεται από διάφορα χαρακτηριστικά όπως είναι π.χ οι αλλαγές του φωτισμού. Πολλές προσπάθειες έχουν καταβληθεί για την αντιμετώπιση των δύο κατευθύνσεων του προβλήματος αυτού.

Σκοπός της μελέτης αυτής είναι η παρουσίαση των αλγορίθμων που χρησιμοποιούνται στη βιβλιογραφία για την αντιμετώπιση του προβλήματος αυτού καθώς και η λεπτομερής καταγραφή των χαρακτηριστικών και των αποτελεσμάτων τους. Στο σημείο αυτό σημειώνεται ότι ο τρόπος σύγκρισης των αλγορίθμων γίνεται με τη χρήση δεδομένων εικόνων (galleries datasets) που έχουν δημιουργηθεί για το σκοπό αυτό, μέσα από τα οποία αξιολογείται και η αποτελεσματικότητα του κάθε αλγορίθμου.

1.2. Δομή της Εργασίας

Το κείμενο κατανέμεται σε εφτά κεφάλαια, καθένα από τα οποία αναλύει και παρουσιάζει ένα διαφορετικό θέμα το οποίο διερευνήθηκε κατά την υλοποίηση της παρούσας εργασίας. Στα κεφάλαια της εργασίας γίνεται αναλυτική παρουσίαση, τόσο του προβλήματος το οποίο πραγματεύεται η εργασία, όσο και των μεθόδων οι οποίες έχουν ακολουθηθεί και παρουσιάζονται στη βιβλιογραφία.

Στο παρόν κεφάλαιο γίνεται μια εισαγωγή στο πρόβλημα της επαναταυτοποίησης του ατόμου, καθώς επίσης παρουσιάζεται και μια σύνοψη της δομής της εργασίας.

Στο δεύτερο και τρίτο κεφάλαιο, παρουσιάζονται οι έννοιες της μηχανικής όρασης, της επεξεργασίας εικόνας και γενικότερα οι θεωρητικές δομές οι οποίες είναι απαραίτητες προκειμένου να μπορέσει ο αναγνώστης να κατανοήσει τα όσα θα παρουσιαστούν στην συνέχεια της εργασίας.

Στο τέταρτο κεφάλαιο γίνεται μια πιο εκτενής ανάλυση του φαινομένου, παρουσιάζοντας τη χρήση, το σκοπό και τη λογική στην οποία βασίζεται. Είναι σημαντικό για να μπορέσει ο αναγνώστης κατανοήσει τους τρόπους διαχωρισμού και εφαρμογής των μεθόδων που παρουσιάζονται στην έρευνα αυτή, καθώς και να καταλάβει τους λόγους για τους οποίους το φαινόμενο αποτελεί ένα σημαντικό κομμάτι της έρευνας.

Το πέμπτο κεφάλαιο, αποτελεί το συνδετικό κρίκο της εργασίας, καθώς παρουσιάζονται μερικά από τα πιο σημαντικά σύνολα δεδομένων εικόνων που χρησιμοποιούνται ευρέως στη βιβλιογραφία για το σκοπό αυτό, καθώς επίσης αναλύονται οι τεχνικές και αλγόριθμοι.

Συνέχεια του πέμπτου κεφαλαίου αποτελεί το έκτο, το οποίο συγκρίνει τις βασικότερες τεχνικές και μεθοδολογίες που έχουν παρουσιαστεί στο πέμπτο κεφάλαιο και παρουσιάζει τα αποτελέσματα σύγκρισης μεταξύ αυτών καθώς και το σχολιασμό τους για τη καλύτερη κατανόηση των αποτελεσμάτων.

Τέλος, το τελευταίο κεφάλαιο συνοψίζει όλα τα στοιχεία που έχουν παρουσιαστεί στην υπόλοιπη εργασία και καταλήγει σε συμπεράσματα και μελλοντικές προσεγγίσεις.

ΚΕΦΑΛΑΙΟ 2. ΜΗΧΑΝΙΚΗ ΟΡΑΣΗ ΚΑΙ ΕΠΕΞΕΡΓΑΣΙΑ ΕΙΚΟΝΑΣ

Στο ακόλουθο κεφάλαιο εισάγονται βασικές έννοιες έτσι ώστε να ενταχθεί ο αναγνώστης σε αυτές και να αντιληφτεί το υπόβαθρο της παρούσας εργασίας. Πιο συγκεκριμένα γίνεται μια εισαγωγή στην έννοια της «Μηχανικής Όρασης» και αναλύονται οι χρήσεις και οι τομείς στους οποίους εφαρμόζεται καθώς επίσης παρουσιάζονται και βασικά θέματα της επεξεργασίας εικόνας.

2.1. Εισαγωγή στη Μηχανική Όραση

Η αίσθηση της όρασης θεωρείται μία από τις σημαντικές αισθήσεις του ανθρώπου καθώς του δίνει τη δυνατότητα να αντιλαμβάνεται και να κατανοεί τον κόσμο γύρω του. Τη δυνατότητα αυτή επιδιώκει να αντιγράψει η μηχανική ή υπολογιστική όραση. Η μηχανική όραση αποτελεί ένα επιστημονικό πεδίο της τεχνητής νοημοσύνης που προσπαθεί αλγοριθμικά να αναπαραστήσει μία τρισδιάστατη σκηνή μέσω της επεξεργασίας των δισδιάστατων εικόνων που λαμβάνει σαν είσοδο από οπτικούς αισθητήρες όπως είναι οι κάμερες, με κόστος την απώλεια μεγάλου όγκου χρήσιμης πληροφορίας.

Πιο συγκεκριμένα, η όραση υπολογιστή έχει ως στόχο τη μετατροπή δεδομένων από μία κάμερα σταθερής εικόνας/βίντεο σε μία απόφαση ή σε κάποιου άλλου είδους απεικόνιση. Όλες αυτές οι μετατροπές γίνονται με σκοπό την επίτευξη ενός συγκεκριμένου στόχου μιας διαδικασίας. Η απόφαση που τελικά θα ληφθεί μπορεί να έχει να κάνει π.χ με την ύπαρξη ενός ατόμου ή το πλήθος των καρκινικών κυττάρων σε μια εικόνα. Μια νέα είδους απεικόνιση μπορεί να σημαίνει π.χ τη μετατροπή μιας έγχρωμης εικόνας σε εικόνα στην κλίμακα του γκρι ή την αφαίρεση του υποβάθρου από αυτή.

Η μηχανική όραση επιδιώκει να εφαρμόσει θεωρίες και μοντέλα για την κατασκευή συστημάτων με δυνατότητα όρασης. Παραδείγματα στα οποία τέτοια συστήματα βρίσκουν εφαρμογή είναι τα εξής:

1. Έλεγχος διαδικασιών (π.χ ένα βιομηχανικό ρομπότ ή ένα αυτόνομο όχημα)
2. Ανίχνευση συμβάντων (π.χ. οπτική επιτήρηση).
3. Οργάνωση πληροφοριών (π.χ. ευρετηριοποίηση βάσεων δεδομένων).
4. Εξομοίωση αντικειμένων και περιβάλλοντος (π.χ. βιομηχανική επιθεώρηση, ιατρική ανάλυση εικόνας ή τοπογραφική εξομοίωση).
5. Αλληλεπίδραση χρηστών με υπολογιστικά συστήματα (π.χ. μια συσκευή επικοινωνίας ανθρώπου / μηχανής).

Η όραση μεταξύ ανθρώπου και μηχανής είναι μία τελείως διαφορετική διαδικασιά για το καθένα. Αυτό συμβαίνει γιατί εκ πρώτης όψεως είναι εύκολο για το ανθρώπινο μάτι να αντιληφθεί και να αναγνωτίσει πού βρίσκεται ένα αντικείμενο μέσα σε μία εικόνα. Όμως πίσω από αυτή τη φαινομενικά εύκολη δουλειά κρύβεται μια πολύπλοκη διαδικασία. Ο ανθρώπινος εγκέφαλος διαιρεί το οπτικό σήμα σε πολλά κανάλια μέσω των οποίων ρέουν διαφορετικά είδη πληροφοριών. Ο εγκέφαλος διαθέτει ένα σύστημα που αναγνωρίζει με τρόπο τέτοιο που εξαρτάται από τον εκάστοτε στόχο, τα σημαντικά μέρη της εικόνας ενώ παράλληλα καταστέλλει την εξέταση άλλων περιοχών. Κατά τη διάρκεια αυτής της διαδικασίας γίνεται μαζική ανατροφοδότηση του οπτικού ρεύματος, η οποία μέχρι τώρα έχει γίνει ελάχιστα κατανοητή. Ταυτόχρονα υπάρχουν ευρείες συνειρμικές είσοδοι από μυϊκούς αισθητήρες ελέγχου, αλλά και από άλλες αισθήσεις που επιτρέπουν στον εγκέφαλο να χρησιμοποιήσεις όλες τις γνώσεις και συσχετίσεις που έχουν προκύψει κατά τη διάρκεια ζωής του ανθρώπου. Η ανατροφοδότηση κάνει κύκλους μέσα στον εγκέφαλο, πηγαίνοντας πίσω σε όλα τα στάδια επεξεργασίας, συμπεριλαμβανομένων των αισθητήρων, στην περίπτωση αυτή των ματιών, τα οποία ελέγχουν μηχανικά το φωτισμό μέσω της ίριδας και συντονίζουν τη λήψη σήματος στην επιφάνεια του αμφιβληστροειδή.

Αντιθέτως τα συστήματα που βασίζονται στη μηχανική όραση είναι ακόμη αρκετά «αφελή». Ως επί το πλείστων, σε αυτά δεν υπάρχει καμία ενσωματωμένη αναγνώριση προτύπων, κανένας αυτόματος έλεγχος εστίασης και διαφράγματος κάμερας και καμία συσχέτιση λόγω ετών εμπειρίας. Από την άλλη μεριά, το μόνο που «βλέπει» ένα σύστημα όρασης είναι ένα πλέγμα αριθμών. Κάθε μεμονωμένος αριθμός από αυτό το πλέγμα περιέχει και ένα επιπλέον στοιχείο θορύβου και έτσι του δίνει ελάχιστες πληροφορίες για τα στοιχεία μιας εικόνας. Στόχος της μηχανικής όρασης είναι να μετατραπεί αυτό το θορυβώδες σύνολο αριθμών στην αντίληψη της πραγματικής εικόνας και των στοιχίων που την αποτελούν.

Ο τομέας της μηχανικής όρασης έχει άμεση σχέση με την ψηφιακή επεξεργασία εικόνας και πολλές φορές συγχέεται για το πού σταματάει ο ένας και πού αρχίζει ο άλλος. Για να γίνει πιο κατανοητό αυτό το όριο μεταξύ αυτών ορίζονται τα παρακάτω τρία επίπεδα επεξεργασίας εικόνας, τα οποία ακολουθούνται από τα συστήματα τεχνητής όρασης για να επιτύχουν το στόχο τους:

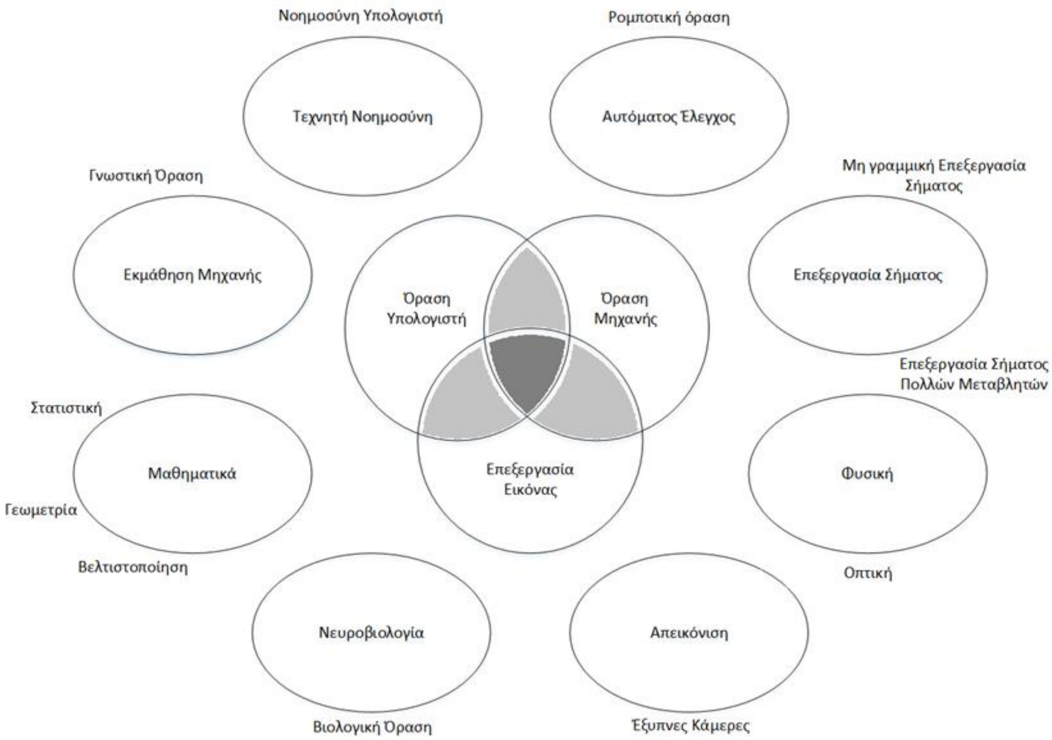
• Χαμηλό επίπεδο: Σε έναν αλγόριθμο χαμηλού επιπέδου, τόσο η είσοδος όσο και η έξοδος είναι εικόνες (π.χ αποθορυβοποίηση, απαλοιφή θολώματος κλπ).

• Ενδιάμεσο επίπεδο: Η είσοδος, σε ένα αλγόριθμο επεξεργασίας ενδιάμεσου επιπέδου, είναι εικόνα αλλά η έξοδος είναι κάποια χαρακτηριστικά ανωτέρου επιπέδου (π.χ. ακμές, περιγράμματα περιοχών κλπ.).

• Υψηλό επίπεδο: Η είσοδος σε ένα αλγόριθμο επεξεργασίας υψηλού επιπέδου είναι εικόνα αλλά η έξοδος είναι πληροφορία (π.χ. αντικείμενα που περιέχονται στην εικόνα).

Η μηχανική όραση αποτελεί συμπληρωματικό και συνδετικό κρίκο μεταξύ πολλών περιοχών. Πολλές φορές αντιμετωπίζεται σαν τμήμα της τεχνητής νοημοσύνης καθώς οι δύο αυτοί τομείς μοιράζονται θέματα που αφορούν την αναγνώριση προτύπων και τις τεχνικές εκμάθησης. Όσον αφορά π.χ το πεδίο της ρομποτικής μπορεί να προσφέρει μέσω ολοκληρωμένων συστημάτων ακριβείς πληροφορίες περιβάλλοντος για την αυτόνομη πλοήγηση ρομπότ μέσα στο χώρο.

Όσον αφορά την επεξεργασία σήματος, η τεχνητή όραση μπορεί να επεκτείνει τις μεθόδους της και από επεξεργασία μονοδιάστατων σημάτων να γίνεται πλέον επεξεργασία τόσο δισδιάστατων όσο και πολυδιάστατων σημάτων. Επιπλέον η τεχνητή όραση μπορούμε να πούμε πως έχει συνάφεια και με τα γραφικά υπολογιστών καθώς και οι δύο ασχολούνται με τη δημιουργία τρισδιάστατων μοντέλων από δεδομένα εικόνων. Τέλος, πρέπει να σημειωθεί ότι τόσο οι μέθοδοι όσο και οι λειτουργίες αυτού του πολυδιάστατου τομέα μελετώνται υπό την οπτική γωνία των αντίστοιχων μαθηματικών πεδίων.



Σχήμα 2 1 Συσχέτιση μηχανικής όρασης με άλλα επιστημονικά πεδία

2.2. Μηχανική Όραση και Εκπαίδευση

Μηχανική μάθηση (machine learning) [18] αναπτύχθηκε από τη μελέτη της αναγνώρισης προτύπων και της υπολογιστικής θεωρίας μάθησης με τις αρχές της τεχνητής νοημοσύνης. Το 1959, ο Arthur Samuel [19] δίνει τον ορισμό της μηχανικής μάθησης ως «Πεδίο μελέτης που δίνει στους υπολογιστές την ικανότητα να μαθαίνουν, χωρίς να έχουν ρητά προγραμματιστεί». Στον ορισμό σημειώνεται πως η μηχανική μάθηση διερευνά τη μελέτη και την κατασκευή αλγορίθμων που μπορούν να μαθαίνουν μέσα από την ανατροφοδότηση δεδομένων και να κάνουν προβλέψεις σχετικά με αυτά, προκειμένου να χρησιμοποιηθούν σε διαφορετικούς τομείς. Οι αλγόριθμοι αυτοί λειτουργούν κατασκευάζοντας μοντέλα τα οποία προκύπτουν από πειραματικά δεδομένα, προκειμένου να κάνουν προβλέψεις βασιζόμενες στα δεδομένα ή να εξάγουν αποφάσεις και να εκφράσουν ένα αποτέλεσμα.

Η μηχανική μάθηση είναι στενά συνδεδεμένη και συχνά συγχέεται με την υπολογιστική στατιστική, που επίσης επικεντρώνεται στην πρόβλεψη μέσω της χρήσης των υπολογιστών. Ο τομέας αυτός συνδυάζεται παράλληλα με τη μαθηματική βελτιστοποίηση, η οποία παρέχει μεθόδους, θεωρία και τομείς εφαρμογών. Η μηχανική μάθηση εφαρμόζεται σε μια σειρά από υπολογιστικές εργασίες, όπου τόσο ο σχεδιασμός όσο και ο ρητός προγραμματισμός των αλγορίθμων είναι ανέφικτος.
Όσον αφορά την ανάλυση δεδομένων, η μηχανική μάθηση αποτελεί μέθοδο που χρησιμοποιείται για την επινόηση πολύπλοκων μοντέλων και αλγορίθμων που έχουν ως σκοπό την πρόβλεψη και την αναγνώριση στοιχείων. Τα αναλυτικά μοντέλα τα οποία συντάσσονται επιτρέπουν στους ερευνητές να παράγουν αξιόπιστες αποφάσεις και αποτελέσματα και να αναδείξουν αλληλοσυσχετίσεις μέσω της μάθησης από σχέσεις, τάσεις στα δεδομένα, αλλά και μέσα από τον διαχωρισμό των δεδομένων τα οποία είναι δύσκολο να κατηγοριοποιηθούν.

2.3. Βασικά Στοιχεία Επεξεργασίας Εικόνας

Σε αυτή την ενότητα θα περιγραφούν βασικές έννοιες της επεξεργασίας εικόνας καθώς και βασικά στοιχεία τα οποία χρησιμοποιούνται για το σκοπό αυτό. Σημαντικό ρόλο για τη ποιότητα μιας εικόνας παίζει η θέση της κάμερας και η βαθμονόμησή της (camera calibration).

2.3.1. Βαθμονόμηση Κάμερας (Camera Calibration)

Πριν από μερικά χρόνια οι πληροφορίες που συλλέγονταν από συστήματα παρακολούθησης βίντεο ήταν κυρίως μόνο ποσοτικές αλλά πρόσφατα η ιδέα της χρήσης αυτών των συστημάτων ως σημαντικών πηγών ποιοτικής πληροφόρησης έγινε ακόμη πιο διάχυτη. Στη πραγματικότητα, με κατάλληλη διαμόρφωση του συστήματος παρακολούθησης, είναι δυνατό να προστεθούν ποιοτικές πληροφορίες όπως είναι η ποσότητα μετρικών που μπορούν να χρησιμοποιηθούν για τη μέτρηση αντικειμένων στη σκηνή ή για τον υπολογισμό αποστάσεων ή ακόμη και για την αυτόματη αναγνώριση προσώπων. Για να ληφθούν ποιοτικές πληροφορίες από τις εικόνες είναι απαραίτητο να βαθμονομηθεί και να ευθυγραμμιστεί η κάμερα η οποία θα χρησιμοποιηθεί (camera calibration). Η βαθμονόμηση της κάμερας είναι η διαδικασία εκτίμησης όλων των παραμέτρων του μοντέλου και μπορεί να χωριστεί σε δύο κύρια βήματα: την εκτίμηση των παραμέτρων που ορίζει το γραμμικό μοντέλο και την εκτίμηση των παραμέτρων μη γραμμικής παραμόρφωσης.

Το ουσιαστικό στοιχείο της διαδικασίας αυτής είναι η εύρεση των πραγματικών παραμέτρων της κάμερας που έλαβε τις φωτογραφίες. Ορισμένες από αυτές τις παραμέτρους είναι το εστιακό μήκος, το μέγεθος της μορφής, το κύριο σημείο και η παραμόρφωση του φακού.

Στην επιστημονική βιβλιογραφία υπάρχουν διαφορετικές προσεγγίσεις για το σκοπό αυτό με βάση διαφορετικές τεχνικές [77] [78]. Μία από τις πιο διαδεδομένες τεχνικές είναι η μέθοδος του Zhang. Η τεχνική αυτή απαιτεί μόνο τη παρατήρηση ενός σχεδίου από την κάμερα, το οποίο πρέπει να ληφθεί τουλάχιστον σε τρεις διαφορετικούς προσανατολισμούς (Σχήμα 2-2). Είτε η κάμερα, είτε το επίπεδο σχέδιο μπορεί να μετακινηθεί ελεύθερα και η κίνηση δε χρειάζεται να είναι γνωστή. Η βαθμονόμηση της κάμερας είναι πολύ απλή και η διαδικασία είναι εύκολη και γρήγορη.

Σχήμα 2-2 Παράδειγμα χρήσης για την βαθμονόμηση της κάμερας. [79]

Η τεχνική αυτή επιτρέπει την εκτίμηση των εσωτερικών παραμέτρων (που είναι κοινές σε όλες τις εικόνες που έχουν αποκτηθεί) και των εξωτερικών παραμέτρων (που είναι διαφορετικές για κάθε εικόνα). Επίσης μπορεί να πραγματοποιηθεί λαμβάνοντας υπόψη ένα μοντέλο γραμμικής κάμερας. Προκειμένου να ληφθούν υπόψη οι μη γραμμικές παράμετροι που περιγράφουν την παραμόρφωση αυτή η τεχνική μπορεί να ενισχυθεί με έναν μη γραμμικό αλγόριθμο ελαχιστοποίησης. Ειδικότερα, η μέθοδος του Zhang λύνει ένα απλό γραμμικό πρόβλημα που εκτιμά τις παραμέτρους του μοντέλου της κάμερας και τα αποτελέσματα αυτής μπορούν να χρησιμοποιηθούν ως πρώτη λύση για την επίλυση ενός προβλήματος μη γραμμικών ελαχιστοποιήσεων που βελτιώνει τα αποτελέσματα λαμβάνοντας υπόψη και τις παραμέτρους παραμόρφωσης. Για την επίλυση του μη γραμμικού προβλήματος, ο Zhang προτείνει μια τεχνική που βασίζεται στο κριτήριο της μέγιστης πιθανότητας, αλλά είναι δυνατόν να χρησιμοποιηθούν όλοι οι διαφορετικοί αλγόριθμοι μη γραμμικών ελαχιστοποιήσεων στη βιβλιογραφία. [79]

2.3.2. Κατάτμηση Εικόνας (Image Segmentation)

Ο όρος κατάτμηση εικόνας αναφέρεται στη διαδικασία διαχωρισμού μιας ψηφιακής εικόνας σε πολλαπλά τμήματα (segments). Είναι μια τεχνική που χρησιμοποιείται ευρέως για την ανίχνευση κίνησης, την εύρεση διαφορών σε δύο πανομοιότυπες εικόνες ή ακόμη και τον υπολογισμό απόστασης αντικειμένου από σημείο.

Πιο συγκεκριμένα, ο διαχωρισμός των εικόνων βασίζεται στην ανάθεση μιας ετικέτας σε κάθε εικονοστοιχείο (pixel) σύμφωνα με τα χαρακτηριστικά του. Εικονοστοιχεία με όμοια οπτικά χαρακτηριστικά, όπως είναι το χρώμα, η ένταση, η υφή, εισάγονται στην ίδια κατηγορία και έχουν κοινή ετικέτα. Ο διαχωρισμός αυτός και η χρήση των ετικετών έχει ως στόχο την πιο εύκολη κατηγοριοποίηση των pixels. Η κατηγοριοποίηση αυτή βοηθά στην περαιτέρω χρήση αλγορίθμων για την εξαγωγή αποτελεσμάτων του πεδίου που πρέπει να διερευνηθεί. Ακόμη, η εικόνα χωρίζεται σε διαφορετικά επίπεδα (layers) τα οποία αφορούν το προσκήνιο και το παρασκήνιο της εικόνας. Τα παραπάνω επίπεδα των εικόνων παρουσιάζονται στην παρούσα ενότητα. Επιπλέον, σημειώνεται ότι ο διαχωρισμός των επιπέδων στοχεύει στην εξαγωγή των αντίστοιχων χαρακτηριστικών.

Αφαίρεση Φόντου (Background subtraction)

Η αφαίρεση φόντου είναι μια μέθοδος ανίχνευσης αντικειμένων που βρίσκονται στο προσκήνιο μιας σκηνής. Πιο συγκεκριμένα, η λογική πίσω από τη μέθοδο αυτή σχετίζεται με το διαχωρισμό των εικόνων σε διάφορα επίπεδα (layers). Κάθε επίπεδο φέρει δικά του χαρακτηριστικά και εξετάζεται ως οντότητα. Η αρχική εικόνα που λαμβάνει ο αλγόριθμος αποτελεί το παρασκήνιο (background) της εικόνας. Οι διαφοροποιήσεις που υπάρχουν από επίπεδο σε επίπεδο εξετάζονται ως μεταβολές. Για την επιλογή της κατάλληλης μεθόδου πρέπει να ληφθούν υπόψιν κάποιοι παράγοντες. Οι παράγοντες αυτοί σχετίζονται με:

- Τον φωτισμό.
- Τις αλλαγές της γεωμετρίας του φόντου.
- Το αντικείμενο του φόντου.
- Την κάμερα η οποία θα χρησιμοποιηθεί για να ληφθεί η εικόνα.

Συνηθέστερη μέθοδος η οποία χρησιμοποιείται από την βιβλιοθήκη OpenCV είναι η μέθοδος αφαίρεσης φόντου με χρήση του προσαρμοζόμενου μοντέλου μίξης κανονικών κατανομών μέσω της συνάρτησης BackgroundSubtractorMOG2(). Οι παράμετροι της συνάρτησης αυτής είναι οι: history, varThreshold και bShadowDetection. Με την παράμετρο history δηλώνεται ο αριθμός των καρέ τα οποία θα συγκρατεί - επεξεργάζεται ο αλγόριθμος. Δηλαδή τα δείγματα t με τα οποία θα υπολογιστούν οι κατανομές. Η επόμενη παράμετρος, varThreshold, αφορά το τετράγωνο της απόστασης Mahalanobis. Τέλος, η παράμετρος bShadowDetection είναι τύπου boolean και παίρνει την τιμή 1 αν εφαρμοστεί η μέθοδος εντοπισμού σκιάς, ή την τιμή 0 αν δεν είναι επιθυμητό κάτι τέτοιο. Ωστόσο, πρέπει να σημειωθεί ότι ο αριθμός των παραμέτρων οι οποίες χρησιμοποιούνται ποικίλει ανάλογα με την έκδοση της βιβλιοθήκης.

Σχήμα 2-3 Αποτέλεσμα χρήσης τεχνικής αφαίρεσης Background. [74]

Οπτική Ροή (Optical flow)

Η οπτική ροή (optical flow) είναι το μοτίβο της φαινομενικής κίνησης των αντικειμένων της εικόνας μεταξύ δύο διαδοχικών καρέ που προκαλείται από την κίνηση του αντικειμένου ή κάμερας. Είναι ένα δύο διαστάσεων (2D) διανυσματικό πεδίο όπου κάθε φορέας μετατόπισης δείχνει τη μετατόπιση των σημείων αναφορικά με το πρώτο και δεύτερο καρέ.

Η μέθοδος της οπτικής ροής μπορεί να χρησιμοποιηθεί για ποικίλες εφαρμογές όπως η ανίχνευση κίνησης, η ανίχνευση ενός κινούμενου αντικειμένου, ακόμη και η ανίχνευση εμποδίου. Η λογική χρήσης της μεθόδου αυτής είναι κοινή σε όποια εφαρμογή κι αν προσαρμοστεί. Για την υλοποίηση του αλγορίθμου πρέπει να ληφθούν υπόψη οι παρακάτω υποθέσεις:

1. Η ένταση του κάθε pixel ενός αντικειμένου δεν αλλάζει μεταξύ διαδοχικών πλαισίων.
2. Τα γειτονικά pixels έχουν παρόμοια κίνηση. [75]

Σχήμα 2-4 Οπτική ροή σε παράδειγμα ανίχνευσης περαστικών. [76]

2.3.3. Περιγραφείς Χρώματος (Color descriptors)

Η σημασία του χρώματος των εικόνων και πιο συγκεκριμένα το χρώμα των εικονοστοιχείων (pixels) αποτελεί σημαντικό παράγοντα κατά την επεξεργασία και λαμβάνεται υπόψη για την εξαγωγή των χαρακτηριστικών. Για να αξιοποιηθούν πλήρως αυτές οι πληροφορίες προτείνονται δύο ξεχωριστές κατηγορίες:

- τα ιστογράμματα χρώματος (color histograms)
- τα σημασιολογικά ονόματα χρωμάτων (Semantic Color Descriptors)

Ιστογράμματα Χρώματος (Color Histograms)

Μια εικόνα αποτελείται από πολλαπλά κανάλια καθένα από τα οποία παρέχει διάφορες πληροφορίες όπως είναι π.χ το χρώμα, το βάθος καθώς και ο θόρυβος που μπορεί να εμπεριέχεται. Ο αριθμός των καναλιών τα οποία φέρει μια εικόνα δεν είναι σταθερός, ωστόσο κάθε εικόνα έχει τουλάχιστον ένα βασικό κανάλι. Το ιστόγραμμα χρώματος (color histograms) αποτελεί αναπαράσταση της κατανομής των χρωμάτων σε ένα συγκεκριμένο κανάλι. Τα αρχικά κανάλια RGB (red, green, blue) είναι επιρρεπή σε αλλαγές φωτισμού, με αποτέλεσμα, να χρησιμοποιούνται κυρίως κανάλια HSV (hue, saturation, lightness) και Lab για να διατηρήσουν τις ομαλές μεταβάσεις χρωμάτων σε πολύπλοκες συνθήκες φωτισμού και αποσπασματικές προβολές της κάμερας.

Έστω ότι σε ένα δεδομένο χώρο χρώματος, το χρωματικό ιστόγραμμα της λωρίδας της εικόνας I που περιέχει N εικονοστοιχεία αντιπροσωπεύεται ως H(I) = [h_1, h_2,…,h_n], όπου h_i = N, το οποίο είναι ο αριθμός ενός εικονοστοιχείου στη λωρίδα εικόνας που ανήκει στο i-ο χρωματικό κώδικα. Για την αναπαράσταση του h_i χρησιμοποιείται ο παρακάτω τύπος:

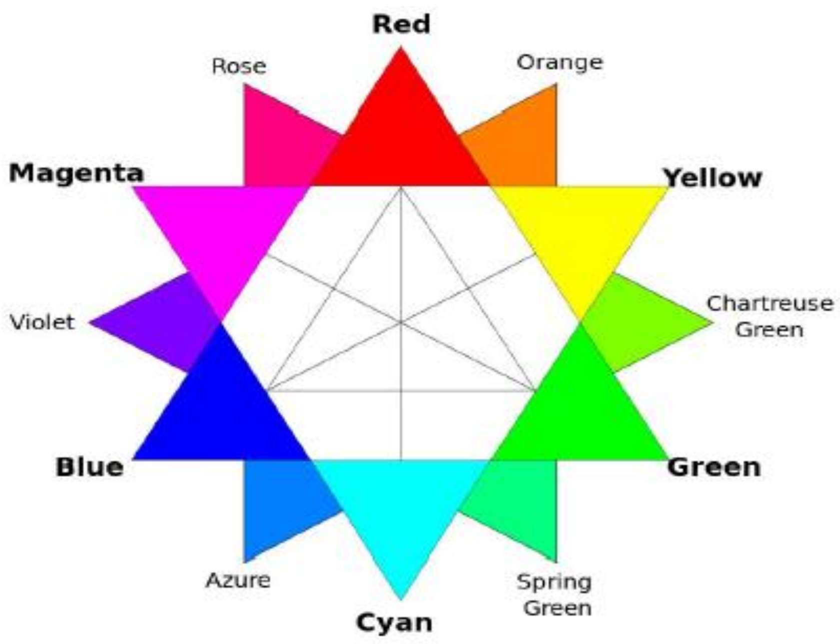
$$h_i = \sum_{j=1}^{N} P_{i|j} \tag{2.1}$$

Για να υπολογιστούν τα ιστογράμματα H(I) σε μια λωρίδα εικόνας, πρώτα μετατρέπεται κάθε εικονοστοιχείο από RGB σε HSV και Lab, κατόπιν γίνεται κβάντηση σε 16-αδικό, ξεχωριστά για κάθε κανάλι, στοιχείο το οποίο οδηγεί σε χαρακτηριστικά 144 διαστάσεων.

Σημασιολογικά Ονόματα Χρωμάτων (Semantic Color Descriptors)

Αντί των ιστογραμμάτων προστίθεται η πληροφορία που αφορά τις τιμές pixel μεταξύ των φωτογραφιών του ίδιου προσώπου οι οποίες μπορεί να δημιουργούν μικρά σύνολα. Παρατηρείται ότι όταν συγκρίνεται ένα ζευγάρι εικόνων χρησιμοποιούνται κυρίως χρωματικοί κώδικες για να ταξινομηθούν τα εικονοστοιχεία αυτής με βάση μια συγκεκριμένη σειρά χρωμάτων και να αγνοηθεί έτσι η μικρή διαφορά που μπορεί να υπάρχει. Για παράδειγμα, έστω ότι υπάρχουν δύο εικονοστοιχεία A (255,0,0) και B (220,0,0), έχουν και τα δύο κόκκινο χρώμα στο μάτι του ανθρώπου. Ωστόσο, οι τιμές RGB έχουν παρουσιάζουν μεγάλο κενό και τα σημασιολογικά ονόματα χρωμάτων θεωρούνται πιο ανθεκτικά και λιγότερο ευαίσθητα στον θόρυβο, ώστε να μπορέσει να ληφθεί η πληροφορία σωστά.

Για την ανάκτηση και ταυτοποίηση των ονομάτων, είναι απαραίτητη η λειτουργία χαρτογράφησης των τιμών που ανήκουν στην κατηγορία HSV σε ονόματα χρωμάτων. Επιλέγεται η μορφοποίηση HSV αντί για τιμές RGB, επειδή τα κανάλια στη μορφοποίηση αυτή είναι πολύ πιο πιθανό να παραμείνουν συνεχόμενα (αριθμοί με συνεχόμενη ροή) όταν αλλάζουν οι συνθήκες. Η εικόνα που ακολουθεί δείχνει την κατανομή των φωτεινών χρωμάτων σε χρώμα-χώρο HSV.



Σχήμα 2-5 Τροχός σήμανσης χρώματος σε χώρο HSV.

2.3.4. Θόρυβος

Ο θόρυβος μιας εικόνας αποτελεί τυχαία διακύμανση της φωτεινότητας ή των πληροφοριών χρώματος και είναι συνήθως μια πτυχή του ηλεκτρονικού θορύβου. Μπορεί να παραχθεί από τον αισθητήρα ή το κύκλωμα ενός αισθητήρα ή μιας ψηφιακής κάμερας. Ο θόρυβος είναι ένα ανεπιθύμητο υποπροϊόν στη λήψη εικόνας που αποκρύπτει τις επιθυμητές πληροφορίες. Αποτελεί σημαντική αλλοίωση στην επεξεργασία εικόνας, καθώς εάν εμφανιστεί σε μεγάλες στιβάδες μπορεί να οδηγήσει σε λανθασμένο αποτέλεσμα αναγνώρισης - ανίχνευσης. Προκειμένου να μειωθεί ο θόρυβος αρκετοί αλγόριθμοι χρησιμοποιούν φίλτρα και άλλες τεχνικές προκειμένου να μπορέσουν να μειώσουν το θόρυβο και να ανακτήσουν - ανιχνεύσουν το επιθυμητό αντικείμενο το οποίο υπάρχει σε μία εικόνα.

2.4. Τομείς Ανίχνευσης και Αναγνώρισης

Στην παράγραφο αυτή σημειώνονται τα σημαντικά στοιχεία της ανίχνευσης και της αναγνώρισης καθώς είναι σημαντικό προτού προχωρήσουμε στην ανάπτυξη της εργασίας και της εξήγησης των αλγορίθμων, να μπορέσει ο αναγνώστης να κατανοήσει τις διαφορές ανάμεσα στις βασικές έννοιες της ανίχνευσης και της αναγνώρισης στοιχείων και αντικειμένων.

2.4.1. Ανίχνευση - Αναγνώριση Αντικειμένου (Object Detection - Recognition)

Η ανίχνευση προτύπων σε μια εικόνα είναι μια απαραίτητη διαδικασία η οποία προσφέρει την εξαγωγή σημαντικών πληροφοριών και στοιχείων προκειμένου να ληφθεί μια απόφαση. Προκειμένου να γίνει ανίχνευση ενός στοιχείου πρέπει να γίνει επεξεργασία μιας εικόνας και εφαρμογή ενός αλγορίθμου αναγνώρισης (π.χ. ένας ταξινομητής εικόνας) ο οποίος θα λαμβάνει ως είσοδο μια επιλεγμένη εικόνα και θα εξάγει το στοιχείο το οποίο έχει ανιχνευτεί.

Για να μπορέσει ένας αλγόριθμος να εξάγει τα αποτελέσματα της ανίχνευσης ή αναγνώρισης απαιτείται η εκπαίδευση του αλγορίθμου προκειμένου να μάθει τις διαφορές μεταξύ των διαφορετικών τάξεων. Ο αλγόριθμος μπορεί μόνο να ανιχνεύσει αντικείμενα / τάξεις που έχει μάθει και όχι αντικείμενα τα οποία υπάρχουν απλά σε κάθε εικόνα. Για να γίνει επιτυχής αναγνώριση ενός αντικειμένου θα πρέπει να γίνει κατάλληλη εξαγωγή των χαρακτηριστικών και χρήση των περιγραφέων εικόνας, ώστε να μπορέσει η εφαρμογή να διακρίνει καλύτερα το εκάστοτε αντικείμενο και να το αναγνωρίσει. Επίσης, θα πρέπει κατά την εκπαίδευση του συστήματος να γίνει σωστή χρήση των ετικετών σήμανσης των αντικειμένων, μια διαδικασία η οποία παίζει σημαντικό ρόλο στη φάση της εκπαίδευσης και βοηθά προκειμένου να γίνει σωστά η διαδικασία της αναγνώρισης.

Η διαδικασία η οποία ακολουθείται κάθε φορά στην αναγνώριση προτύπων είναι η ακόλουθη:

1. Επιλογή τύπου δεδομένων (π.χ εικόνες, μουσικά αρχεία κλπ).
2. Δημιουργία συνόλου εκπαίδευσης
3. Αξιολόγηση του συνόλου εκπαίδευσης
4. Επιλογή κατάλληλων χαρακτηριστικών τα οποία θα εξαχθούν για ανάλυση.
5. Εισαγωγή διανύσματος ετικέτας (label) για κάθε στοιχείο του συνόλου δεδομένων
6. Μετατροπή κάθε αντικείμενου σε διάνυσμα χαρακτηριστικών.
7. Εφαρμογή αλγορίθμου εκπαίδευσης
8. Αξιολόγηση της διαδικασίας.
9. Εισαγωγή στοιχείου για αναγνώριση - κατηγοριοποίηση.

Η διαδικασία η οποία περιγράφεται παραπάνω ακολουθείται κάθε φορά που πρέπει να γίνει μια αναγνώριση προτύπου. Το σημαντικότερο κομμάτι αποτελούν τα δεδομένα της εκπαίδευσης καθώς θα πρέπει να απεικονίζουν με σαφήνεια το στοιχείο το οποίο θέλει ο χρήστης να αναγνωρίσει. Οι αλγόριθμοι αναφορικά με την εκπαίδευση των δεδομένων και την εξαγωγή των χαρακτηριστικών είναι αρκετοί. Ο κάθε αλγόριθμος που εφαρμόζεται πετυχαίνει διαφορετική απόδοση για διαφορετικά δεδομένα συνόλου εκπαίδευσης που χρησιμοποιούνται κάθε φορά.
Για την απόδοση του αλγορίθμου αναγνώρισης για κάθε ομάδα εικόνων που θα χρησιμοποιηθούν είναι αναγκαίο να υπάρχει ένας έλεγχος από μεθόσους που να πιστοποιούν αν ένα αντικείμενο ανήκει στο επιθυμητό σύνολο ή όχι (ground truth). Με βάση τη παραπάνω γνώση ένας αλγόριθμος για την εκπαίδεσυη του εκάστοτε αλγορίθμου ταξινομητή (classifier) μπορεί να πετύχει πολύ καλύτερα αποτελέσματα.
Για τη διαδικασία της δημιουργίας του συνόλου ground truth υπάρχουν τρεις βασικές προσεγγίσεις:

• 1^{η} προσέγγιση: η διαδικασία αυτή δημιουργείται από συστήματα που παρέχουν ακρίβεια μεγαλύτερη τουλάχιστον μιας τάξης μεγέθους από τη μέθοδο της αξιολόγησης.
• 2^{η} προσέγγιση: δημιουργία δεδομένων ground truth με χρήση γραφικών. Η προσέγγιση αυτή δε προτιμάται τόσο συχνά για οικονομικούς κυρίω λόγους αλλά και από άποψη χρόνου. Επίσης τα αποτέλεσμα της δεν είναι τόσο αξιόπιστα.
• 3^{η} προσέγγιση: χαρακτηρισμός των δεδομένων με τη χρήση του ανθρώπινου παράγοντα. Η παραπάνω διαδικασία είναι αρκετά χρονοβόρα

2.4.2. Ανίχνευση Πεζών (Pedestrian Detection)

Η ανίχνευση πεζών (Pedestrian Detection) [28] αποτελεί σημαντική διαδικασία σε οποιοδήποτε σύστημα παρακολούθησης βίντεο, καθώς παρέχει τις θεμελιώδεις πληροφορίες για τη σημασιολογική κατανόηση της εικόνας. Έχει προφανή επέκταση π.χ στις εφαρμογές της αυτοκινητοβιομηχανίας λόγω της δυνατότητας βελτίωσης των συστημάτων ασφαλείας. Παρά τις προκλήσεις, η ανίχνευση πεζών εξακολουθεί να παραμένει ένας σημαντικός και ενεργός ερευνητικός χώρος τα τελευταία χρόνια. Οι προσεγγίσεις που υπάρχουν μέχρι στιγμής είναι οι εξής:

- Ολιστική ανίχνευση (Holistic detection)

Οι ανιχνευτές εκπαιδεύονται για την αναζήτηση πεζών σε εικόνες και βίντεο με τη σάρωση ολόκληρου του πλαισίου (frame). Ο ανιχνευτής σημειώνει κάποια «απειλή» εάν τα χαρακτηριστικά της εικόνας μέσα στο τοπικό παράθυρο αναζήτησης πληρούν ορισμένα κριτήρια. Ορισμένες μέθοδοι χρησιμοποιούν παγκόσμια (global) χαρακτηριστικά, ενώ άλλοι χρησιμοποιούν τοπικά χαρακτηριστικά, όπως τα ιστογράμματα προσανατολισμένων κλίσεων (HOG). Το μειονέκτημα αυτής της προσέγγισης είναι ότι η απόδοση μπορεί εύκολα να επηρεαστεί από τις διαφορές συνθήκες του περιβάλλοντος και κυρίως του φωτισμού.

- Ανίχνευση βάσει μέρους (Part-based detection)

Οι πεζοί μοντελοποιούνται ως συλλογές τμημάτων (collections of parts) στην περίπτωση αυτή. Τα υποθετικά τμήματα αρχικά δημιουργούνται με την εκμάθηση τοπικών χαρακτηριστικών, τα οποία περιλαμβάνουν τα στοιχεία γωνίας (edgelet) από μία εικόνα και τα χαρακτηριστικά προσανατολισμού. Στη συνέχεια συνδυάζονται για να σχηματίσουν την καλύτερη συναρμολόγηση υφιστάμενων «υποθετικών πεζών». Παρόλο που η προσέγγιση αυτή είναι ελκυστική, η ανίχνευση βάσει μέρους είναι δύσκολη διαδικασία. Η εφαρμογή αυτής της προσέγγισης ακολουθεί μια τυποποιημένη διαδικασία για την επεξεργασία των δεδομένων εικόνας και την εκτέλεση ταξινόμησης σε όλες τις πιθανές θέσεις και τέλος τη δημιουργία του τελικού συνόλου πλαισίων οριοθέτησης (b-boxes).

- Ανίχνευση βασισμένη σε διαγράμματα (Patch-based detection)

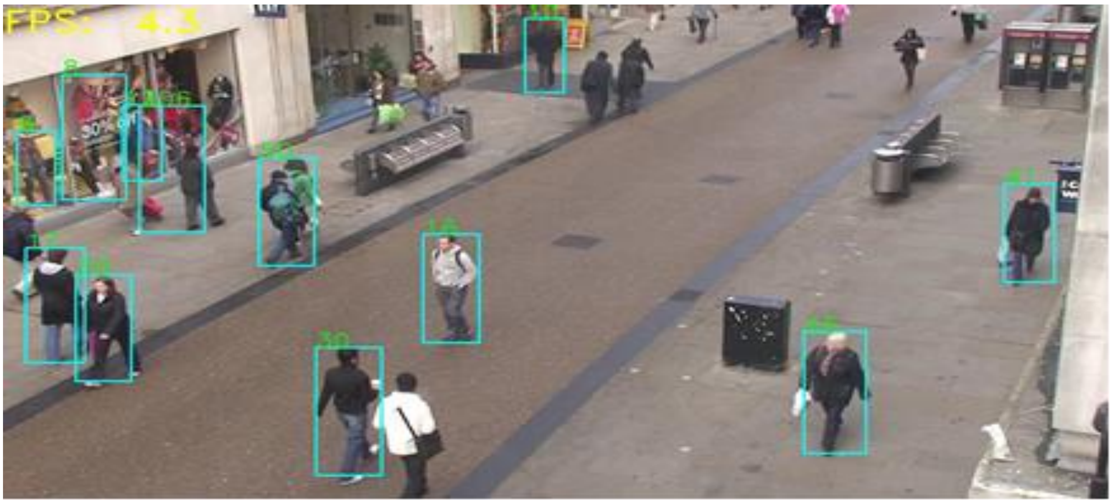
Η διαδικασία αυτή συνδυάζει τόσο την ανίχνευση όσο και την κατάτμηση. Κατά τη διαδικασία της ανίχνευσης χρησιμοποιούνται τα εξαγόμενα τοπικά χαρακτηριστικά για να ταιριάζουν με τις καταχωρήσεις του βιβλίου κωδίκων και ατόμων και σε κάθε επανάληψη του αλγορίθμου εκπέμπει μία ψήφο για τις υποθέσεις των πεζών. Τα τελικά αποτελέσματα ανίχνευσης μπορούν να ληφθούν με περαιτέρω εξέλιξη αυτών των υποθέσεων. Το πλεονέκτημα αυτής της προσέγγισης είναι ο μικρός αριθμός εικόνων που απαιτούνται για την εκπαίδευση του μοντέλου.

• Ανίχνευση με βάση την κίνηση (Motion-based detection)

Όταν οι συνθήκες επιτρέπουν (σταθερή κάμερα, συνθήκες σταθερού φωτισμού κ.λπ.), η αφαίρεση υποβάθρου μπορεί να βοηθήσει στην ανίχνευση πεζών. Η αφαίρεση του φόντου ταξινομεί τα εικονοστοιχεία των ροών βίντεο είτε ως φόντο, όπου δεν εντοπίζεται καμία κίνηση, είτε στο προσκήνιο, όπου ανιχνεύεται κίνηση. Αυτή η διαδικασία υπογραμμίζει τις σιλουέτες (τα συνδεδεμένα στοιχεία στο προσκήνιο) κάθε κινούμενου στοιχείου της σκηνής, συμπεριλαμβανομένων των ανθρώπων. Δεδομένου ότι οι μέθοδοι που εξετάζουν το σύνολο της σιλουέτας και εκτελούν μια ενιαία ταξινόμηση είναι γενικά πολύ ευαίσθητοι σε ατέλειες σχήματος (shape defects), έχει προταθεί μια μερική μέθοδος που χωρίζει τις σιλουέτες σε ένα σύνολο μικρότερων περιοχών για να μειωθεί η επίδραση των ατελειών.

• Ανίχνευση με χρήση πολλαπλών καμερών (Detection using multiple cameras)

Η τεχνική αυτή χωρίζει το έδαφος σε μη-επικαλυπτόμενα πλέγματα παράγοντας ένα χάρτη με μια εκτίμηση πιθανότητας να καταλαμβάνεται από ένα άτομο. Λαμβάνοντας τα συγχρονισμένα ρεύματα βίντεο από διαφορετικές γωνίες, η μέθοδος αυτή μπορεί να συνδυάσει αποτελεσματικά την ανίχνευση ατόμων παρά τις σημαντικές αποκλείσεις και τις αλλαγές φωτισμού



Σχήμα 2-6 Αποτελέσματα λειτουργίας ανίχνευσης πεζών.

ΚΕΦΑΛΑΙΟ 3. ΔΙΑΔΙΚΑΣΙΕΣ ΑΝΑΓΝΩΡΙΣΗΣ

3.1. Διαδικασία Αναγνώρισης

Η αναγνώριση ενός αντικειμένου και συγκεκριμένα ενός ατόμου σε μια στατική εικόνα ή βίντεο αποτελεί στοιχείο κομβικής σημασίας. Τα βήματα τα οποία απαιτούνται για την διαδιακασία αυτή είναι τα ακόλουθα (Σχήμα 3-1):

- Λήψη βίντεο και εικόνων
- Ανίχνευση ανθρώπου στην εικόνα
- Ανίχνευση ροής του αντικειμένου το οποίο κινείται
- Εξαγωγή χαρακτηριστικών του αντικειμένου το οποίο έχει ανιχνευθεί
- Σύγκριση χαρακτηριστικών και παραγωγή αποτελέσματος αναγνώρισης.

Σχήμα 3-1 Βήματα διαδικασίας αναγνώρισης ατόμου.

Προκειμένου να γίνει σωστά η αναγνώριση θα πρέπει να έχει εκπαιδευτεί ο αλγόριθμος με βάση ένα δείγμα εικόνων (gallery ή dataset), προκειμένου να μπορέσει να ανιχνεύσει το άτομο και στη συνέχεια να το ταυτοποιήσει με βάση το σύνολο εικόνων που είχε εκπαιδευτεί αρχικά. Στις ακόλουθες ενότητες παρουσιάζονται βασικά στοιχεία και αλγόριθμοι οι οποίοι χρησιμοποιούνται κατά τη διαδικασία των παραπάνω βημάτων.

3.2. Ανίχνευση / Εξαγωγή Χαρακτηριστικών (Feature Detection / Extraction)

Στην παρούσα ενότητα θα περιγραφούν οι τρόποι εξαγωγής των στοιχείων ενδιαφέροντος τα οποία υπάρχουν σε μια εικόνα. Δεν υπάρχει καθολικός ή ακριβής ορισμός του τι αποτελεί χαρακτηριστικό και ο ακριβής ορισμός εξαρτάται συχνά από το πρόβλημα ή τον τύπο της εφαρμογής. Δεδομένου ότι ένα χαρακτηριστικό ορίζεται ως ένα «σημείο ενδιαφέροντος» μέρος μιας εικόνας, τα χαρακτηριστικά αυτά χρησιμοποιούνται ως σημείο εκκίνησης για πολλούς αλγορίθμους μηχανικής όρασης. Κατά συνέπεια, η επιθυμητή ιδιότητα για έναν ανιχνευτή χαρακτηριστικών είναι η επανάληψη δηλαδή το ίδιο χαρακτηριστικό θα ανιχνευθεί σε δύο ή περισσότερες διαφορετικές εικόνες της ίδιας σκηνής.

Η ανίχνευση χαρακτηριστικών (feature detection) [27] είναι μια λειτουργία επεξεργασίας εικόνας χαμηλού επιπέδου. Εκτελείται συνήθως ως η πρώτη ενέργεια σε μια εικόνα και εξετάζει κάθε εικονοστοιχείο (pixel) για να δει αν υπάρχει μοτίβο το οποίο να αποτελεί χαρακτηριστικό. Αν αυτό είναι μέρος ενός μεγαλύτερου αλγορίθμου, τότε ο αλγόριθμος τυπικά θα εξετάσει μόνο την εικόνα στην περιοχή των χαρακτηριστικών που επιθυμεί. Ως προϋπόθεση για την ανίχνευση χαρακτηριστικών απαιτείται η εξομάλυνση της εικόνας εισόδου π.χ από έναν Gaussian πυρήνα σε μια αναπαράσταση χώρου κλίμακας. Μέσα από την διαδικασία αυτή υπολογίζονται μία ή περισσότερες εικόνες χαρακτηριστικών, που συχνά εκφράζονται σε όρους τοπικών λειτουργιών παραγώγων εικόνας.

Περιστασιακά, εάν η ανίχνευση χαρακτηριστικών είναι υπολογιστικά δαπανηρή και υπάρχουν χρονικοί περιορισμοί, μπορεί να χρησιμοποιηθεί ένας αλγόριθμος υψηλότερου επιπέδου για την καθοδήγηση του σταδίου ανίχνευσης χαρακτηριστικών, έτσι ώστε να αναζητούνται μόνο ορισμένα τμήματα της εικόνας και να εξαχθούν μόνο τα απαραίτητα χαρακτηριστικά. Πολλοί αλγόριθμοι μηχανικής όρασης χρησιμοποιούν την ανίχνευση χαρακτηριστικών ως αρχικό βήμα, με αποτέλεσμα να αναπτυχθεί ένας πολύ μεγάλος αριθμός ανιχνευτών χαρακτηριστικών.

Η εξαγωγή χαρακτηριστικών (feature extraction) [21] είναι μια διαδικασία επεξεργασίας εικόνας η οποία έχει απασχολήσει έντονα την ερευνητική κοινότητα. Αποτελεί μια ιδιαίτερη διαδικασία κατά την οποία εφαρμόζονται αλγόριθμοι επεξεργασίας με σκοπό τη λήψη πληροφοριών και την εξαγωγή σημαντικών πληροφοριών από μια εικόνα. Στην παρούσα ενότητα παρουσιάζονται οι κύριοι αλγόριθμοι οι οποίοι χρησιμοποιούνται ευρέως από αρκετές εφαρμογές για την εξαγωγή χαρακτηριστικών από εικόνες. Για να γίνει σωστά και ολοκληρωμένα η εξαγωγή χαρακτηριστικών, θα πρέπει να χρησιμοποιηθεί ο κατάλληλος αλγόριθμος, παράλληλα με τη χρήση φίλτρων για την επεξεργασία της εικόνας, εφόσον έχει συνδυαστεί με τον κατάλληλο περιγραφέα.

3.3. Περιγραφείς Εικόνων (Image Descriptors)

Ένας περιγραφέας χαρακτηριστικών (feature descriptor) [26] είναι ένας αλγόριθμος που χρησιμοποιείται σε μια εικόνα για να εξάγει τα διανύσματα χαρακτηριστικών αυτής. Οι περιγραφείς χαρακτηριστικών κωδικοποιούν ενδιαφέρουσες πληροφορίες σε μια σειρά αριθμών και λειτουργούν ως ένα είδος αριθμητικού «δακτυλικού αποτυπώματος» που μπορεί να χρησιμοποιηθεί για να διαφοροποιήσει το ένα χαρακτηριστικό από το άλλο. Στην ιδανική περίπτωση, αυτή η πληροφορία θα ήταν αμετάβλητη κάτω από το μετασχηματισμό εικόνας, έτσι ώστε να βρεθεί το χαρακτηριστικό ακόμη και αν η εικόνα μεταμορφωθεί ή παραμορφωθεί με κάποιο τρόπο. Ένα παράδειγμα τέτοιου περιγραφέα χαρακτηριστικών είναι ο SIFT, ο οποίος κωδικοποιεί πληροφορίες σχετικά με την τοπική κλίση της γειτονιάς μιας εικόνας σε αριθμούς διανύσματος χαρακτηριστικών. Άλλα τέτοια παραδείγματα είναι ο HOG και ο SURF.

Η βασική ιδέα είναι να εντοπιστούν πρώτα τα σημεία ενδιαφέροντος ή οι περιοχές ενδιαφέροντος και στη συνέχεια να υπολογιστούν οι αμετάβλητοι περιγραφείς χαρακτηριστικών σε καθεμία από αυτές. Μόλις υπολογιστούν οι περιγραφείς χαρακτηριστικών, η αντιστοιχία χαρακτηριστικών μεταξύ διαφορετικών εικόνων μπορεί να δημιουργηθεί αυτόματα υπό κάποια μετρική ομοιότητας, π.χ. την ευκλείδεια απόσταση. Όσον αφορά την περιγραφή των χαρακτηριστικών, μια σειρά μεθόδων έχουν προταθεί στη βιβλιογραφία. Οι περισσότεροι και πιο δημοφιλείς μέθοδοι είναι αυτοί που βασίζονται σε ιστογράμματα, τα οποία επιτυγχάνουν καλύτερη απόδοση από άλλα είδη περιγραφικών στοιχείων όπως ο περιγραφέας βάσει φίλτρου (filter descriptor) κ.λπ. Ωστόσο, υπάρχουν αρκετοί περιγραφείς οι οποίοι είναι πλήρως ή μερικώς ανθεκτικοί σε πολλές από τις παραλλαγές και παραμορφώσεις και δεν μπορούν να χειριστούν τις περίπλοκες αλλαγές σε φωτισμό συμπεριλαμβανομένης της διόρθωσης γάμμα, αντανακλάσεων, αλλαγές στο χρόνο έκθεσης κλπ.

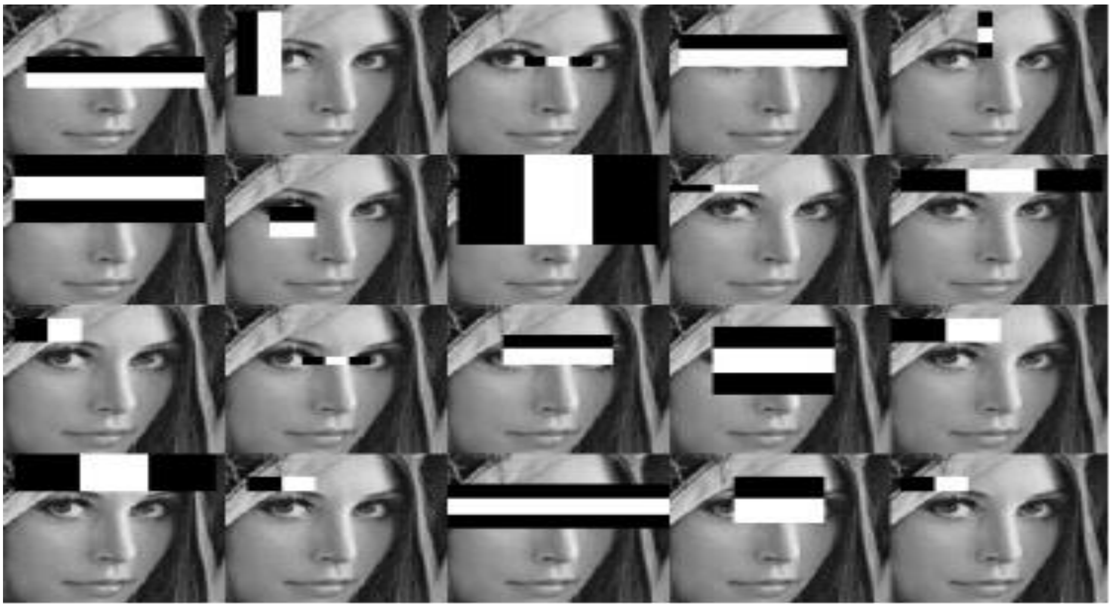
Στην παρούσα ενότητα παρουσιάζονται οι βασικοί περιγραφείς οι οποίοι χρησιμοποιούνται κατά κόρον στις μεθόδους επαναταυτοποίησης ατόμων.

3.3.1. Μετασχηματισμοί Haar

Τα χαρακτηριστικά Haar χρησιμοποιούνται στην αναγνώριση αντικειμένων. Οφείλουν το όνομά τους στην διαισθητική τους ομοιότητα με τα κύματα Haar και χρησιμοποιήθηκαν στον πρώτο ανιχνευτή προσώπου πραγματικού χρόνου. [31]

Ιστορικά, η επεξεργασία εικόνας με τις τιμές των εικονοστοιχείων RGB σε κάθε εικόνα έκαναν την εργασία υπολογισμού χρονοβόρα και δαπανηρή. Μια δημοσίευση των Παπαγεωργίου και άλλων [34]. παρουσίασε ένα εναλλακτικό σύνολο χαρακτηριστικών βασισμένο σε κύματα Haar αντί για τις συνήθεις εντάσεις εικόνας. Οι Viola και Jones [31] προσάρμοσαν την ιδέα της χρήσης κυματοειδών μετασχηματισμών Haar και ανέπτυξαν τα λεγόμενα χαρακτηριστικά Haar (Haar features).

Ένα χαρακτηριστικό γνώρισμα Haar θεωρεί γειτονικές ορθογώνιες περιοχές σε μια συγκεκριμένη θέση με ένα παράθυρο ανίχνευσης, συνοψίζοντας τις εντάσεις των εικονοστοιχείων σε κάθε περιοχή και υπολογίζοντας τη διαφορά μεταξύ αυτών των αθροισμάτων.Αυτή η διαφορά στη συνέχεια χρησιμοποιείται για την κατηγοριοποίηση τμημάτων μιας εικόνας. Για να γίνει κατανοητή η χρήση των Haar χαρακτηριστικών, θεωρείται πως υπάρχει μια βάση δεδομένων εικόνων με ανθρώπινα πρόσωπα. Είναι κοινή παρατήρηση ότι η περιοχή των ματιών είναι πιο σκούρα από την υπόλοιπη περιοχή του προσώπου. Επομένως, ένα κοινό χαρακτηριστικό γνώρισμα Haar για την ανίχνευση προσώπου είναι ένα σύνολο δύο παρακείμενων ορθογωνίων που βρίσκονται πάνω και κάτω από το μάτι. Η θέση αυτών των ορθογωνίων καθορίζεται σε σχέση με ένα παράθυρο ανίχνευσης που ενεργεί σαν ένα πλαίσιο οριοθέτησης του αντικειμένου στόχου δηλ. του προσώπου σε αυτή τη περίπτωση. [34]



Σχήμα 3-2 Χαρακτηριστικά Haar που χρησιμοποιούνται από μια εικόνα.

Στη φάση ανίχνευσης του πλαισίου των αντικειμένων (με τη μέθοδο των Viola-Jones), ένα παράθυρο συγκεκριμένου μεγέθους μετακινείται πάνω από την εικόνα εισόδου και για κάθε υποενότητα της εικόνας υπολογίζεται το χαρακτηριστικό γνώρισμα Haar. Αυτή η διαφορά στη συνέχεια συγκρίνεται με ένα όριο που χωρίζει τα μη αντικείμενα από τα αντικείμενα. Επειδή ένα τέτοιο χαρακτηριστικό γνώρισμα Haar είναι μόνο ένας ασθενής ταξινομητής, ένας μεγάλος αριθμός χαρακτηριστικών τύπου Haar είναι απαραίτητος για την περιγραφή ενός αντικειμένου με επαρκή ακρίβεια. Το βασικό πλεονέκτημα ενός χαρακτηριστικού τύπου Haar έναντι των περισσότερων άλλων χαρακτηριστικών είναι η ταχύτητα υπολογισμού του. Λόγω της χρήσης ολοκληρωμένων εικόνων, ένα χαρακτηριστικό γνώρισμα Haar οποιουδήποτε μεγέθους μπορεί να υπολογιστεί σε σταθερό χρόνο.

3.3.2. Περιγραφέας SURF

Ο περιγραφέας Surf [15] έχει κατασκευαστεί ώστε να αποτελέσει ένα σημαντικά γρήγορο σχήμα ανίχνευσης και περιγραφής σημείων. Συνοπτικά, επιτυγχάνει:

- επαναληπτικότητα, δηλαδή τη σταθερή εύρεση των ίδιων σημείων ανεξάρτητα με τις περιστροφές του αντικειμένου.
- διακριτότητα των περιοχών που αποτελούν γωνιακά σημεία και κηλίδες (blobs).
- ευρωστία στο θόρυβο και στις αλλαγές φωτεινότητας. Με σκοπό την ανίχνευση χαρακτηριστικών σημείων ο αλγόριθμος υιοθετεί μια γρήγορη προσέγγιση του πίνακα Hessian σε συνδυασμό με τις εικόνες ολοκλήρωσης (integral). Οι εικόνες αυτές [16] αποτελούν ένα γρήγορο σχήμα υπολογισμού της συνέλιξης με διάφορα τετραγωνικά φίλτρα, καθώς συνεισφέρουν στην ανεξαρτητοποίηση του χρόνου υπολογισμού από το μέγεθος του φίλτρου.

Ο Εσσιανός πίνακας (Hessian matrix) σε μία εικόνα, δοσμένου ενός σημείου (x,y) μίας κλίμακας σ ορίζεται ώς

$$\mathrm{H}(x,\sigma) = \begin{pmatrix} L_{xx}(x,\sigma) & L_{xy}(x,\sigma) \\ L_{xy}(x,\sigma) & L_{yy}(x,\sigma) \end{pmatrix} \tag{3.1}$$

με τα L να αποτελούν τις συνελίξεις τις εικόνας I στο σημείο αυτό και με τις αντίστοιχες Gaussian παραγώγους δεύτερης τάξης. Η γρήγορη προσέγγιση του Hessian πίνακα στηρίζεται στη χρήση φίλτρων 9 × 9 με $\sigma = 1.2$, τα οποία προσεγγίζουν τις Gaussian παραγώγους δεύτερης τάξης.

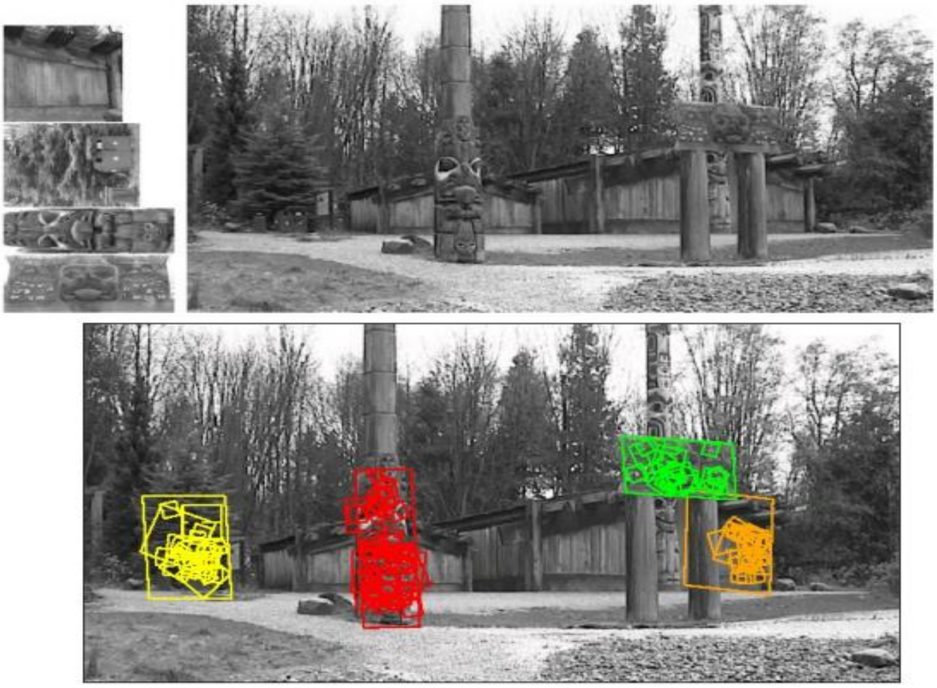
Ο περιγραφέας Surf βασίζεται στην κατανομή πρώτης τάξης κυματιδίων Haar με σκοπό να προσδιορίσει το περιεχόμενο της έντασης γύρω από το κάθε σημείο ενδιαφέροντος. Υπολογίζει την κατανομή των κυματιδίων γύρω από τα σημεία ενδιαφέροντος σε διαφορετικές κλίμακες και προσθέτει τις κατανομές ώστε να προκύψει ο κυρίαρχος προσανατολισμός του σημείου. Για το χαρακτηριστικό διάνυσμα του σημείου ενδιαφέροντος ορίζεται μια τετράγωνη περιοχή με μέγεθος 20s, όπου s η αντίστοιχη κλίμακα. Στην συνέχεια το διάνυσμα αυτό υποδιαιρείται σε περιοχές 4 × 4 και ανάλογα με την κατανομή των κυματιδίων Haar ορίζεται ο προσανατολισμός του κάθε τμήματος. Το τελικό διάνυσμα προκύπτει από την ένωση των προσανατολισμών κάθε περιοχής.

3.3.3. Περιγραφέας SIFT

Ο μετασχηματισμός μεταβλητής κλίμακας (SIFT) [23] είναι ένας περιγραφέας εικόνας για αντιστοίχιση και αναγνώριση ομοιοτήτων βάσει εικόνας. Ο αλγόριθμος αυτός αναπτύχθηκε από τον David Lowe (1999, 2004). Αυτός ο περιγραφέας καθώς και οι σχετικοί περιγραφείς εικόνων χρησιμοποιούνται για μεγάλο αριθμό σκοπών στη μηχανική όραση, που σχετίζονται με την αντιστοίχιση σημείων μεταξύ διαφορετικών προβολών μιας 3-D σκηνής και αναγνώρισης αντικειμένων βασισμένης στην προβολή αυτή.

Ο περιγραφέας SIFT είναι αμετάβλητος όσον αφορά τους μετασχηματισμούς κλιμάκωσης στον τομέα της εικόνας και «ανθεκτικός» στις παραλλαγές φωτισμού. Όπως είναι γνωστό ο φωτισμός είναι ο μεγαλύτερος εχθρός της επεξεργασίας εικόνας, καθώς εισάγει θόρυβο μέσα στην εικόνα με αποτέλεσμα να αλλοιωθεί το αποτέλεσμα της επεξεργασίας, αλλά και η διαδικασία μάθησης. Από πειραματικά αποτελέσματα, ο περιγραφέας SIFT αποδείχθηκε πολύ χρήσιμος στην πράξη για την αντιστοίχιση εικόνων και την αναγνώριση αντικειμένων σε συνθήκες του πραγματικού κόσμου.

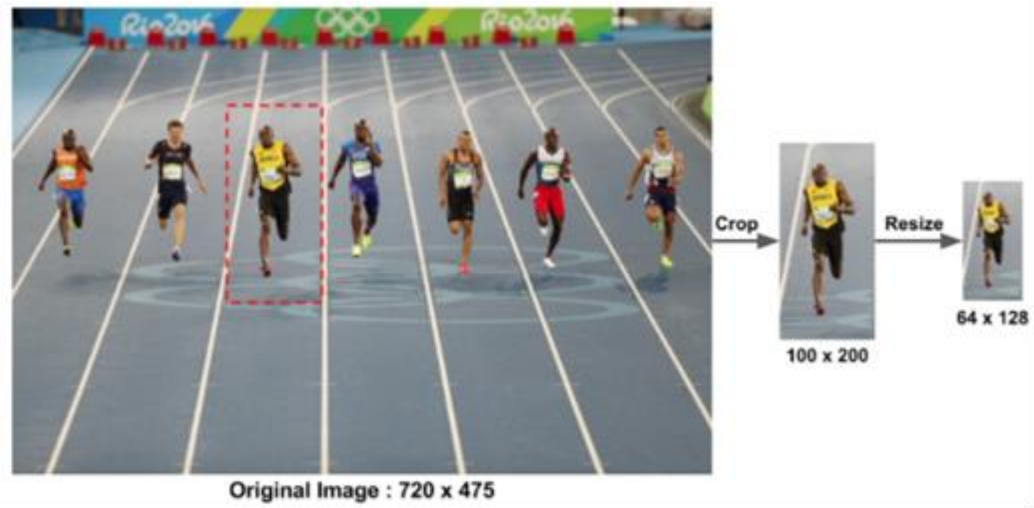
Στην αρχική διατύπωση, ο περιγραφέας SIFT περιείχε μια μέθοδο για την ανίχνευση σημείων ενδιαφέροντος από μια εικόνα επιπέδου γκρίζου στην οποία συγκεντρώθηκαν στατιστικές τοπικών διαβαθμίσεων των εντάσεων της. Στόχος ήταν να δοθεί μια περιληπτική περιγραφή των τοπικών δομών της εικόνας σε μια γειτονιά γύρω από κάθε σημείο ενδιαφέροντος, με σκοπό την χρήση αυτού του περιγραφικού στοιχείου για την αντιστοίχιση των σημείων ενδιαφέροντος μεταξύ διαφορετικών εικόνων. Αργότερα, ο περιγραφέας SIFT εφαρμόστηκε επίσης σε πυκνά πλέγματα (πυκνό SIFT) τα οποία έχουν αποδειχθεί ότι οδηγούν σε καλύτερη απόδοση για εργασίες όπως κατηγοριοποίηση αντικειμένων, ταξινόμηση υφής, ευθυγράμμιση εικόνας και βιομετρία.



Σχήμα 3-3 Αποτελέσματα ανίχνευσης στοιχείων με τον αλγόριθμο SIFT. [23]

3.3.4. Ιστόγραμμα Διαβαθμίσεων (Histogram of Oriented Gradients ή HOG)

Τυπικά, ένας περιγραφέας χαρακτηριστικών μετατρέπει μια εικόνα μεγέθους: πλάτος x ύψος x 3 (κανάλια) σε ένα διάνυσμα / πίνακα χαρακτηριστικών μήκους n. Στην περίπτωση του περιγραφέα χαρακτηριστικών Histogram of oriented gradients (HOG), σε μια εικόνα εισόδου με μέγεθος 64 x 128 x 3 το διάνυσμα χαρακτηριστικών εξόδου είναι μεγέθους 3780. Τα μήκη αυτά χρησιμοποιούνται ως βασικοί παράμετροι και προτείνονται για την καλύτερη εξαγωγή των χαρακτηριστικών, ωστόσο ποικίλουν ανάλογα με το μέγεθος και τα κανάλια τα οποία φέρουν ή θα χρησιμοποιηθούν από μια εικόνα. Ως παράδειγμα δίνεται μια εικόνα μεγέθους 720 × 475 (Σχήμα 3-4). Από την εικόνα επιλέγεται ένα άτομο και στη συνέχεια κόβετε σε μέγεθος 100 × 200 για τον υπολογισμό του περιγραφέα χαρακτηριστικών HOG. Προκειμένου να μπορέσει να χρησιμοποιηθεί καλύτερα η εικόνα, γίνεται μια αλλαγή στο μέγεθος αυτής σε 64 × 128 και στη συνέχεια εφαρμόζετε ο περιγραφέας HOG.



Σχήμα 3-4 Διαδικασία περικοπής εικόνων στην εφαρμογή του περιγραφέα HOG.

Στον περιγραφέα χαρακτηριστικών HOG, κύριο ρόλο παίζει η παραγωγή των χαρακτηριστικών γνωρισμάτων κατανομής (ιστογράμματα) και διευθύνσεων κλίσεων (προσανατολισμένες κλίσεις). Οι διαβαθμίσεις (παράγωγα x και y) μιας εικόνας είναι χρήσιμες καθώς το μέγεθος των διαβαθμίσεων είναι μεγάλο γύρω από τις ακμές και τις γωνίες (περιοχές με απότομες αλλαγές έντασης). Σημειώνεται ότι οι ακμές και οι γωνίες μιας εικόνας φέρουν περισσότερες πληροφορίες σχετικά με το σχήμα του αντικειμένου από τις επίπεδες περιοχές.

Η διαδικασία απόσπασης του περιγραφέα μπορεί να χωριστεί σε τέσσερα βήματα:

1. Τον υπολογισμό της κλίσης του φορέα για κάθε ένα από τα εικονοστοιχεία
2. Την ομαδοποίηση των εικονοστοιχείων σε κελιά (cells)
3. Τη δημιουργία μπλοκ από ομάδες κελιών.
4. Τη δημιουργία του περιγραφέα.

Για να υπολογιστεί ένας περιγραφέας HOG, πρέπει πρώτα να υπολογιστούν οι οριζόντιες και κάθετες κλίσεις σε μια εικόνα. Ο υπολογισμός των κλίσεων γίνεται μέσα από τον καθορισμό των τιμών με βάση τα εικονοστοιχεία της εικόνας, προκειμένου να υπολογιστεί το ιστόγραμμα των κλίσεων. Αυτό επιτυγχάνεται εύκολα φιλτράροντας την εικόνα με τους ακόλουθους πυρήνες (Σχήμα 3-5).

-1	0	1

-1
0
1

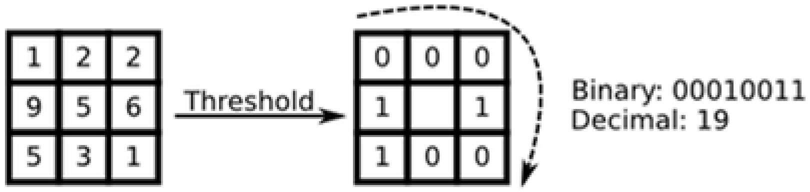
Σχήμα 3-5 Πυρήνες που χρησιμοποιούνται για το φιλτράρισμα των εικόνων.

Ο περιγραφέας αυτός είναι αρκετά χρονοβόρος στην παραγωγή αποτελέσματος, ωστόσο έχει αυξημένα αποτελέσματα όταν χρησιμοποιείται στην ανίχνευση και αναγνώριση ενός αντικειμένου. Ωστόσο, η επιλογή του σωστού περιγραφέα εξαρτάται από την εκάστοτε περίπτωση - πρόβλημα καθώς και από τα στοιχεία - αντικείμενα τα οποία τίθενται προς ανίχνευση.

3.3.5. Αλγόριθμος Local Binary Patterns (LBP)

Ο αλγόριθμος LBP κατηγοριοποιεί και δίνει ετικέτες στα pixel μιας εικόνας με δεκαδικούς αριθμούς, οι οποίοι ονομάζονται τοπικά δυαδικά μοτίβα (Local binary patterns) ή αλλιώς LBP. Τα μοτίβα αυτά κωδικοποιούν την τοπική δομή γύρω από κάθε pixel, όπως φαίνεται στην εικόνα που ακολουθεί (3-6). Ουσιαστικά η διαδικασία που εκτελείται είναι η εξής: Κάθε pixel συγκρίνεται με τα οκτώ γειτονικά του σε μια γειτονιά 3x3 αφαιρώντας την τιμή του κεντρικού pixel. Η σύγκριση γίνεται δεξιόστροφα. Εάν η τιμή της έντασης του ενός pixel είναι μεγαλύτερη ή ίση με το γείτονά της, τότε η τιμή του pixel αυτού αντικαθίσταται με τον αριθμό 1, ενώ σε αντίθετη περίπτωση με τον αριθμό 0. Η εικόνα που θα προκύψει θα είναι ασπρόμαυρη και θα διακρίνονται με λευκό χρώμα οι πιο έντονες περιοχές της.

Σημειώνεται ότι, η τιμή που φέρει το κάθε pixel αφορά την ένταση την οποία έχει. Για να λάβουμε την τιμή αυτής της περιοχής σε δεκαδικό αριθμό, απλά βάζουμε τα pixel στη σωστή σειρά και μετατρέπουμε τον δυαδικό αριθμό σε δεκαδικό. Η παραπάνω διαδικασία παρουσιάζεται στην εικόνα που ακολουθεί:

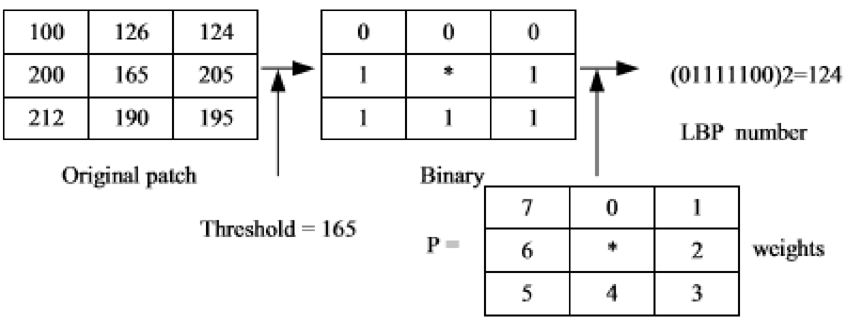


Σχήμα 3-6 Μετατροπή μιας συστάδας pixel με βάση τον αλγόριθμο LBP.

Η συνάρτηση η οποία ακολουθεί παρουσιάζει τον υπολογισμό με τον οποίο παράγεται το τελικό LBP αποτέλεσμα.

$$LBP_{P,R} = \sum_{i=0}^{P-1} s(g_i - g_c)2^i. \qquad s(x) = \begin{cases} 1, x > 0 \\ 0, x \leq 0 \end{cases} \tag{3.2}$$

Το παράδειγμα που ακολουθεί παρουσιάζει τη λειτουργία του LBP αλγορίθμου και τη μετατροπή μιας συστάδας pixel. Στον αλγόριθμο χρησιμοποιείται και το όριο (threshold) το οποίο έχει τεθεί στην τιμή 165. Οι τιμές των pixel κάτω από το όριο αυτό μετατρέπονται σε 0, ενώ αντίθετα οι άλλες τιμές παίρνουν την τιμή 1. Το τελικό αποτέλεσμα που προκύπτει είναι ο αριθμός 124. [6]



Σχήμα 3-7 Παράδειγμα χρήσης του αλγορίθμου LBP.

3.3.6. Μέτρα Απόστασης (Distance Measures)

Στην επεξεργασία εικόνας, το μέτρο απόστασης (distance measure) υπολογίζει την απόσταση κάθε σημείου αντικειμένου από το πλησιέστερο όριο. Ο μετασχηματισμός αυτός είναι ένα σημαντικό εργαλείο στην μηχανική όραση, την επεξεργασία εικόνας και την αναγνώριση μοτίβου. Στη μετασχηματισμένη απόσταση, η δυαδική εικόνα καθορίζει την απόσταση από το καθένα εικονοστοιχείο (pixel) στο πλησιέστερο μη μηδενικό pixel (οι τιμές λαμβάνονται με βάση την ένταση του εκάστοτε pixel).

Προκειμένου να γίνει σωστή σύγκριση χαρακτηριστικών ή ιστογραμμάτων που λαμβάνονται από δύο εικόνες ή τμήματα μιας εικόνας, μερικές μετρήσεις απόστασης μπορεί να χρησιμοποιηθούν. Δεδομένου ότι οι δύο διανύσματα $X = (x_1,x_2,\ldots,x_n)$ και $Y = (y_1,y_2,\ldots,y_n)$ αντιστοιχούν σε δύο διακριτά ιστογράμματα, μπορούν να χρησιμοποιηθούν οι παρακάτω μετρήσεις για να προσδιοριστεί η ομοιότητα των εικόνων. Οι μετρήσεις αυτές καθορίζουν συνήθως την απόσταση μεταξύ δύο δειγμάτων για την αξιολόγηση της διαφοράς τους, η οποία με τη σειρά της μπορεί να χρησιμοποιηθεί σαν μέτρο ομοιότητας.

- Ευκλείδεια απόσταση

Η Ευκλείδεια απόσταση είναι η απόσταση ευθείας γραμμής μεταξύ δύο pixel και αξιολογείται χρησιμοποιώντας τον κανόνα του Ευκλείδη. Η απόσταση σταθμού μέτρησης υπολογίζει τη διαδρομή μεταξύ των pixel βάσει τεσσάρων συνδεδεμένων γειτονιών των οποίων τα άκρα αγγίζουν με μια μονάδα και τα pixel που αγγίζουν διαγώνια είναι δύο μονάδες. Ανάλογα με τον υπολογισμό παράγεται και το αντίστοιχο αποτέλεσμα.

$$d(x,y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \tag{3.3}$$

- Τετραγωνική απόσταση (Chi-square distance)

Η απόσταση αυτή είναι χρήσιμη κατά τη σύγκριση δύο εικόνων ή ιστογραμμάτων σημείο προς σημείο [25]. Η ανομοιομορφία του εκάστοτε σημείου δεν είναι βοηθητική, με αποτέλεσμα τα μέτρα να συγκρίνουν μόνο τις αντίστοιχες τιμές σε ένα ιστόγραμμα. Στις περισσότερες περιπτώσεις γίνεται η υπόθεση ότι η περιοχή των ιστογραμμάτων είναι ευθυγραμμισμένη. Το όνομα αυτής της μέτρησης για την απόσταση προέρχεται από τον Pearson's [80] η οποία χρησιμοποιείται για να δείξει την πιθανότητα μιας εικόνας να αντλείται από μια άλλη. Σε ορισμένες περιπτώσεις, π.χ. ενώ συγκρίνονται ιστογράμματα. η διαφορά μεταξύ των μικρότερων στοιχείων γίνεται πιο σημαντική από τη διαφορά μεταξύ των μεγάλων ομάδων. Ο τύπος που χρησιμοποιείται για τη μέτρηση αυτή είναι:

$$\mathrm{X}^2(\mathrm{X},\mathrm{Y}) = \frac{1}{2}\sum_i \frac{(\mathrm{X}_i - Y_i)^2}{(X_i + Y_i)} \tag{3.4}$$

• Απόσταση Mahalanobis

Η απόσταση Mahalanobis είναι βασισμένη σε συσχετισμούς μεταξύ των μεταβλητών τιμών από τις οποίες τα διαφορετικά στοιχεία μπορούν να προσδιοριστούν και να αναλυθούν. Είναι ένας χρήσιμος τρόπος ορισμού και υπολογισμό της ομοιότητας από ένα άγνωστο σύνολο δειγμάτων σε έναν γνωστό. Διαφέρει από Ευκλείδεια απόσταση δεδομένου ότι λαμβάνει υπόψη τους συσχετισμούς ενός συνόλου στοιχείων.

Τυπικά, η απόσταση Mahalanobis από μια ομάδα τιμών, η οποία ορίζεται με τον ακόλουθο τύπο:

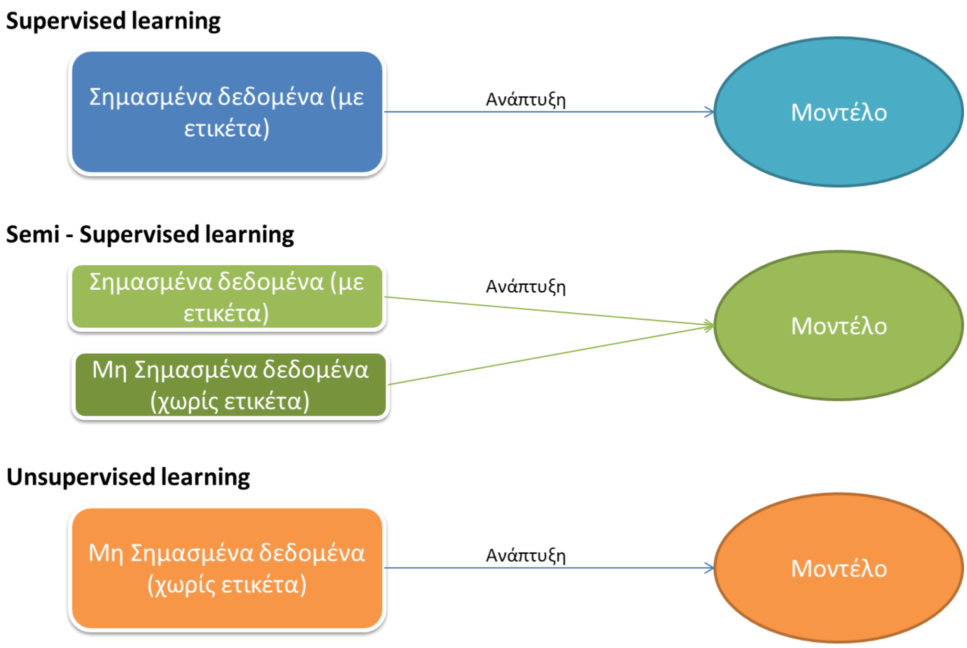
$$d\ (x,y) = \sqrt{(x-y)C^{-1}(x-y)^T} \tag{3.5}$$

$$C = \begin{pmatrix} 0.9167 & -0.6667 \\ -0.6667 & 0.6667 \end{pmatrix} \tag{3.6}$$

3.4. Διαδικασίες Εκπαίδευσης

Οι εργασίες μηχανικής μάθησης ταξινομούνται σε τρεις μεγάλες κατηγορίες, ανάλογα με τη φύση του «σήματος» εκπαίδευσης ή την «ανατροφοδότηση» που είναι διαθέσιμα σε ένα σύστημα εκμάθησης. Ανάλογα με τον τύπο των δεδομένων και του προβλήματος στο οποίο θα χρησιμοποιηθούν τα μοντέλα τα οποία θα αναπτυχθούν μέσα από την διαδικασία της μάθησης, επιλέγεται και η κατάλληλη κατηγορία για την εκπαιδευτική διαδικασία και συλλέγονται τα κατάλληλα δείγματα τα οποία θα αποτελέσουν την κύρια βάση δεδομένων για το σύστημα.

Στην ενότητα που ακολουθεί παρουσιάζονται αναλυτικά τα στοιχεία για κάθε κατηγορία, όπως φαίνονται και συνοπτικά στην ακόλουθη εικόνα.



Σχήμα 3-8 Διαχωρισμός διαδικασιών εκπαίδευσης.

3.4.1. Εποπτευόμενη Μάθηση (Supervised Learning)

Η εποπτευόμενη μάθηση (supervised learning) [14] αφορά το κομμάτι της εκπαίδευσης, όπου υπάρχουν οι μεταβλητές εισόδου (X) και μια μεταβλητή εξόδου (Y) και χρησιμοποιείτε ένας αλγόριθμος ο οποίος θα χαρτογραφήσει την έξοδο λαμβάνοντας ως δεδομένη την είσοδο. Η συνάρτηση η οποία αντιστοιχεί στο κομμάτι αυτό της εκπαίδευσης είναι η ακόλουθη:

$$Y = f(X) \tag{3.7}$$

Απώτερος στόχος αυτού του τρόπου εκπαίδευσης είναι η επίτευξη ενός καλύτερου αποτελέσματος στη διαδικασία της χαρτογράφησης – σύγκρισης των δεδομένων, όταν υπάρχουν νέα δεδομένα εισόδου (X). Γνωρίζοντας τα δεδομένα αυτά, μπορεί ο αλγόριθμος να προβλέψει τις μεταβλητές εξόδου (Y).

Το όνομα της μεθόδου αυτής προκύπτει από τη διαδικασία του αλγορίθμου που μαθαίνει από το σύνολο δεδομένων εισόδου (κατάρτισης), το οποίο μπορεί να θεωρηθεί ως δάσκαλος που εποπτεύει τη διαδικασία μάθησης σε ένα σύστημα εκπαίδευσης. Η διαδικασία η οποία ακολουθείται αφορά ουσιαστικά τη συνεχή προσπάθεια του αλγορίθμου να μάθει όσο περισσότερα μπορεί για το σύνολο εκπαίδευσης, μέσα από συνεχείς ερωτήσεις και συγκρίσεις, γνωρίζοντας ήδη από την αρχή της διαδικασίας αυτής τα σωστά στοιχεία. Η μάθηση στην περίπτωση αυτή σταματά όταν ο αλγόριθμος επιτυγχάνει ένα αποδεκτό επίπεδο απόδοσης.

Το δείγμα το οποίο εισάγεται για την εκπαίδευση του συστήματος και την ανάπτυξη του μοντέλου πρέπει να φέρει τις κατάλληλες τιμές - ετικέτες οι οποίες αφορούν τη κατηγορία στην οποία ανήκουν. Π.χ τα στοιχεία τα οποία ανήκουν σε μία κατηγορία φέρουν το όνομα «κόκκινο» και αντίστοιχα τα στοιχεία μιας άλλης κατηγορίας το όνομα «μπλε». Οι ετικέτες αυτές θα χρησιμοποιηθούν στην επόμενη διαδικασία, εκείνη της αναγνώρισης, όπου ο αλγόριθμος θα κληθεί ανάλογα με το δείγμα το οποίο έχει εισαχθεί, και με τη διαδικασία την οποία έχει εκπαιδευτεί, να μπορέσει να αναγνωρίσει τα εκάστοτε αντικείμενα εισόδου.

Τα προβλήματα τα οποία υπάρχουν στην διαδικασία αυτή μπορούν να διαχωριστούν περαιτέρω σε προβλήματα παλινδρόμησης (regression) και ταξινόμησης (classification).

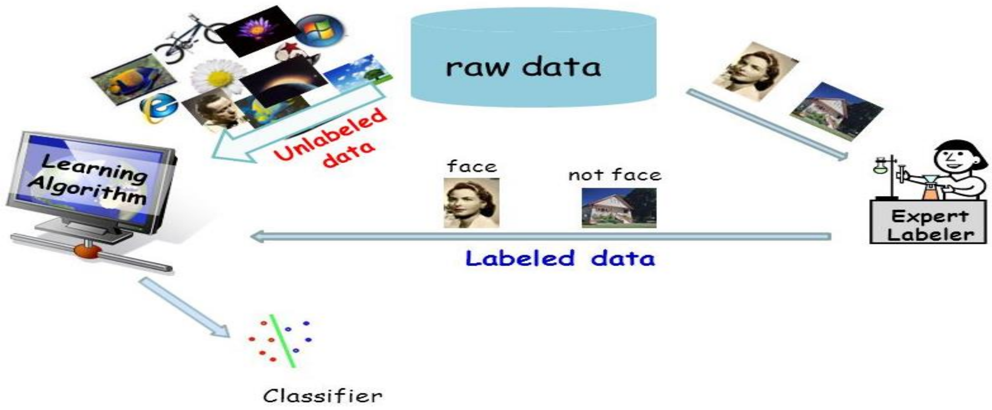
- Ταξινόμηση (Classification): Τα προβλήματα ταξινόμησης σχετίζονται με την επιλογή μιας συγκεκριμένης κατηγορίας, όπως για παράδειγμα την ταξινόμηση των στοιχείων σε κατηγορίες π.χ «κόκκινο» ή «μπλε».

- Παλινδρόμηση (Regression): Ένα πρόβλημα παλινδρόμησης θεωρείται όταν η μεταβλητή εξόδου είναι μια πραγματική τιμή όπως π.χ «ευρώ» ή «βάρος». Ορισμένοι συνηθισμένοι τύποι προβλημάτων που στηρίζονται στην παλινδρόμηση περιλαμβάνουν την πρόβλεψη σύστασης και τη χρήση χρονοσειρών, αντίστοιχα.

3.4.2. Ημι-Εποπτευόμενη Μάθηση (Semi-supervised Learning)

Η ημι-εποπτευόμενη μάθηση (Semi-supervised learning) [17] έχει αρκετά κοινά με την προαναφερθείσα κατηγορία μάθησης (supervised learning). Στην περίπτωση αυτή μια ποσότητα δεδομένων εισόδου (X) είναι επισημασμένα, δηλαδή φέρουν ετικέτες. Ένα τέτοιο παράδειγμα είναι π.χ ένα αρχείο φωτογραφιών όπου μόνο μερικές από τις εικόνες είναι επισημασμένες (σκύλος, γάτα, πρόσωπο) ενώ η πλειοψηφία αυτών δεν έχει επισημανθεί.

Πολλά προβλήματα τα οποία χρησιμοποιούν τεχνικές μάθησης εμπίπτουν στις παραπάνω κατηγορίες. Ωστόσο η χρήση τους σε προβλήματα πραγματικού κόσμου αποτελεί μια απαιτητική διαδικασία. Αυτό οφείλεται στο γεγονός ότι η επισήμανση των δεδομένων μπορεί να είναι μια δαπανηρή και χρονοβόρα διαδικασία, καθώς μπορεί επίσης να απαιτεί πρόσβαση σε εμπειρογνώμονες του εκάστοτε τομέα στον οποίο ανήκουν τα δεδομένα. Αντίθετα, τα μη επισημασμένα δεδομένα μπορούν εύκολα να συλλεχθούν και δεν απαιτούν τόσο μεγάλο χρόνο. Η διαδικασία εκπαίδευσης που ακολουθείται και στη περίπτωση αυτή, καθώς και οι αλγόριθμοι, είναι κοινή με τις παραπάνω διαδικασίες, με την διαφορά ότι τα αποτελέσματα αναγνώρισης ενδέχεται να μην τόσο αυξημένα.



Σχήμα 3-9 Διαδικασία Semi-supervised learning.

3.4.3. Μη-Εποπτευόμενη Μάθηση (Unsupervised Learning)

Η μη-εποπτευόμενη μάθηση (unsupervised learning) αφορά τη διαδικασία εκπαίδευσης μόνο με δεδομένα εισόδου (X) και χωρίς να υπάρχει καμία αντίστοιχη μεταβλητή εξόδου. Ο στόχος αυτής της εκπαιδευτικής διαδικασίας χωρίς επίβλεψη είναι να μοντελοποιήσει την υποκείμενη δομή ή τη διανομή στα δεδομένα προκειμένου να εξάγει περισσότερα χαρακτηριστικά και να γνωρίσει περισσότερα για τα δεδομένα.

Το όνομα της διαδικασίας αυτής, αλλά και η λογική της έρχεται σε αντίθεση με την εποπτευόμενη μάθηση που προαναφέρθηκε, καθώς στην περίπτωση αυτή δεν υπάρχουν σωστές απαντήσεις και δεν υπάρχει δάσκαλος. Οι αλγόριθμοι αφήνονται στις δικές τους γνώσεις, οι οποίες προκύπτουν από την εξαγωγή χαρακτηριστικών που αναφέρθηκαν παραπάνω, προκειμένου να ανακαλύψουν την ενδιαφέρουσα δομή των δεδομένων.

Τα δείγματα εισόδου τα οποία εισάγονται στην διαδικασία αυτή δεν φέρουν κάποια ετικέτα, καθώς η ανίχνευση και αναγνώριση των αντικειμένων βασίζεται κυρίως στην διαδικασία της σωστής και ολοκληρωμένης εξαγωγής χαρακτηριστικών. Τα προβλήματα στο unsupervised learning μπορούν να διαχωριστούν περαιτέρω σε προβλήματα ομαδοποίησης (Clustering) και συσχέτισης (Association).

- Ομαδοποίηση (Clustering): Ένα τέτοιο πρόβλημα σχετίζεται με τις εγγενείς ομαδοποιήσεις των δεδομένων, όπως η ομαδοποίηση πελατών με παρόμοια αγοραστική συμπεριφορά. Στην διαδικασία της ομαδοποίησης, κατατάσσονται τα στοιχεία σε κατηγορίες ανάλογα με κάποιο από τα χαρακτηριστικά τους. Συνήθως στην διαδικασία αυτή χρησιμοποιείται ο αλγόριθμος k-means.
- Συσχέτιση (Association): Τα προβλήματα αυτής της κατηγορίας αφορούν την ανάπτυξη κανόνων συσχέτισης μεταξύ των δεδομένων, όπως π.χ τα άτομα που αγοράζουν το Χ προϊόν έχουν επίσης την τάση να αγοράζουν και το Υ. Στην περίπτωση αυτή ο αλγόριθμος Apriori είναι ο συχνότερα χρησιμοποιούμενος αλγόριθμος, και παρουσιάζει σύμφωνα με τη βιβλιογραφία αυξημένα αποτελέσματα.

3.4.4. Ανοιχτή / Κλειστή Επαναταυτοποίηση (Open / Closed Re-identification)

Το πρόβλημα της επαναταυτοποίησης του ατόμου (Person Re-identification) εμπεριέχει τόσο την ανίχνευση, όσο και την αναγνώριση ενός ατόμου, μέσα από ένα σύνολο δειγμάτων. Η διαδικασία αυτή μπορεί γενικά να κατηγοριοποιηθεί σε δύο τύπους ως [20]:

- Επαναταυτοποίηση κλειστού τύπου (Closed set re-identification)

- Επαναταυτοποίηση ανοιχτού τύπου (Open set re-identification)

Η περίπτωση της επαναταυτοποίησης κλειστού τύπου γίνεται με την εκπαίδευση του συστήματος προκειμένου να υπάρχει γνώση σχετικά με την ομάδα ατόμων που τίθεται προς ανίχνευση, σχηματίζοντας έτσι σύνολα «ατόμων». Στην περίπτωση της στην κατηγορίας αυτής η ανίχνευση - αναγνώριση του ατόμου γίνεται μέσα από τα στοιχεία τα οποία έχει εκπαιδευτεί ο εκάστοτε αλγόριθμος. Επομένως μία ερώτηση για ένα άτομο που βρίσκεται εκτός αυτού του κλειστού σετ εικόνων θα είχε ένα λανθασμένο αποτέλεσμα, χωρίς να υποδηλώνεται ότι είναι άγνωστο πρόσωπο. Η χρήση επομένως της μεθόδου αυτής σε δημόσιους χώρους (όχι κλειστό περιβάλλον) δεν προτείνεται στη περίπτωση αυτή.

Αντίθετα, ένα σύστημα επαναταυτοποίησης ανοικτού τύπου μπορεί να προσδιορίσει με επιτυχία εάν το άτομο το οποίο τίθεται προς παρατήρηση είναι ένα αντικείμενο το οποίο εισέρχεται πρώτη φορά ή εάν έχει παρατηρηθεί ξανά στο παρελθόν. Ουσιαστικά, αν ένα μη αναγνωρίσιμο άτομο εισέλθει στο σύστημα, τότε το σύστημα καλείται να τον ανιχνεύσει προσθέτοντας το προφίλ των χαρακτηριστικών του νέου επισκέπτη στο σύνολο των γνωστών ατόμων για να διευκολύνει τον επαναπροσδιορισμό του σε κάποια μελλοντική επίσκεψη. Αν και υπάρχουν αρκετά συστήματα αναγνώρισης τέτοιου τύπου δεν είναι επίσης κατάλληλα για δημόσιους χώρους καθώς παρουσιάζουν σημαντικά μειονεκτήματα στην ακρίβεια και την απόδοση τους, λόγω της μη επαρκούς ανάλυσης δεδομένου ότι τα πρόσωπα ενδέχεται να καταλαμβάνουν μια πολύ μικρή περιοχή στο πλαίσιο παρατήρησης.

3.5. Βασικές Τεχνικές Ταξινόμησης

Προκειμένου να επιτευχθεί η διαδικασία της ανίχνευσης ή της αναγνώρισης ενός αντικειμένου, θα πρέπει να υπάρχουν οι κατάλληλοι αλγόριθμοι οι οποίοι χρησιμοποιούνται για την σύγκριση και την ταξινόμηση των δειγμάτων στις αντίστοιχες κατηγορίες, προκειμένου να παρουσιαστεί το τελικό αποτέλεσμα. Οι ταξινομητές οι οποίοι παρουσιάζονται στην παρούσα ενότητα είναι οι βασικότεροι πάνω στους οποίους βασίζονται άλλες επεκτάσεις που θα παρουσιαστούν στα επόμενα κεφάλαια.

3.5.1. Μηχανές Διανυσμάτων Στήριξης (Support Vector Machines, SVM)

Οι μηχανές διανυσμάτων στήριξης (SVM) υλοποιούν ένα μοντέλο επιβλεπόμενης μάθησης. Δεδομένου ενός συνόλου χαρακτηριστικών προσπαθούν να εντοπίσουν το βέλτιστο υπερεπίπεδο με στόχο να αποδώσουν σωστά ετικέτες και να ταξινομήσουν τα χαρακτηριστικά σε κατηγορίες. Έστω ένα δυαδικό πρόβλημα ταξινόμησης με αντικείμενα που ανήκουν σε μία από τις δύο διαφορετικές κλάσεις $\{x_i, y_i\}$, όπου x_i οι διανυσματικές τιμές του συνόλου προς ταξινόμηση και $y_i = \{-1{,}1\}$ η ετικέτα που δέχεται κάθε τιμή και καθορίζει την κατηγορία που ανήκει. Σκοπός είναι η εύρεση ενός υπερεπιπέδου που διαχωρίζει το χώρο των διανυσμάτων x_i με τέτοιο τρόπο, ώστε τα διανύσματα της ίδιας κλάσης να περιέχονται στην ίδια πλευρά. Η βασική εξίσωση που περιγράφει το υπερπίπεδο έχει τη μορφή:

$$W \cdot x + b = 0 \tag{3.8}$$

όπου w το διάνυσμα που περιέχει τα βάρη που καθορίζουν την κλίση και b το κατώφλι για την μετακίνηση του επιπέδου.

Οι τιμές που βρίσκονται πιο κοντά στο υπερ-επίπεδο αποτελούν τα διανύσματα στήριξης. Για κάθε επίπεδο που ορίζει η εκάστοτε κλάση περιγράφεται ως:

$$W \cdot x + b = +1 \tag{3.9}$$

Για την πρόβλεψη της κατηγορίας κάθε νέου στοιχείου x που εισάγεται στο σύνολο, με δεδομένο το w λαμβάνουμε υπόψη μία συνάρτηση κόστους με τη μορφή:

$$\mathrm{E}(w) = \frac{\lambda}{2} \| w \|^2 + \frac{1}{n} \sum_{i=1}^{n} \max \left\{ 0.1 - y_i (\langle w, x \rangle) \right\} \tag{3.10}$$

3.5.2. Fisher Linear Discriminant Analysis (FLDA)

Η κύρια ιδέα της Fisher Linear Discriminant Analysis (FLDA) [30] είναι η αναζήτηση μιας γραμμής προβολής που χωρίζει τα αντικείμενα σε κατηγορίες. Αυτή η τεχνική αναζητά κατευθύνσεις στα δεδομένα που έχουν μεγαλύτερη διακύμανση και στη συνέχεια προβάλλει τα δεδομένα σε αυτό. Με αυτόν τον τρόπο παράγεται μια χαμηλότερη διάσταση αναπαράστασης δεδομένων που αφαιρεί ορισμένες από τις κατευθύνσεις που παρουσιάζουν θόρυβο. Υπάρχουν πολλά δύσκολα ζητήματα σχετικά με το ποιες κατευθύνσεις πρέπει κάποιος να επιλέξει. Η ανάλυση αυτή βασίζεται στην PCA [31], η οποία αποτελεί μια μια τεχνική μείωσης των διαστάσεων (dimensionality reduction) χωρίς επίβλεψη και ως εκ τούτου δε περιλαμβάνει πληροφορίες ετικέτας των δεδομένων. Για παράδειγμα, έστω ότι υπάρχουν 2 συστάδες σαν δοκοί σε 2 διαστάσεις, μια δοκός έχει y = 1 και η άλλη y = -1. Οι δοκοί τοποθετούνται παράλληλα και πολύ κοντά μεταξύ τους, όσο είναι η διακύμανση του συνολικού όγκου των δεδομένων, αγνοώντας τις ετικέτες. Για τη ταξινόμηση, αυτό θα ήταν μια πολύ κακή προβολή, διότι όλες οι ετικέτες μοιάζουν ομοιόμορφες και καταστρέφουμε έτσι τις χρήσιμες πληροφορίες. Μια πολύ πιο χρήσιμη προβολή είναι η ορθογώνιατ, δηλ. προς την κατεύθυνση της ελάχιστης συνολικής διακύμανσης, η οποία θα διαχωρίζει τέλεια το τα δεδομένα. Προκειμένου να γίνει μια βέλτιστη απεικόνιση των χαρακτηριστικών τα οποία έχουν εξαχθεί και να μην υπάρχει καταστροφή κανενός είδους πληροφορίας, η Fisher-LDA εξετάζει τη μεγιστοποίηση του ακόλουθου στόχου:

$$J(w) = \frac{w^T S_B w}{w^T S_w w} \tag{3.11}$$

όπου S_B είναι ο «μεταξύ των τάξεων» scatter matrix και S_W είναι ο «εντός των κλάσεων» scatter matrix. Σημειώνεται ότι λόγω του γεγονότος ότι οι μήτρες σκέδασης είναι ανάλογες με τη συν-διακύμανση, μπορούν να οριστούν πίνακες συν-διακύμανσης (matrices) J προκειμένου να υπάρξει σταθερότητα της αναλογικότητας. Οι ορισμοί των πινάκων σκέδασης στην περίπτωση αυτή είναι:

$$S_B = \sum_c (\mu_c - \bar{x})(\mu_c - \bar{x})^T \tag{3.12}$$

$$S_W = \sum_c \sum_{i \in c} (x_i - \bar{\mu}_c)(x_i - \bar{\mu}_c)^T \tag{3.13}$$

Στη συνέχεια κάθε δείγμα προβάλλεται σε διαχωριστική γραμμή με τη διατύπωση που εισάγεται σαν μετρική, ώστε να μπορέσει να γίνει η σύγκρισή τους.

3.5.3. Μοντέλα Μίξης Κανονικών Κατανομών (Gaussian Mixture Models)

Η Μίξη Κανονικών Κατανομών (για συντομία GMM) [73] χρησιμοποιείται κατά κόρον σε τεχνικές του Person Re-identification. Σε αρκετές περιπτώσεις η κανονική κατανομή παρουσιάζει περιορισμούς λόγω της αδυναμίας σωστής ομαδοποίησης δεδομένων. Για να ξεπεραστεί αυτή η δυσκολία, χρησιμοποιείται το άθροισμα με βαρή K συνιστωσών που περιγράφονται στην παρακάτω εξίσωση:

$$p(x \mid \Theta) = \sum_{k=1}^{k} w_k g(x \mid \mu_\kappa, \Sigma_i) \tag{3.14}$$

Στην εξίσωση αυτή το x είναι ένα διάνυσμα δεδομένων συνεχόμενων τιμών και D-διαστάσεων. Ως w_k, με k = 1, … , Κ συμβολίζονται τα βάρη της μίξης $g(x|\mu_k, \Sigma_i)$ όπου i = 1,…,K είναι οι πυκνότητες πιθανότητες των συνιστωσών κανονικής κατανομής. Το μ_k είναι το διάνυσμα μέσης τιμής και Σ_k ο πίνακας συν-διασποράς (covariance matrix) [7].

Η εξίσωση αυτή, με τις παραμέτρους που τη χαρακτηρίζουν (w_k, μ_k, Σ_k) οι οποίες συμβολίζονται με Θ, χρησιμοποιείται ως στατιστικό μοντέλο το οποίο ομαδοποιεί τα δεδομένα. Κάθε ομάδα δεδομένων αναπαριστάται από μια κανονική κατανομή που χαρακτηρίζεται από τις παραμέτρους της. Σκοπός μας είναι η βελτιστοποίηση της ομοιότητας του στατιστικού μοντέλου με τα πραγματικά δεδομένα. Για τη καλύτερη κατανόηση των GMM χρήσιμη θα ήταν η υπενθύμιση της κανονικής ή Gaussian συνάρτησης πυκνότητας μιας μεταβλητής:

$$p(\vec{x}) = \frac{1}{\sqrt{2\pi\sigma}} e^{[1/2(\vec{x}-\vec{\mu}/\sigma)2} \tag{3.15}$$

όπου μ είναι η μέση τιμή των τιμών του x, σ η τυπική απόκλιση και σ^2 η διασπορά των τιμών του x σε σχέση με τη μέση τιμή μ. Η εξίσωση αυτή στη γενική της μορφή για πολλές μεταβλητές είναι:

$$p(\vec{x}) = \frac{1}{(2\pi)^{d/2}\sqrt{|\Sigma|}} e^{[-1/2(\vec{x}-\vec{\mu})^{T}\Sigma^{-1}(\vec{x}-\vec{\mu})]} \tag{3.16}$$

Για να προσδιοριστούν οι παράμετροι της GMM ώστε να επιτευχθεί η βέλτιστη ομαδοποίηση των δεδομένων κάνουμε χρήση του αλγορίθμου EM(Expectation-Maximization) που βασίζεται στο θεώρημα Bayes:

$$P(\omega_i \mid x) = \frac{P(\omega_i)\,p(x \mid \omega_i)}{p(x)} = \frac{P(\omega_i)\,p(x \mid \omega_i)}{\Sigma_{\omega_i}\,p(x \mid \omega_i)P(\omega_i)} \tag{3.17}$$

Όπου $P(\omega_i)$ είναι η αρχική πιθανότητα (η πιθανότητα του μοντέλου πριν λάβουμε υπόψιν τα δεδομένα) [9], $p(x|\omega_i)$ η πιθανοφάνεια, $P(\omega_i|x)$ η αναθεωρημένη πιθανότητα (η πιθανότητα του μοντέλου ω_i αφού έχουμε λάβει υπόψιν τα δεδομένα x) και $p(x)$ η περιθωριακή πιθανότητα (marginal probability) των δεδομένων, δηλαδή η πιθανότητα των δεδομένων που είναι ανεξάρτητη από κάποια υπόθεση ή αλλιώς το άθροισμα όλων των συνδυασμένων πιθανοτήτων (joint probabilites).

ΚΕΦΑΛΑΙΟ 4. ΠΑΡΟΥΣΙΑΣΗ ΤΟΥ PERSON RE-IDENTIFICATION

4.1. Σκοπός και Ιστορικά Στοιχεία

Η επαναταυτοποίηση ατόμου (Person Re-identification) αποτελεί μια από τις βασικές διεργασίες ενός συστήματος αυτοματοποιημένης βιντεο-επιτήρησης (video surveillance) και έχει καταστεί αντικείμενο έντονης έρευνας τα τελευταία χρόνια. Έχοντας ως δεδομένο μια εικόνα/βίντεο ενός ατόμου που έχει ληφθεί από μια κάμερα παρακολούθησης, η επαναταυτοποίηση είναι η διαδικασία αναγνώρισης του ίδιου ατόμου από εικόνες/βίντεο που έχουν ληφθεί από διαφορετικές κάμερες ή οπτικές γωνίες. Εκτός από την βιντεο-επιτήρηση (video surveillance) η επαναταυτοποίηση ατόμου (Person Re-identification) έχει επιπλέον εφαρμογές σε διάφορους τομείς όπως τη ρομποτική, τα συστήματα πολυμέσων, την εγκληματολογική έρευνα κ.α

Η επαναταυτοποίηση ατόμου (Person Re-identification) αποτελεί μια αρκετά απαιτητική διαδιακασία λόγω της οπτικής ασάφειας και της χωροχρονικής αβεβαιότητας που παρουσιάζονται κατά τη λήψη των εικόνων/βίντεο από τις κάμερες παρακολούθησης αλλά και των πολλαπλών χαρακτηριστικών τα οποία παρουσιάζονται κατά την επεξεργασία αυτών όπως είναι οι παραλλαγές φωτισμού, θέσης και θέασης του ατόμου. Τα στοιχεία αυτά συχνά συνδυάζονται με εικόνες/βίντεο χαμηλής ανάλυσης και έντονο θόρυβο κάνωντας τη διαδικασία της επαναταυτοποίησης ατόμων ακόμη πιο δύσκολη.

Οι χωρικές ή χρονικές συνθήκες για τον περιορισμό του προβλήματος είναι δύσκολο να καταγραφούν. Το καθήκον του προσδιορισμού ενός αντικειμένου που εμφανίζεται στο οπτικό πεδίο μιας κάμερας και της αναγνώρισης του ίδιου αντικειμένου ξανά στην ίδια ή άλλη κάμερα ονομάζεται «επαναταυτοποίηση αντικειμένου» [41]. Η έρευνα σχετικά με την επανεξέταση και ανίχνευση ατόμων (Person Re-identification) ξεκίνησε με τη παρακολούθηση πολλαπλών φωτογραφικών μηχανών [32] και την ανίχνευση των ατόμων. Από τότε έχουν αναπτυχθεί αρκετές σημαντικές τεχνικές και αλγόριθμοι. Το 1997, οι Huang και Russell [33] πρότειναν έναν Bayesian αλγόριθμο - τρόπο για την εκτίμηση της πρόβλεψης και την εμφάνιση των αντικειμένων σε μια κάμερα παρατήρησης προκειμένου να υπάρξουν δεδομένες αποδείξεις. Το μοντέλο εμφάνισης αυτό περιλαμβάνει πολλαπλά χωρο-χρονικά χαρακτηριστικά όπως το χρώμα, το μήκος, το ύψος και το πλάτος του οχήματος το οποίο τέθηκε προς παρατήρηση, η ταχύτητα και ο χρόνος παρατήρησης.

Το 2005 δόθηκε από τους Wojciech Zajdel, Zoran Zivkovic και Ben J. A. Krise από το Πανεπιστήμιο του Άμστερνταμ [38] ο όρος «επαναπροσδιορισμός/ επαναταυτοποίηση ατόμου» (Person Re-identification). Οι Zajedel et al. αποσκοπούν σε έρευνά τους στο να «επαναπροσδιορίσουν» ένα άτομο όταν φεύγει από το οπτικό πεδίο και εισέρχεται ξανά κάποια άλλη χρονική στιγμή. Στη μέθοδο τους αυτή, μια μοναδική λανθάνουσα ετικέτα λαμβάνεται υπόψιν για κάθε πρόσωπο και ένα δυναμικό Bayesian δίκτυο όπου παράγεται και κωδικοποιείται η πιθανωτική σχέση μεταξύ των ετικετών και των χαρακτηριστικών που εξάγονται. Η ταυτότητα ενός εισερχόμενου προσώπου καθορίζεται από τους διανομές ετικετών οι οποίοι υπολογίζονται κατά προσέγγιση.

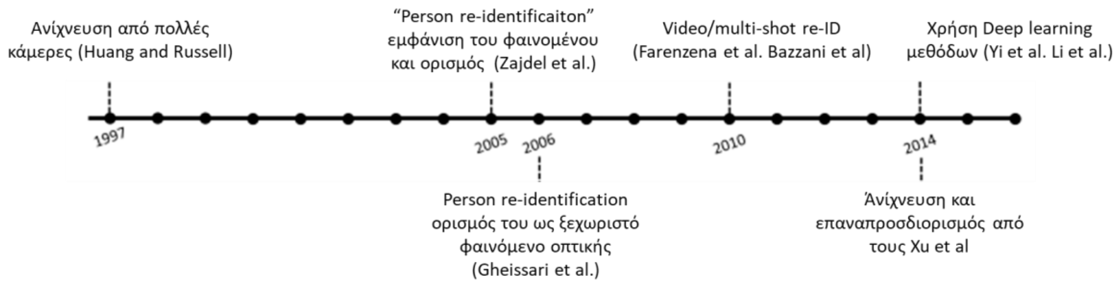
Ένα χρόνο αργότερα το 2006, οι Gheissari et al. [81] ασχολήθηκαν με την επαναταυτοποίηση του ατόμου με την εφαρμογή αλγορίθμου χωρικής και χρονικής κατάτμησης εικόνων, για την ανίχνευση νέων στοιχείων. Η οπτική αντιστοίχιση με βάση το χρώμα και τα ιστογράμματα edgel πραγματοποιούνται είτε από ένα αρθρωτό μοντέλο πεζών ή τη επισήμανση του σημείου Hessian-Affine. Τα πειράματα που παρουσιάζονται στην παραπάνω έρευνα, διεξήχθησαν σε ένα σύνολο δεδομένων με 44 άτομα. Οι κάμερες που χρησιμοποιήθηκαν ήταν 3 με επικάλυψη οπτικού πεδίου. Σημειώνεται ότι αν και οι Gheissari et al. σχεδίασαν μια χωρική-χρονική μέθοδο τμηματοποίησης χρησιμοποιώντας εικόνες οι οποίες είχαν ληφθεί από βίντεο, οι πληροφορίες οι οποίες λαμβάνονται δεν αφορούν ούτε την εξαγωγή χαρακτηριστικών, ούτε τις διαδικασίες ταιριάσματος (matching). Για το λόγο αυτό η μέθοδος αυτή ανήκει στην κατηγορία των αλγορίθμων που βασίζονται στην επεξεργασία εικόνων. Η εργασία αυτή είναι άξια αναφοράς και αποτέλεσε σημαντικό στοιχείο για την επαναταυτοποίηση του ατόμου και την αρχή της ως ανεξάρτητη εργασία μηχανικής όρασης.

Αρχικά οι περισσότεροι αλγόριθμοι οι οποίοι αναπτύχθηκαν και παρουσιάστηκαν είχαν ως στόχο την χρήση σε βίντεο, ωστόσο από το 2008 και έπειτα τα περισσότερα έργα εστιάζουν στο ταίριασμα εικόνων. Το 2010, προτάθηκαν βιβλιογραφικά δύο έργα [82], [83] τα οποία παρουσιάζουν πολλαπλούς σκοπούς του re-ID, στα οποία τα πλαίσια είναι τυχαία επιλεγμένα. Το χρώμα είναι ένα κοινό χαρακτηριστικό που χρησιμοποιείται και στα δύο έργα, και οι Farenzena et al. [83] χρησιμοποιούν επιπλέον ένα μοντέλο τμηματοποίησης για την ανίχνευση του ατόμου σε πρώτο πλάνο. Για τη μέτρηση της απόστασης, και τα δύο έργα υπολογίζουν την ελάχιστη απόσταση μεταξύ των πλαισίων οριοθέτησης σε δύο σύνολα εικόνων. Οι Bazzani et al. χρησιμοποίησαν επίσης την απόσταση Bhattacharyya για τα χαρακτηριστικά χρώματος. Από το στοιχείο αυτό φαίνεται ότι η χρήση πολλαπλών καρέ ανά άτομο βελτιώνει αποτελεσματικά την εκδοχή λήψης χαρακτηριστικών ενός πλαισίου [82], [83] και την ακρίβεια επαναταυτοποίησης του ατόμου.

Η αύξηση των αλγορίθμων που χρησιμοποιούν deep learning τεχνικές στην ταξινόμηση εικόνων [84] εξαπλώνεται και από το 2014 ξεκινά να εφαρμόζεται και στο Person Re-identification, όταν οι Yi et al. [85] και Li et al. [86] και οι δύο χρησιμοποιούν ένα νευρωνικό δίκτυο [87] για να προσδιορίσουν αν ένα ζευγάρι εικόνων εισόδου ανήκουν στο ίδιο άτομο. Ο λόγος για την επιλογή του τυποποιημένου μοντέλου είναι ότι ο αριθμός των δειγμάτων μάθησης για κάθε ταυτότητα είναι περιορισμένος (συνήθως δύο). Εκτός από ορισμένες παραλλαγές των ρυθμίσεων παραμέτρων, οι κύριες διαφορές είναι ότι η εργασία [85] προσθέτει μια συνάρτηση κόστους στο δίκτυο, ενώ η εργασία [86] χρησιμοποιεί ένα διαφορετικό διαχωρισμό σώματος. Τα πειραματικά σύνολα δεδομένων δεν επικαλύπτονται στις δύο αυτές εργασίες, επομένως οι δύο μέθοδοι δεν είναι άμεσα συγκρίσιμες. Παρόλο που η απόδοσή τους δεν είναι σταθερή ακόμη και στα μικρά σύνολα δεδομένων, οι μέθοδοι deep learning έχουν γίνει από τότε μια δημοφιλής επιλογή για τη χρήση αυτή.

Ενώ η πλειοψηφία των έργων που χρησιμοποιούν πλαίσια οριοθέτησης (b-boxes) στα πειράματά τους, που παράγονται από σταθερό ανιχνευτή, είναι απαραίτητο να μελετηθεί χωριστά και η επίδραση των ανιχνευτών πεζών στην ακρίβεια ταυτοποίησης. Το 2014, οι Xu et al. [88] ασχολήθηκαν με το θέμα αυτό, συνδυάζοντας τις βαθμολογίες ανίχνευσης πεζών και μεθόδων επαναταυτοποίησης ατόμων. Στην εργασία αυτή παρουσιάζεται το σύνολο δεδομένων CAMPUS, εξετάζοντας από κοινού την ανίχνευση πεζών και την χρήση της μεθόδου επαναταυτοποίησης, παράγοντας υψηλότερα ποσοστά στην ακρίβεια ανάκτησης ατόμου από τη χρήση τους ξεχωριστά.

Στην εικόνα που ακολουθεί παρουσιάζονται χρονολογικά οι κυριότερες περίοδοι του φαινομένου, καθώς και η εξέλιξή τους χρονολογικά.



Σχήμα 4-1 Χρονολογικό διάγραμμα ανάπτυξης του φαινομένου Re-id.

4.2. Διαδικασία Ροής του Person Re-ID (Pipelining)

Αρκετοί ερευνητές ασχολήθηκαν με την ανάλυση βίντεο από πολλαπλές κάμερες οι οποίες λαμβάνουν εικόνα από διαφορετικά σημεία. Τα δίκτυα κλειστής τηλεόρασης (Closed circuit TV – CCTV) αντιμετωπίζουν πολλές προκλήσεις όπως η περιορισμένη δυνατότητα παρακολούθησης που οδηγεί στην έλλειψη προσοχής σε σημαντικά γεγονότα και στόχους συμπεριλαμβανομένης και της υπερφόρτωσης δεδομένων από το μεγάλο πλήθος μηχανών παρατήρησης. Κατά συνέπεια, υπάρχει μια ξεχωριστή ανάγκη για μια τεχνολογία η οποία θα λογίζει την ανακούφιση του βάρους των ανθρώπινων πόρων και θα έχει ως στόχο την αύξηση των περιορισμένων ανθρώπινων δυνατοτήτων.

Ο ανανεωμένος μηχανισμός του Person Re-identification λαμβάνει ως είσοδο είτε κομμάτια είτε οριοθετημένα πλαίσια εικόνων που περιέχουν διαχωρισμένες εικόνες μεμονωμένων ατόμων, όπως αυτές δημιουργούνται από ένα σύστημα οπτικής παρακολούθησης. Για την αυτόματη ανίχνευση - ταυτοποίηση ανθρώπων σε διαφορετικές τοποθεσίες με την πάροδο του χρόνου που λαμβάνονται από διαφορετικές κάμερας παρακολούθησης, η διαδικασία που ακολουθείται συνήθως περιλαμβάνει τα ακόλουθα 5 βήματα:

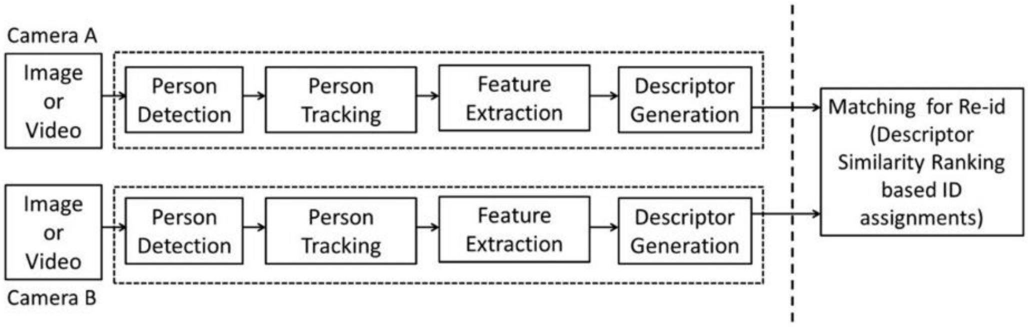
1.Ανίχνευση ατόμου (Person Detection)

2.Παρακολούθηση ατόμου με χρήση ανιχνευτών πεζών (Person Tracking)

3.Εξαγωγή χαρακτηριστικών (Feature Extraction)

4.Δημιουργία περιγραφέων (Descriptor Generation) για την περιγραφή των χαρακτηριστικών ενός ατόμου.

5.Φάση ταιριάσματος (matching) των περιγραφέων ατόμων με τη χρήση κάποιας μετρικής ομοιότητα μεταξύ των εικόνων.



Σχήμα 4-2 Pipeline του person re-identification

Αυτά τα βήματα επεξεργασίας δημιουργούν ορισμένες απαιτήσεις στον αλγόριθμο και στο σχεδιασμό του συστήματος. Αυτό οδήγησε τόσο στην ανάπτυξη νέων όσο και στην εκμετάλλευση των υπαρχόντων τεχνικών αντιμετώπισης των προβλημάτων της αναπαράστασης χαρακτηριστικών, του μοντέλου ταιριάσματος και της δημιουργίας πλαισίου. Οι σύγχρονες προσεγγίσεις στο re-identification συνήθως εκμεταλλεύονται τη χαμηλή στάθμη χαρακτηριστικών όπως χρώμα [34], υφή, χωρική δομή [35] ή συνδυασμό αυτών [36]. Αυτό οφείλεται στο γεγονός ότι αυτά τα χαρακτηριστικά μπορούν να μετρηθούν σχετικά εύκολα και αξιόπιστα, και παρέχουν ένα λογικό επίπεδο διακρίσεων μεταξύ των ατόμων. Επίσης μπορούν να κωδικοποιηθούν περαιτέρω με τη χρήση σταθερού μήκους περιγραφέων, όπως π.χ. με τη μορφή ιστογραμμάτων [36] και covariances [37]. Σημαντικά στοιχεία του person re-identification αποτελούν τα ακόλουθα:

- Καταστάσεις κάμερας: Οι καταστάσεις στις οποίες μπορεί να βρεθεί ένα σύστημα παρακολούθησης είναι: 1)χρήση απλής κάμερας, 2)κάμερες με ίδιο οπτικό πεδίο, 3)κάμερες με χρήση βαθμονόμησης, 4) κάμερες χωρίς της χρήση βαθμονόμησης, 5)κάμερες με αλληλοεπικαλυπτόμενα πεδία 6)κάμερες με μη-αλληλοεπικαλυπτόμενα πεδία.
- Εικόνες λήψης: Οι μέθοδοι επαναταυτοποίησης μπορούν να χωριστούν σε μεθόδους που χρησιμοποιούν εικόνες απλής λήψης είτε σε εικόνες πολλαπλής λήψης. Οι μέθοδοι πολλαπλής λήψης περιέχουν περισσότερες πληροφορίες για ένα άτομο από αυτές της μονής λήψης πετυχαίνοντας καλύτερα αποτελέσματα αλλά αυξάνοντας παράλληλα και τη χρήση των υπολογισμών.
- Σύνολα χαρακτηριστικών - Υπογραφή εικόνας (signature): Οι μέθοδοι ένα συγκεκριμένο σύνολο χαρακτηριστικών από μία εικόνα. Κάθε εικόνα περιγράφεται από ένα τέτοιο σύνολο χαρακτηριστικών το οποίο αποτελεί και την υπογραφή της.
- Ανάλυση του σώματος: Η ανάλυση του σώματος ενός ανθρώπου ή και ενός αντικειμένου δίνει τη δυνατότητα της πιο γρήγορης και σωστής εξαγωγής χαρακτηριστικών τα οποία χρησιμοποιούνται από τις μεθόδους. Η εκ των προτέρων γνώση που έχουμε σχετικά με της δομή του ανθρώπινου σώματος και το πως αυτό εμφανίζεται σε εικόνες μας παρέχει χρήσιμες πληροφορίες για τη σωστή διευθέτηση των χαρακτηριστικών του αντικειμένου μέσα σε μια εικόνα.
- Μοντέλο Μάθησης: Οι περισσότερες σύγχρονες μέθοδοι επαναταυτοποίησης κάνουν χρήση κάποιου αλγορίθμου μηχανικής μάθησης είτε για τη καλύτερη περιγραφή μιας εικόνας είτε για την εκμάθηση κάποιας μετρικής μάθησης.
- Σενάριο: Το σενάριο ορίζει τον τόπο, το χρόνο, το χώρο και το σκοπό της επαναταυτοποίησης ενός ατόμου. Ανάλογα με τα δεδομένα τα οποία υπάρχουν κάθε μέθοδος, χρησιμοποιεί τις κατάλληλες τεχνικές και αλγορίθμους.

Μόλις επιτευχθεί η κατάλληλη αντιπροσώπευση, ο πλησιέστερος γείτονας [35] ή οι αλγόριθμοι αντιστοίχισης που βασίζονται σε μοντέλο, όπως η κατάταξη διανυσμάτων υποστήριξης [36] γίνεται το ταίριασμα μεταξύ των ατόμων. Σε κάθε περίπτωση, μια μετρική απόστασης (π.χ. Euclidean) πρέπει να επιλεγεί για να μετρηθεί η ομοιότητα μεταξύ δύο δειγμάτων οι οποίες μπορεί να είναι βελτιστοποιημένες διακριτικά για να μεγιστοποιήσει την απόδοση επαναταυτοποίησης. Γεφυρώνοντας αυτά τα δύο στάδια, μερικές μελέτες επιχείρησαν επίσης να εξάγουν και να εκπαιδεύσουν τα συστήματά τους με διακριτικά χαρακτηριστικά χαμηλού επιπέδου απευθείας από τα δεδομένα.

Οι διαφορετικές προσεγγίσεις χρησιμοποιούν ελαφρώς διαφορετικές ταξινομήσεις στην κατηγοριοποίηση των υπάρχουσων μεθόδων re-identification.

Γενικά, μόνο όταν ένα ζεύγος εικόνων ταιριάζει, η μέθοδος αυτή του person re-identification θεωρείται ως μέθοδος επαναταυτοποίησης ξεχωριστών δειγμάτων. Εάν η αντιστοίχιση διεξάγεται μεταξύ δύο ομάδων εικόνων, π.χ. πλαίσια που λαμβάνονται από δύο ξεχωριστές κάμερες, η μέθοδος είναι γνωστή ως προσέγγιση αναγνώρισης πολλαπλών σημείων. Μια προσέγγιση κατηγοριοποιείται ως ελεγχόμενη μέθοδος αν είναι πριν από την εφαρμογή και εκμεταλλεύεται τα δείγματα που έχουν επισημανθεί για τη ρύθμιση των παραμέτρων του μοντέλου ως μετρήσεις απόστασης, το βάρος των χαρακτηριστικών ή τα όρια απόφασης. Διαφορετικά, θεωρείται ως μια μη-ελεγχόμενη προσέγγιση εάν αφορά την εξαγωγή ισχυρών οπτικών χαρακτηριστικών και δεν βασίζεται σε δεδομένα εκπαίδευσης. Η διάκριση αυτών των ορίων αφορά τις μεθόδους που μαθαίνουν κατά την εκπαίδευση των δεδομένων αλλά δεν βασίζονται στο σχολιασμό των δεδομένων.

4.3. Διαχωρισμός Μεθόδων Επαναταυτοποίησης

Υπάρχουν σημαντικές κατηγορίες και τύποι μεθοδολογιών οι οποίες σχετίζονται με το Person Re-identification. Το κομμάτι αυτό παρουσιάζει τον διαχωρισμό των μεθόδων αυτών για να μπορέσει ο αναγνώστης να κατανοήσει καλύτερα την παρουσίαση των μεθόδων που ακολουθεί. Ο διαχωρισμός γίνεται σε αυτές που χρησιμοποιούν πληροφορίες και παραμέτρους της υπόστασης του συστήματος και σε αυτές που χρησιμοποιούν το περιεχόμενο της κάθε εικόνας για την αντιστοίχιση ατόμων σε σκηνές από διαφορετικές κάμερες:

- Μέθοδοι συνάφειας

Αυτές οι μέθοδοι βασίζονται σε εξωτερικές πληροφορίες για τα συμφραζόμενα είτε για την αντιστοίχιση κλάδων ή την εξαγωγή χαρακτηριστικών για το Re-ID. Σε αυτή τη κατηγορία οι μέθοδοι χωρίζονται αντίστοιχα σε αυτές που χρησιμοποιούν πληροφορίες 1) γεωμετρικών παραμέτρων της κάμερας ή σε αυτές που ενσωματώνουν και αξιοποιούν τις 2) παραμέτρους βαθμονόμησης της κάθε κάμερας

1) Γεωμετρία κάμερας

Οι πρώτες μέθοδοι πάνω στο Person Re-id εστίαζαν κυρίως στις έννοιες των χωρικών και χρονικών σχέσεων μεταξύ των καμερών για τη μείωση των σφαλμάτων επαναταυτοποίησης, περιορίζοντας το μέγεθος του σετ των εικόνων. Ένα παράδειγμα είναι η χρήση ενός παραμετρικού πλέγματος επιφάνειας ως χάρτη που προβάλλεται σε όλες τις κάμερες και όπου το αντικείμενο είναι ορατό επιτρέποντας τη μοντελοποίηση των διαδρομών ενός ατόμου όταν περνάει μέσα από το δίκτυο από κάμερες. Οι χάρτες από διαφορετικά κομμάτια συγκρίνονται χρησιμοποιώντας το σταθμισμένο άθροισμα της μέτρησης τετραγωνικών αποστάσεων. Ωστόσο, για να δημιουργηθεί η υπογραφή της εμφάνισης το αντικείμενο πρέπει να είναι ορατό σε τουλάχιστον 3 κάμερες με αλληλεπικαλυπτόμενες προβολές ταυτόχρονα.

2) Βαθμονόμηση κάμερας

Η βαθμονόμηση της κάθε κάμερας αξιοποιείται για την εξαγωγή μοναδικών και διακριτικών χαρακτηριστικών για την αύξηση των οπτικών περιγραφικών στοιχείων που χρησιμοποιούνται για το Person Re-id. Σε ένα σύστημα με αλληλοεπικαλυπτόμενα πεδία όρασης η βαθμονόμηση της κάμερας επιτρέπει την ανάκτηση των συντεταγμένων του ατόμου στο τρισδιάστατο χώρο και άρα την αποδοτικότερη χρήση της πληροφορίας που προσφέρει ένα τέτοιο δίκτυο για την επαναταυτοποίηση των ατόμων. Το βασικό μειονέκτημα αυτών των μεθόδων είναι πως σε ένα μεγάλο δίκτυο από κάμερες, η βαθμονόμηση όλων των καμερών δεν είναι πάντα εφικτή. Πρόκειται για μια ερευνητική περιοχή χωρίς έντονη δραστηριότητα τα τελευταία χρόνια

- Μέθοδοι περιεχομένου

Έχουν προταθεί αρκετές προσεγγίσεις που βασίζονται εξ ολοκλήρου στην ανάλυση των οπτικών περιγραφικών δεικτών και δεν ενσωματώνουν καθόλου εξωτερικές πληροφορίες συμφραζομένων για την υποβοήθηση της διαδικασίας. Οι μέθοδοι που ασχολούνται με το περιεχόμενο της εικόνας διακρίνονται σε 1)ενεργές και 2)παθητικές. Οι περισσότερες από τις πρόσφατες έρευνες επικεντρώνονται σε μη-συμφραζόμενες μεθόδους. Μια δημοφιλής ταξινόμηση μέσα σε αυτή την τάξη βασίζεται στο αν χρησιμοποιείται μία εικόνα (single shot) ή πολλαπλές εικόνες (multi-shot) για να δημιουργηθούν και να συγκριθούν οι περιγραφείς εμφάνισης.

1) Ενεργές

Αυτές οι μέθοδοι ορίζονται ως ενεργές καθώς εκμεταλλεύονται το περιεχόμενο των εικόνων κάνοντας χρήση εποπτευόμενων ή μη τεχνικών εκμάθησης για την εξαγωγή ή το ταίριασμα των περιγραφέων. Οι μέθοδοι αυτοί και αυτές με τη σειρά τους διαιρούνται σε τρεις κατηγορίες, τη i) βαθμονόμησης χρώματος την ii) εκμάθηση περιγραφέα και τη iii) εκμάθηση μετρική απόστασης :

i) Βαθμονόμηση Χρώματος

Στη κατηγορία αυτή ανήκουν οι μέθοδοι που προσπαθούν να μοντελοποιήσουν τη σχέση της μίας κάμερας με την άλλη σε χρωματικό επίπεδο. Για να γίνει αυτό χρειάζονται ένα στάδιο εκμάθησης για να μπορούν να αναπτύξουν το μοντέλο βαθμονόμησης που πρέπει να ενημερώνεται συχνά για να λαμβάνει υπόψη όλες τις επιθυμητές σχέσεις. Αυτό μπορεί να γίνει με τη χρήση συναρτήσεων μεταφοράς της φωτεινότητας της εικόνας (Brightness Trunsfer Function, BTF) μεταξύ κάθε ζεύγους φωτογραφικών μηχανών και αποκτάται από τα δεδομένα εκπαίδευσης. Μόλις γίνει γνωστή μια τέτοια αντιστοίχιση μεταξύ των φωτογραφικών μηχανών, το πρόβλημα του Person Re-ID μειώνεται σε ένα από τα αντίστοιχα μοντέλα μετασχηματισμένης εμφάνισης (transformed appearance models). Ωστόσο, μια τέτοια χαρτογράφηση δεν είναι μοναδική και αλλάζει από το πλαίσιο στο πλαίσιο ανάλογα με διάφορους παράγοντες όπως ο φωτισμός, η γεωμετρία της σκηνής, το εστιακό μήκος, ο χρόνος έκθεσης και το άνοιγμα κάθε κάμερας. Επομένως, ένα ενιαίο BTF δεν μπορεί να χρησιμοποιηθεί για συνεπή αντιστοίχιση μοντέλων.

ii) Εκμάθηση Περιγραφέα

Σε αυτή την κατηγορία οι μέθοδοι προσπαθούν να εντοπίσουν τα διακριτικά χαρακτηριστικά ενός ατόμου ή να χρησιμοποιήσουν ένα στάδιο εκμάθησης για να δημιουργήσουν περιγραφικά λεξικά χαρακτηριστικών (bag-of-features) τα οποία θα βοηθήσουν έτσι ώστε να γίνει περισσότερο αντιπροσωπευτική η εμφάνιση ενός ατόμου και κατ'έκταση η υπογραφή (signature) του. Πρόκειται για μια κατηγορία με αρκετές εφαρμογές τα τελευταία χρόνια. Πολλές από αυτές τις μεθόδους χωρίζουν την ανθρώπινη σιλουέτα σε τμήματα χρησιμοποιώντας αλγόριθμο κατάλληλο για το σκοπό αυτό (shape context algorithm) που βασίζεται στο λεξικό σχήματος (shape dictionary) και που έχει μάθει εκ των προτέρων προσπαθώντας να ανακτήσει τα κατάλληλα χαρακτηριστικά. Αυτό το μοντέλο αντιστοίχισης χαρακτηριστικών σε σύνολα λέξεων ονομάζεται «μοντέλο απο σύνολα λέξεων» (Bag of Words model) ή

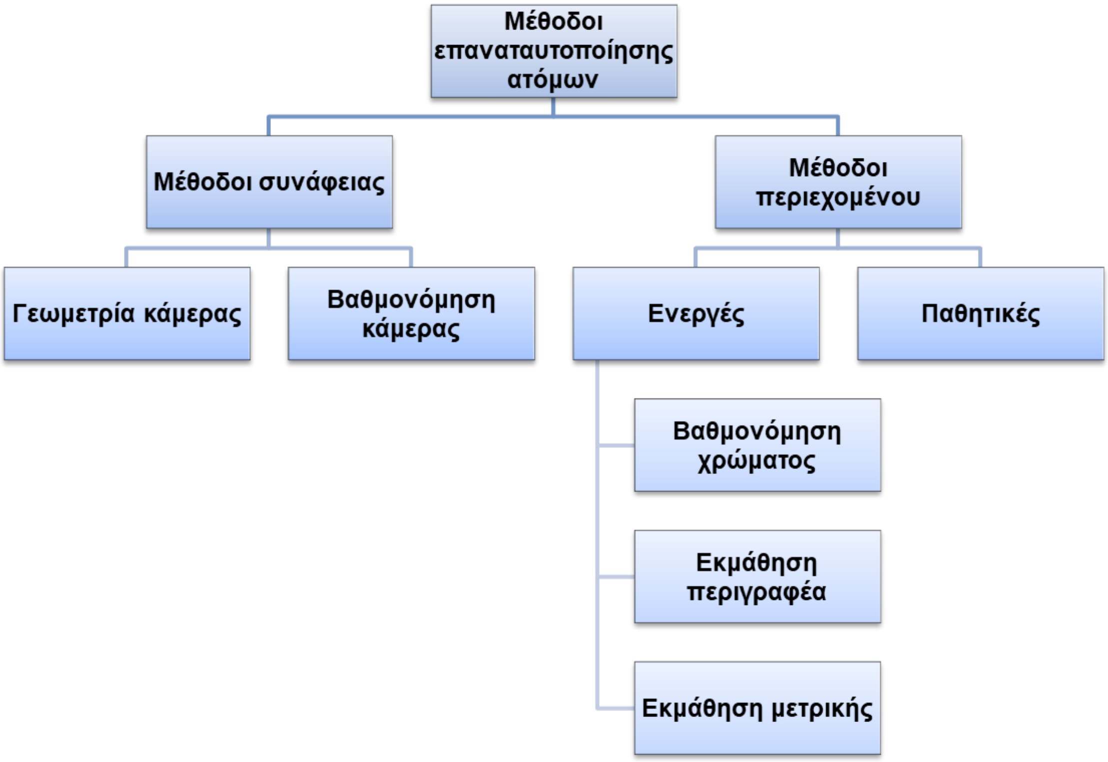
BoW [39]. Αυτός ο διαχωρισμός έχει οδηγήσει στη δημιουργία ενός ανώτερου επιπέδου χαρακτηριστικών των εικόνων, το επίπεδο των ιδιοτήτων (attributes). Τα attributes μπορούν να χαρακτηριστούν ως λέξεις-κλειδιά τα οποία μπορούν να περιγράφουν αντίκειμενα μια εικόνας όπως είναι π.χ το φύλο, η ενδυμασία, το χρώμα των μαλλιών ή κάποιου αξεσουάρ τα οποία είναι μοναδικά και που μπορούν να διακρίνουν ένα δεδομένο θέμα από ένα άλλο, ακόμη και στη περίπτωση που οοι συνολικές εμφανίσεις τους είναι αρκετά όμοιες.

iii) Εκμάθηση Μετρικής Απόστασης

Στη κατηγορία αυτή οι μέθοδοι μετατοπίζουν την εστίαση από τις προσπάθειες που βασίζονται στην επιλογή χαρακτηριστικών για τη βελτίωση του Person Re-ID με την εκμάθηση κατάλληλων μετρητών απόστασης που μπορούν να μεγιστοποιήσουν την ακρίβεια ταιριάσματος, ανεξάρτητα από την επιλογή της εμφάνισης της αναπαράστασης. Οι αλγόριθμοι αυτοί προσπαθούν να μάθουν μια μέτρηση στο χώρο που καθορίζεται από τα χαρακτηριστικά μιας εικόνας που διατηρώντας αυτά που προέρχονται από την ίδια τάξη πιο κοντά, ενώ τα χαρακτηριστικά από διαφορετικές κατηγορίες πιο απομακρυσμένα. Στο πλαίσιο του Person Re-ID, ο στόχος είναι η μάθηση μιας μετρικής απόστασης στον χώρο εμφάνησης που μεγιστοποιεί την απόσταση μεταξύ των περιγραφέων διαφορετικών ανθρώπων και να ελαχιστοποιεί την απόσταση για περιγραφείς του ίδιου προσώπου. Η μετρική μάθηση γίνεται με επίβλεψη και υπό περιορισμούς ανά ζεύγη. Τα χαρακτηριστικά εκπαίδευσης συνδυάζονται με περιγραφείς εμφάνισης και οι ετικέτες εκπαίδευσης είναι είτε θετικές είτε αρνητικές ανάλογα με το αν οι περιγραφείς εμφάνισης ανήκουν στο ίδιο άτομο ή σε διαφορετικό αντίστοιχα.

2) Παθητικές

Τέλος οι μέθοδοι αυτής της κατηγορίας ασχολούνται με το σχεδιασμό οπτικών περιγραφών για τον χαρακτηρισμό της εμφάνισης ενός ατόμου και τη σύγκρισή του με τον υπολογισμό των μέτρων ομοιότητας για την επίτευξη του Person Re-ID. Αυτές οι μέθοδοι ορίζονται ως παθητικές, καθώς δεν βασίζονται σε τεχνικές εποπτευόμενης ή μη μάθησης για την εξαγωγή και την αντιστοίχιση των περιγραφέων. Σε αυτή τη κατηγορία ανήκουν αρκετές γνωστές και σύγχρονες μέθοδοι όπως η μέθοδος SDALF η οποία κατατμεί την εικόνα ενός ατόμου σε περιοχές ανάλογα με τη δομή του σώματος π.χ κεφάλι, άνω μέρος του σώματος, κάτω μέρος σώματος δημιουργώντας έτσι μια υπογραφή, δηλ. ένα σύνολο από χαρακτηριστικά κάθε μέρους του. Βασικό πλεονέκτημα αυτών των μεθόδων είναι η χρήση τους χωρίς την ανάγκη κάποιας τεχνικής εκμάθησης.



Σχήμα 4-3 Διαχωρισμός μεθόδων επαναταυτοποίησης.

4.4. Μετρικές Αξιολόγησης (Evaluation metrics)

Πολλά πρότυπα σύνολα δεδομένων αναφοράς αντικατοπτρίζουν ένα σενάριο «κλειστού κόσμου», π.χ. ακριβώς όπως οι δύο προβολές κάμερας με μία ακριβώς εμφάνιση κάθε ατόμου ανά κάμερα και με 1:1 ακριβή αντιστοιχία ταυτότητας μεταξύ των καμερών. Αυτό έρχεται σε αντίθεση με ένα περισσότερο ρεαλιστικό σενάριο «ανοικτού κόσμου», όπου τα άτομα σε κάθε κάμερα μπορεί να έχουν εικόνα μόνο εν μέρει και να μην λαμβάνουν υπόψη την επικάλυψη και τον αριθμό των καμερών, το μέγεθος του περιβάλλοντος και τον αριθμό των ανθρώπων οι οποίοι μπορεί να είναι άγνωστοι και διάσπαρτοι σε πολύ μεγαλύτερη κλίμακα. Έτσι, ο χώρος αναζήτησης είναι ένα άγνωστο μέγεθος και περιέχει έναν δυνητικά απεριόριστο αριθμό υποψηφίων στόχων. Η μέθοδος αναγνώρισης και ανίχνευσης των στόχων σε τέτοια ανοικτά περιβάλλοντα μπορεί να κλιμακωθεί σε αυθαίρετα μεγάλα επίπεδα, που να καλύπτουν τεράστιες χωρικές περιοχές, όχι μόνο διαφορετικά κτίρια, αλλά και διαφορετικές πόλεις, χώρες ή ακόμη και ηπείρους, οδηγώντας σε ένα τεράστιο όγκο δεδομένων.

Υπάρχουν διάφορες μετρήσεις που είναι χρήσιμες για την ποσοτικοποίηση της αποτελεσματικότητας ενός συστήματος επαναταυτοποίησης. Οι δύο πιο κοινές μετρήσεις είναι η ακρίβεια κατάταξης-1 (rank-1 ή r-1) και η καμπύλη CMC. Κατά την αξιολόγηση των αλγορίθμων που χρησιμοποιούνται στο re-identification, η σωρευτική καμπύλη αντιστοίχισης χαρακτηριστικών (CMC) χρησιμοποιείται συνήθως και αντιπροσωπεύει την πιθανότητα εμφάνισης μιας ταυτότητας ερωτήματος σε διαφορετικά μεγέθη (N) και λίστες υποψηφίων και είναι ακριβής ως μέθοδος αξιολόγησης μόνο όταν υπάρχει μία επαρκής ποσότητα σύγκρισης Η μέτρηση αυτή είναι αποδεκτή, στην πράξη, όταν οι άνθρωποι ενδιαφέρονται περισσότερο για να επαληθεύσουν τα δεδομένα τα οποία έχουν εξαχθεί από την ανίχνευση.

Για την έρευνητική ακεραιότητα οι Zheng et al. [39] προτείνουν για αξιολόγηση τη χρήση της μέσης ακρίβειας (mAP) η οποία δείχνει ποιο ποσοστό των αποτελεσμάτων ανήκουν επιτυχώς στη κατηγορία. Το κίνητρο είναι ότι θα πρέπει να είναι σε θέση να δημιουργηθεί ένα τέλειο σύστημα επαναταυτοποίησης που να επιστρέφει όλα τα επιτυχή αποτελέσματα στον χρήστη. Η υπόθεση μπορεί να είναι ότι δύο συστήματα είναι εξίσου ικανά να εντοπίζουν την ταύτιση και την αναγνώριση των εικόνων, αλλά έχουν διαφορετική δυνατότητα ανάκτησης. Σε πολλά σενάρια η CMC δεν διαθέτει επαρκείς διακριτικές ικανότητες σε αντίθεση με τη mAP. Συνεπώς, εάν χρησιμοποιηθούν παράλληλα η mAP και η CMC, για ένα σύνολο δεδομένων όπου υπάρχουν πολλαπλές εικόνες από πολλές κάμερες τότε θα μπορέσουν να προκύψουν πιο ασφαλή αποτελέσματα.

Η ακρίβεια της κατάταξης-1 (rank-1 ή r-1) αναφέρεται στη συμβατική έννοια της ακρίβειας ταξινόμησης: το ποσοστό των εικόνων ανίχνευσης που ταιριάζουν απόλυτα με τις αντίστοιχες εικόνες οι οποίες έχουν χρησιμοποιηθεί για την εκπαίδευση των αλγορίθμων. Σε αντίθετη περίπτωση αν οι αποστάσεις μπορούν να ταξινομηθούν σε αύξουσα σειρά λέμε πως η εικόνα βρέθηκε στα k πρώτα αποτελέσματα και άρα είναι βαθμού-k (rank-k). Η υψηλή ακρίβεια της μετρικής αυτής είναι γνωστό ότι είναι δύσκολο να επιτευχθεί λόγω των σημαντικών προβλήματα κατά τη διαδικασία του Person Re-id.

Το ερώτημα είναι πόσο ψηλά εμφανίζονται συνήθως οι τιμές στην κατάταξη των μεθόδων αυτών. Η καμπύλη CMC συνοψίζει την πιθανότητα του επιτυχούς αποτελέσματος να εμφανίζεται στην κορυφή 1, 2,…, N της ταξινομημένης λίστας (το πρώτο σημείο στο CMC η καμπύλη είναι η ακρίβεια της κατάταξης 1). Άλλες μετρήσεις που μπορούν να εξαχθούν από την CMC περιλαμβάνουν την κλιμακωτή περιοχή κάτω από την καμπύλη και την αναμενόμενη κατάταξη. [40]

4.5. Ανοιχτά Ζητήματα στο Person Re-ID

Μπορεί η μελέτη του προβλήματος του Person Re-id να είναι σε άνθιση τα τελευταία χρόνια, ωστόσο υπάρχουν ακόμη ανοιχτά ζητήματα που απασχολούν την επιστημονική κοινότητα. Στην παρούσα ενότητα παρουσιάζονται κάποιες πληροφορίες σχετικά με τα πιο σημαντικά ανοιχτά ζητήματα τα οποία σχετίζονται με το φαινόμενο αυτό.

1. Αυξημένος όγκος δεδομένων

Ο σχολιασμός των δεδομένων - εικόνων μεγάλης κλίμακας αποτελούσε ανέκαθεν θέμα για την εκπαίδευση των συστημάτων. Το πρόβλημα αυτό είναι ακόμα πιο δύσκολο όταν γίνεται αναφορά στο φαινόμενο του person re-identification, διότι εκτός από το σχεδιασμό ενός πλαισίου οριοθέτησης ενός πεζού, πρέπει να του ανατεθεί κι ένα αναγνωριστικό. Η αντιστοίχιση ταυτότητας δεν είναι ασήμαντη, αφού ένας πεζός μπορεί να εισέλθει ξανά στα πεδία θέασης της κάμερας (Field Of View, FOV) ή να εισέλθει σε άλλη κάμερα παρατήρησης πολύ καιρό μετά την πρώτη εμφάνιση του. Αυτό δυσχεραίνει την ανίχνευση - αναγνώριση, καθώς είναι δαπανηρό για δύο διαδικασίες να προσπαθούν να ξανα-εξάγουν και να ταυτοποιήσουν τα ήδη εξαγώμενα αναγνωριστικά. Αυτές οι δυσκολίες εξηγούν εν μέρει γιατί τυπικά τα τρέχοντα σύνολα δεδομένων έχουν πολύ περιορισμένο αριθμό εικόνων για κάθε αναγνωριστικό. Τα τελευταία δύο χρόνια διαπιστώθηκε η δημιουργία πολλών εικόνων δεδομένων μεγάλης κλίμακας, π.χ. Market-1501 [89], PRW [90], LSPS [91] και MARS [92], αλλά εξακολουθούν να μην ικανοποιούν τις απόψεις των πρακτικών εφαρμογών. Βιβλιογραφικά έχουν παρουσιαστεί δύο εναλλακτικές στρατηγικές προκειμένου να βοηθήσουν στην παράκαμψη του προβλήματος των δεδομένων.

Κατ 'αρχάς, ο τρόπος χρήσης των σχολιασμών από τα σύνολα δεδομένων εντοπισμού και ανίχνευσης παραμένει υπό-έρευνα. Σε σύγκριση με το reID, οι σχολιασμοί εντοπισμού και ανίχνευσης δεν απαιτούν ανάθεση ταυτότητας όταν ένα άτομο επαν-εισέρχεται στο FOV. Στην [90], παρουσιάζεται η προσθήκη περισσότερων δεδομένων εκπαίδευσης πεζών και παρασκηνίων στο στάδιο R-CNN, η οποία ωφελεί την εκπαίδευση του περιγραφέα. Στις έρευνες [93], [94], χρησιμοποιούνται οι σχολιασμοί από ανεξάρτητα σύνολα δεδομένων για την αναπαραγωγή των εικόνων επαναταυτοποίησης. Δεδομένου ότι τα χαρακτηριστικά μπορούν να σχολιαστούν μέσα από τη συνεργασία περιγραφέων και την αποθήκευση δεδομένων και να έχουν καλή ικανότητα γενίκευσης, πρέπει να σημειωθεί ότι τα δεδομένα αυτά χρησιμοποιούνται ως καλές εναλλακτικές λύσεις για την έλλειψη δεδομένων ταυτότητας. Ως εκ τούτου, οι πόροι είναι πολύτιμοι για την κατάρτιση των συστημάτων επαναταυτοποίησης όταν δεν υπάρχουν στοιχεία κατάρτισης.

Η δεύτερη στρατηγική αφορά την μεταφορά εκμάθησης που μεταφέρει ένα εκπαιδευμένο μοντέλο από την πηγή στον τομέα-στόχο. Η εποπτευόμενη μάθηση (supervised learning) απαιτεί μεγάλο αριθμό ετικετών δεδομένων, τα οποία περιορίζουν το σύστημα όταν χρησιμοποιούνται πολλαπλές κάμερες. Στην [95] προτείνεται ένα μη επιτηρούμενο μοντέλο για την ανίχνευση πλαισίων και την αντιστοίχιση ταυτότητας προκειμένου να γίνεται ταυτόχρονη η αφαίρεση φόντου.

Στο [96] προτείνεται μια μέθοδος με χαμηλή επίβλεψη, η οποία απαιτεί πλήρη σχολιασμό από άλλα σύνολα δεδομένων επαναταυτοποίησης. Στις έρευνες [97], [98] προτείνεται μια μη επιτηρούμενη μάθηση, στην οποία το σύνολο δεδομένων στόχου δεν έχει επισημανθεί. Στην έρευνα των Ma et al. [97] χρησιμοποιείται μια SVM κατάταξη μεταξύ τομέων, ενώ οι Peng et al. [98] διατυπώνουν το πρόβλημα μεταφοράς ως λεξικό στην εργασία μάθησης, το οποίο μαθαίνει τις κοινές αμετάβλητες λανθάνουσες μεταβλητές και είναι εκπαιδευμένο βασικά με το σύνολο δεδομένων «στόχου». Αυτές οι μέθοδοι δείχνουν ότι είναι εφικτό να υπάρξει ένα δίκαιο μοντέλο για το re-identification και ότι είναι επωφελές να εξάγωνται διακριτικά στοιχεία από τα μη εποπτευόμενα δεδομένα. Η χρήση μοντέλων CNN σε άλλα σύνολα δεδομένων επαναταυτοποίησης μπορεί να είναι πιο δύσκολη, διότι οι τεχνικές deep learning παρέχουν καλή προσαρμογή και επιτυγχάνουν καλύτερα αποτελέσματα. Οι Xiao et al. [89] συγκεντρώνουν μια σειρά δεδομένων επαναταυτοποίησης και ταυτόχρονα εκπαιδεύουν ένα μοντέλο αναγνώρισης για το σύνολο δεδομένων στόχου. Συμπερασματικά παρουσιάζεται ότι η χρήση μεθόδων μάθησης μετρικής απόστασης (distance metric) [92], [93] μπορεί επίσης να διορθώσει το αποτέλεσμα μεταφοράς σε κάποιο βαθμό, αλλά και πως η μη εποπτευόμενη μάθηση εξακολουθεί να είναι ένα ανοιχτό θέμα για τα μοντέλα τα οποία χρησιμοποιούνται στο deep learning.

2. Επαναπροσδιορισμός των αποτελεσμάτων του re-identification

Η διαδικασία της επαναταυτοποίησης μπορεί να θεωρηθεί ως εργασία ανάκτησης, στην οποία η κατάταξη είναι ένα σημαντικό βήμα για τη βελτίωση της ακρίβειας της ταυτοποίησης. Η κατάταξη μπορεί να γίνει είτε εξετάζοντας το αντικείμενο επαναληπτικά, είτε πλήρως αυτόματα. Όταν εμπλέκεται η σήμανση του ανθρώπου, οι Liu et al. [99] προτείνουν τη μέθοδο βελτιστοποίησης μετά την κατάταξη (POP) η οποία επιτρέπει σε έναν χρήστη να παρέχει ένα αρνητικό δείγμα εικόνων από την αρχική λίστα του συνόλου δεδομένων. Η αραιή ανατροφοδότηση από τον άνθρωπο επιτρέπει την αυτόματη επιλογή χαρακτηριστικών. Σε μία βελτίωση, οι Wang et al. [100] σχεδιάζουν το μοντέλο (HVIL), το οποίο δεν απαιτεί προκαθορισμένα δεδομένα εκπαίδευσης και μαθαίνει από την ανθρώπινη ανατροφοδότηση για να παρέχει ενημέρωση στιγμιότυπου. Ένας αριθμός διαδοχικών μοντέλων HVIL εμπλέκονται σε ένα μοντέλο ενιαίου συνόλου για χρήση όταν η ανθρώπινη ανατροφοδότηση δεν είναι πλέον διαθέσιμη. Με παρόμοιο τρόπο, οι Martinel et al. [101] προτείνουν να βρεθούν οι πιο σχετικές εικόνες για κάθε ερώτημα, και να παραχθεί η ανθρώπινη ετικέτα. Οι ετικέτες σύμφωνα με τη μέθοδο αυτή χρησιμοποιούνται για να ενημερώσουν το μοντέλο της επαναταυτοποίησης. Οι μέθοδοι αυτόματης αναδιάταξης έχουν επίσης μελετηθεί σε πολλά έργα. Οι Zheng et al. [102] προτείνουν μια μέθοδο προσαρμογής ερωτήματος για το συνδυασμό των αποτελεσμάτων κατάταξης διαφόρων συστημάτων Re-ID. Συγκεκριμένα, χρησιμοποιείται το σχήμα των αρχικών καμπυλών βαθμονόμησης και υποστηρίζεται ότι η καμπύλη παρουσιάζει ένα σχήμα "L" για ένα καλό χαρακτηριστικό. Στο [103], διάφορες μετρήσεις γίνονται με βάση την άμεση βελτιστοποίηση της καμπύλης CMC. Οι Garcia et al. [104] αναλύουν τις πληροφορίες στο πλαίσιο της κατάταξης. Αυτό συνδυάζεται περαιτέρω με μια επανακαθορισμένη μετρική εκπαίδευσης. Οι Leng et al. [105] χρησιμοποιούν την ιδέα των πλησιέστερων γειτόνων [106] για να βελτιώσουν την αρχική κατάταξη με βάση τις σχέσεις στα βήματα εκτός σύνδεσης.

Η αναδιάταξη εξακολουθεί να είναι μια ανοικτή κατεύθυνση στο person re-identification, ενώ έχει μελετηθεί εκτενώς σε άλλες περιπτώσεις. Το σενάριο εφαρμογής μπορεί να απεικονιστεί ως εξής. Κατά την αναζήτηση ενός προσώπου ενδιαφέροντος, είναι πιθανό οι εικόνες κάτω από ορισμένες κάμερες να είναι πολύ δύσκολο να βρεθούν λόγω έντονων διακυμάνσεων της εικόνας. Αλλά ίσως είναι πιο εύκολο να εντοπιστούν οι κάμερες των οποίων οι εικόνες είναι περισσότερο όμοιες με τις εικόνες λήψης του.

3. Επαναταυτοποίηση ατόμων σε ανοιχτό περιβάλλον (open-world reID)

Οι περισσότερες έρευνες στη βιβλιογραφία μπορούν να θεωρηθούν ως μια εργασία προσδιορισμού. Οι ταυτότητες των ερωτημάτων θεωρείται ότι υπάρχουν στην βάση δεδομένων και οι εργασίες αποσκοπούν στον προσδιορισμό του αναγνωριστικού του ερωτήματος. Αντίθετα, τα συστήματα αναδημοσίευσης ανοικτού κόσμου μελετούν το πρόβλημα της επαλήθευσης του ατόμου.

Στη βιβλιογραφία, η επαναταυτοποίηση ατόμου σε ανοιχτό περιβάλλον βρίσκεται ακόμη σε πρώιμο στάδιο και προτείνονται διάφορα έργα που βοηθούν στον καθορισμό αυτού του στόχου. Στο [107] οι Zheng et al. σχεδίασαν ένα σύστημα που αποτελείται από μια λίστα εικόνων (gallery) με πολλές γνωστές ταυτότητες και έναν αριθμό ανιχνευτών. Η εργασία τους στοχεύει στην επίτευξη υψηλού ποσοστού αναγνώρισης στόχων (TTR) και χαμηλού ποσοστού εσφαλμένης αναγνώρισης στόχων (FTR), ο οποίος υπολογίζει τον ρυθμό του αριθμού των ερωτημάτων που επαληθεύονται ως ταυτότητες στόχου ως προς τον συνολικό αριθμό των ερωτημάτων. Στο [108], οι Liao et al. διαιρούν την εκ νέου ταυτοποίηση του ανοιχτού κόσμου σε δύο δευτερεύουσες εργασίες, δηλ. ανίχνευση και ταυτοποίηση. Το πρώτο μέρος αποφασίζει αν υπάρχει η ταυτότητα ανιχνευτή στη γκαλερί και ο τελευταίος εκχωρεί το αναγνωριστικό. Συνεπώς, προτείνονται δύο διαφορετικές μετρήσεις αξιολόγησης, ο ρυθμός ανίχνευσης και ταυτοποίησης (DIR) και ο ψευδής ρυθμός αποδοχής (FAR), βάσει του οποίου μπορεί να ληφθεί μια καμπύλη λειτουργίας (ROC).

Κατά συνέπεια, από τεχνικής άποψης, ο κρίσιμος στόχος είναι να βελτιωθεί η ακρίβεια αντιστοίχισης, με βάση την οποία μπορούν να σχεδιαστούν πιθανοτικά μοντέλα για μεθόδους ανίχνευσης (επαλήθευσης). Επιπλέον, όταν απαιτείται ακρίβεια στον προσδιορισμό, το open-world reID θα πρέπει επίσης να εξετάσει τη δυναμική της βάσης εικόνων που θα χρησιμοποιηθεί. Σε ένα δυναμικό σύστημα με συνεχώς εισερχόμενα πλαίσια οριοθέτησης, μια νέα ταυτότητα θα προστεθεί στη «λίστα παρακολούθησης» εάν δεν ανήκει σε καμία υπάρχουσα ταυτότητα γκαλερί και αντίστροφα. Η εγγραφή νέων ταυτοτήτων δυναμικά επιτρέπει την αυτόματη κατασκευή βάσεων δεδομένων και διευκολύνει τη διαδικασία επανεγγραφής σε μια προκαθορισμένη συλλογή.

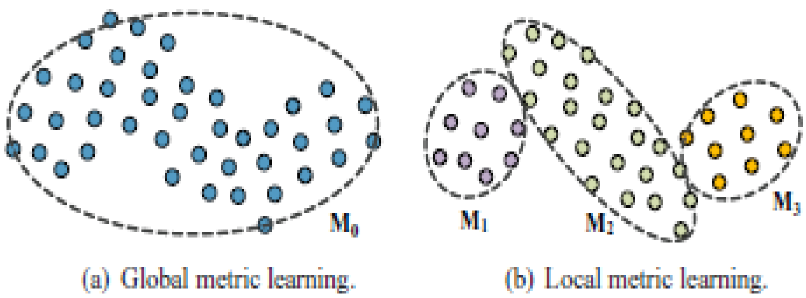
ΚΕΦΑΛΑΙΟ 5. DATASETS / ΤΕΧΝΙΚΕΣ ΚΑΙ ΜΕΘΟΔΟΙ PERSON RE-ID

5.1. Βιβλιογραφικές Αναφορές

Οι αναφορές αυτές αποτελούν την παρουσίαση τεχνικών και μεθοδολογιών οι οποίες έχουν αναπτυχθεί για την ανίχνευση και την επαναταυτοποίηση του ατόμου. Στο κεφάλαιο αυτό θα παρουσιαστούν τα επιμέρους χαρακτηριστικά των αλγορίθμων και των τεχνικών που χρησιμοποιούνται, όπως είναι οι ταξινομητές, οι βάσεις δεδομένων και γενικότερα οι αλγόριθμοι εκπαίδευσης τους. Πολλές υπάρχουσες προσεγγίσεις προσπαθούν να οικοδομήσουν μια ισχυρή αναπαράσταση χαρακτηριστικών που είναι διακριτική και επαρκής να περιγράψει την εμφάνιση ενός ατόμου κάτω από διάφορες συνθήκες [1]. Οι ερευνητές της εργασίας αυτής [1] προτείνουν τη χρήση των ταξινομητών AdaBoost προκειμένου να επιλεγούν τα «κατάλληλα» χαρακτηριστικά από ένα σύνολο χρωμάτων και χαρακτηριστικών υφής.

Σε μια άλλη σημαντική ερευνητική εργασία, ο Venice Erin Liong [2] προτείνει μια μέθοδο τοπικής μετρικής μάθησης (RLML) για την επαναταυτοποίηση και την ανίχνευση του ατόμου. Η μέθοδος αυτή εισάγει ένα συνδυασμό μιας παγκόσμιας μέτρησης και ενός συνόλου τοπικών μετρήσεων για την αναπαράσταση των δεδομένων που συνήθως ποικίλλουν τοπικά. Ως εκ τούτου, το μοντέλο ταιριάζει καλά σε διαφορετικές περιοχές και αποφεύγει ταυτόχρονα την υπερφόρτωση. Το Σχήμα 5.1

απεικονίζει το πλεονέκτημα της χρήσης μιας τοπικής μέτρησης για την αναπαράσταση των δεδομένων. Για κάθε ολικό (global) και τοπικό (local) μοντέλο, παρουσιάζεται η διαφορά κάθε θετικού ζεύγους και κάθε αρνητικού ζεύγους ως μοντέλο κατανομής Gauss υπό το Bayesian πλαίσιο και βελτιστοποιείται η αναλογία πιθανότητας, προκειμένου να παραχθεί μια διακριτική απόσταση μέτρησης. Για να εκτιμηθούν καλύτερα οι παραλλαγές εντός και εκτός τάξης στο κάθε μοντέλο, όταν υπάρχει μικρός αριθμός δειγμάτων κατάρτισης, αποσυντίθεται σε δύο υπο-διαστήματα. Με αυτόν τον τρόπο, η μάθηση της μετρικής απόστασης μπορεί να εξαγάγει περισσότερες σχετικές πληροφορίες. Η μέθοδος αξιολογείται σε τέσσερα ευρέως χρησιμοποιούμενα σύνολα δεδομένων, VIPeR, i-LIDS, GRID, και CAVIAR4REID και τα πειραματικά αποτελέσματα καταδεικνύουν την αποτελεσματικότητα της.



Σχήμα 5-1 Τοπικά και διευρυμένα συστήματα εκπαίδευσης. [1]

Μια ακόμη βιβλιογραφική έρευνα, παρουσιάζει μια διαφορετική αλγοριθμική προσέγγιση επί του θέματος. Ο Hamdoun O. [3], παρουσιάζει μια μέθοδο που ακολουθεί το κλασικό σχήμα ανίχνευσης-αναγνώρισης-επαναταυτοποίησης (DRI). Ο αλγόριθμος ο οποίος παρουσιάζεται μπορεί να χωριστεί σε δύο βασικά στάδια:

- τη φάση μάθησης
- τη φάση αναγνώρισης.

Η μάθηση συνίσταται στην αξιοποίηση της παρακολούθησης ενός δεδομένου ατόμου σε μια ακολουθία από κάμερες, προκειμένου να εξαχθούν τα σημεία ενδιαφέροντος και οι δείκτες περιγραφείς τους οι οποίοι είναι αναγκαίοι για την κατασκευή του μοντέλου που αποτελείται από μια υπογραφή για κάθε άτομο. Η ανίχνευση σημείου ενδιαφέροντος και ο υπολογισμός του περιγραφέα γίνεται χρησιμοποιώντας σημεία κλειδιά (key-points) που είναι διαθέσιμα με τη χρήση της βιβλιοθήκης Camellia [4], η οποία χρησιμοποιείται για την επεξεργασία εικόνων.

Η ανίχνευση και ο χαρακτηρισμός των βασικών σημείων Camellia και οι λειτουργίες, οι οποίες θα περιγραφούν εφαρμόζουν μια πολύ γρήγορη παραλλαγή εμπνευσμένη από τον αλγόριθμο SURF [9]. Ο αλγόριθμος αυτός αποτελεί μια αποτελεσματική εναλλακτική λύση στον πιο κλασικά και ευρέως χρησιμοποιούμενο ανιχνευτή σημείων ενδιαφέροντος και αλγόριθμο περιγραφής SIFT [8]. Τα βασικά σημεία ή σημεία ενδιαφέροντος που χρησιμοποιούνται είναι μια βελτιστοποιημένη υλοποίηση μιας παραλλαγής του αλγορίθμου SURF, χρησιμοποιώντας τον υπολογισμό μόνο με ακέραια στοιχεία, που τον κάνει ιδιαίτερα κατάλληλο και δελεαστικό για την ενσωμάτωση και τη χρήση του σε ένα κλειστό κύκλωμα τηλεόρασης.

Τα βήματα αναγνώρισης χρησιμοποιούν τον εντοπισμό των στοιχείων που πρέπει να ταυτοποιηθούν για το κάθε άτομο και τα μοντέλα κατασκευάζονται κατά τη διάρκεια της μάθησης, προκειμένου να προσδιοριστεί αν το άτομο το οποίο απεικονίζεται είναι όμοιο με τα άτομα τα οποία έχουν ήδη "ταυτοποιηθεί". Πιο συγκεκριμένα, η μέθοδος [5] μπορεί να αναλυθεί στα παρακάτω 5 βήματα:

1.Ανάπτυξη μοντέλων: Δημιουργείται μια υπογραφή για κάθε ανίχνευση και παρακολούθηση του ατόμου. Προκειμένου να μεγιστοποιηθεί η ποσότητα των μη λανθάνων πληροφοριών, δεν χρησιμοποιούνται όλα τα διαδοχικά καρέ, αλλά γίνεται δειγματοληψία των εικόνων, κάθε μισό δευτερόλεπτο. Η υπογραφή και τα χαρακτηριστικά του ατόμου αντιστοιχούν στη συσσώρευση περιγραφικών σημείων ενδιαφέροντος των εικόνων αυτών.

2.Δημιουργία ερωτήματος: Το ερώτημα για το άτομο-στόχος είναι βασισμένο σε πολλές ομοιόμορφα χρονικά διαχωρισμένες εικόνες.

3. Σύγκριση περιγραφικών στοιχείων: Η μετρική που χρησιμοποιείται για τη μέτρηση της ομοιότητας του ατόμου με τα σημεία ενδιαφέροντος των περιγραφέων, είναι το άθροισμα των απόλυτων διαφορών.

4.Γρήγορη αντιστοίχιση: Μια ισχυρή και πολύ γρήγορη αντιστοίχιση μεταξύ των περιγραφέων και των εικόνων γίνεται από τη βιβλιοθήκη Camellia, η οποία υλοποιεί μια Best Bin First (BBF) αναζήτηση σε ένα δέντρο KD [5] που περιέχει όλα τα μοντέλα.

5. Ταυτοποίηση: Η συσχέτιση του ερωτήματος με ένα από τα μοντέλα γίνεται με μια προσέγγιση ψηφοφορίας π.χ κάθε σημείο ενδιαφέροντος που εξάγεται από το ερώτημα, συγκρίνεται με όλα τα μοντέλα που είναι αποθηκευμένα στο δέντρο KD και προστίθεται μια ψήφος σε κάθε μοντέλο που περιέχει ένα αρκετά καλό ποσοστό ταυτοποίησης. Η τελική ταυτοποίηση γίνεται βάσει το υψηλότερο μοντέλο το οποίο έχει ψηφιστεί.

Ακόμη μια βιβλιογραφική αναφορά με σημαντικά αποτελέσματα, η οποία έχει αξιολογηθεί με διαφορετικές βάσεις δεδομένων, είναι η έρευνα του Sakrapee Paisitkriangkrai [6]. Στην έρευνα αυτή εισάγονται αποτελεσματικές προσεγγίσεις για τα βάρη μάθησης και τις λειτουργίες απόστασης. Η πρώτη προσέγγιση βελτιστοποιεί τη σχετική απόσταση και η δεύτερη προσέγγιση μεγιστοποιεί το rank-k, στην οποία το k επιλέγεται να είναι μικρό, π.χ., k <10. Ο καθορισμός της τιμής του k πρέπει να είναι μικρός, καθώς είναι κρίσιμος για πολλούς λόγους.

Η παραπάνω εργασία παρουσιάζει δύο προσεγγίσεις. Η πρώτη προσέγγιση η οποία αναλύεται στην εργασία, στοχεύει στη μεγιστοποίηση της σχετικής απόστασης μεταξύ εικόνων διαφορετικών ατόμων και στην ελαχιστοποίηση μεταξύ εικόνων του ίδιου ατόμου έτσι ώστε η καμπύλη (CMC) [7] να προσεγγίζει έναν ελάχιστο αριθμό. Η δεύτερη προσέγγιση βελτιστοποιεί άμεσα την πιθανότητα οποιαδήποτε από αυτές τις κορυφαίες k αντιστοιχίες να είναι σωστές χρησιμοποιώντας δομημένη μάθηση. Οι προσεγγίσεις που βασίζονται στο σύνολο είναι εξαιρετικά δομημένες και μπορεί να συνδυαστούν με γραμμικές και μη μετρήσεις. Επιπλέον, παρουσιάζονται και δύο πειράματα, τα αποτελέσματα των οποίων δείχνουν ότι η προσέγγισή αυτή επιτυγχάνει υψηλές επιδόσεις στα περισσότερα σύνολα δεδομένων που αφορούν τα δεδομένα επαναταυτοποίησης ατόμων στα οποία αξιολογούνται. Επιπλέον, η προσέγγιση του συνόλου είναι συμπληρωματική σε όλες τις υπάρχουσες μεθόδους οι οποίες υποστηρίζουν την εξ’αποστάσεως εκπαίδευση.

Η κοινώς χρησιμοποιούμενη τεχνική στην ανίχνευση αντικειμένων είναι η αντιστοίχιση του ιστογράμματος των αντικειμένων που χρησιμοποιείται από τους Gandi και Trivedi [6]. Για την ταυτοποίηση των αντικειμένων, με τη χρήση χαρακτηριστικών υφής αντικειμένων και την εξαγωγή σημείων ενδιαφέροντος κάνουν λόγο σε ορισμένα άλλα έργα οι Javed et al. [5] οι οποίοι συνδυάζουν παραμέτρους αντικειμένων κίνησης με μοντέλα εμφάνισης αντικειμένων. Δύο πιο κοινές μορφές προβλημάτων ταυτοποίησης αντικειμένων είναι ο προσδιορισμός ενός κινούμενου οχήματος και το πρόσωπο ενός ατόμου. Το πρώτο είναι πιο απλή διαδικασία [48] λόγω της ακαμψίας των οχημάτων, των χώρων στον οποίο κινούνται και του ομοιόμορφου χρώματος που έχουν. Στην περίπτωση της ταυτοποίησης των οχημάτων τα πιο κοινά χαρακτηριστικά που χρησιμοποιούνται είναι: μέγεθος, ταχύτητα, η θέση λωρίδας, οι πληροφορίες χρώματος και ο χρόνος παρατήρησης.

Έχουν παρουσιαστεί διάφορα μοντέλα για την επαναταυτοποίηση του ατόμου στη βιβλιογραφία. Συνήθως διαφέρουν από τα σύνολα χαρακτηριστικών που χρησιμοποιούνται για την απεικόνιση των εικόνων και τη στρατηγική που χρησιμοποιείται για την τελική απόφαση του ταιριάσματος των πεζών.

Σε μία από τις βασικότερες μελέτες, οι Gheissari et al. [15] καθόρισαν μια αμετάβλητη υπογραφή βασισμένη σε ένα συνδυασμό κανονικοποιημένων χρωμάτων και ιστογράμμων edgel. Εισήγαγαν έναν νέο αλγόριθμο κατακερματισμού για να δημιουργήσουν προεξέχοντα άκρα τα οποία είναι εύρωστα στις αλλαγές που παρουσιάζονται στην εμφάνιση των ενδυμάτων ενός ατόμου. Οι πληροφορίες χρώματος καταγράφηκαν σε ιστογράμματα με βάση την απόχρωση και τον κορεσμό. Η τελική επαναταυτοποίηση επιτεύχθηκε με την αξιολόγηση των αποστάσεων με ιστόγραμμα ανά ζεύγη.

Οι Hamdoun et al. [39] πρότειναν μια άλλη υπογραφή βασισμένη σε περιγραφείς σημείων ενδιαφέροντος που λαμβάνονται από ένα σύνολο εικόνων που προκύπτουν δημιουργώντας ένα πολύ-καναλικό μοντέλο κάθε πεζού, συσσωρεύοντας σημεία ενδιαφέροντος χρονοσειρών χρησιμοποιώντας μια αποδοτική παραλλαγή του SURF [11]. Για το ταίριασμα των πεζών, πραγματοποίησαν την αναζήτηση Best Bin First σε ένα δέντρο KD που περιέχει όλα τα μοντέλα. Ο σημαντικότερος περιορισμός αυτής της εργασίας είναι το γεγονός ότι απαιτεί πολλαπλές λήψεις κάθε ατόμου σε περισσότερες από δύο κάμερες για την αποτελεσματική εκμετάλλευση υπογραφών χρονοσειρών. Ένα άλλο πρόβλημα είναι ότι απαιτείται ένα σύνολο εικόνων υψηλής ποιότητας για την ανίχνευση πολύτιμων σημείων ενδιαφέροντος, κάτι που σπάνια συμβαίνει με τα αρχεία παρακολούθησης μέσω κάμερας.

Παρόλο που πολλά χαρακτηριστικά χρώματος και υφής είχαν χρησιμοποιηθεί σε υπολογισμό απόστασης εικόνας για αντιστοίχιση προσώπων, οι Gray και Tao [45] εισήγαγαν ένα αυτοματοποιημένο πρόγραμμα επιλογής το οποίο εντόπιζε τα πιο αντιπροσωπευτικά σύνολα χαρακτηριστικών. Δημιούργησαν ένα συνοδευτικό σύνολο δεδομένων, που ονομάζεται ViPeR. Σε αυτό το δύσκολο σύνολο δεδομένων, έδειξαν ότι ένα σύνολο αντικειμενικών αναλλοίωτων χαρακτηριστικών μπορεί να επιλεγεί με μια αποτελεσματική προσέγγιση μάθησης βασισμένο στη μέθοδο-αλγόριθμο AdaBoost. Μέσα από την έρευνά τους αποδεικνύουν ότι το πιο πολύτιμο χαρακτηριστικό γνώρισμα για την αναγνώριση ατόμων είναι το ιστόγραμμα Hue. Η μέθοδος τους αυτή συνήθως αναφέρεται ως ELF (Ensemble of Local Features) στη βιβλιογραφία.

Οι Farenzena et al. [52] ανέφεραν τα καλύτερα αποτελέσματα που έχουν επιτευχθεί στο σύνολο δεδομένων VIPeR, το οποίο είναι μέχρι σήμερα το πιο δύσκολο διαθέσιμο σύνολο δεδομένων με ένα μόνο πλάνο. Με την αυστηρή προ-επεξεργασία των εικόνων, εξήγαγαν ολόκληρο το σώμα από το φόντο και το χώρισαν αυτόματα σε τρία σημασιολογικά μέρη: το κεφάλι, τον κορμό και τα πόδια. Σε κατακόρυφα διαχωρισμένες εικόνες, συσσωρεύτηκαν οριζόντια χαρακτηριστικά με γνώμονα τη συμμετρία. Εισήγαγαν επίσης ένα νέο σχήμα αναπαράστασης χαρακτηριστικών με βάση την παρουσία εξαιρετικά επαναλαμβανόμενων δομημένων σημείων. Το συνολικό μοντέλο ονομάζεται SDALF (Symmetry Driven Accumulation of Local Features).

Μια σύγκριση χαρακτηριστικών βάσει σημείων ενδιαφέροντος για το Person Re-identification παρουσιάστηκε από τους Bauml et al. [53]. Παρομοίως, μια σύγκριση χαρακτηριστικών χρωματικού ιστογράμματος μπορεί να βρεθεί στην έρευνα των Gray et al. [54]. Σύμφωνα με τα πειραματικά αποτελέσματα, τα χαρακτηριστικά χρώματος ξεπερνούν τα υπόλοιπα με βάση την καμπύλη των χαρακτηριστικών σωρευτικής αντιστοίχισης (CMC), η οποία αποτελεί ένα κοινό πειραματικό μέτρο της ακρίβειας στην επαναταυτοποίηση του ατόμου.

Οι Bak et al. [55] [56] είναι δύο άλλα παραδείγματα περιγραφικών τεχνικών που χρησιμοποιούν ένα σύνολο τοπικών χαρακτηριστικών που εξάγονται από την εικόνα και στη συνέχεια χρησιμοποιούν ένα μέτρο απόστασης για το ταίριασμά τους. Η προσέγγιση συνδυάζει τα χαρακτηριστικά γνωρίσματα υφής Haar με κυρίαρχο περιγραφέα το χρώμα αξιολογώντας την απόδοση των περιοχών χωρικής συνάφειας με την κατάτμηση της εικόνας σε μέρη του σώματος.

Ο Cai και ο Pietikainen [57] πρότειναν μια νέα προσέγγιση, εμπνευσμένη από μοντέλα αντιστοίχισης εγγράφων, τα οποία αντιπροσώπευαν κάθε εικόνα από τις μετρήσεις ενός συνόλου χρωμάτων. Οι λέξεις-κλειδιά δημιουργούνται εφαρμόζοντας έναν απλό αλγόριθμο k-mean σε παράθυρα 3x3 σε ένα σύνολο εικόνων κατάρτισης. Επίσης τα ιστογράμματα απόχρωσης θα μπορούσαν να επιτύχουν με αρκετά καλή ακρίβεια στην αντιπροσώπευση των προσωπικών στοιχείων.

Μέσα από αυτή την βιβλιογραφική αναφορά προσπαθούμε να επιτύχουμε τη κατανόηση της βασικής προσέγγισης που ακολουθείται σε κάθε ερευνητική εργασία με τη χρήση των βάσεων δεδομένων καθώς και των αλγορίθμων οι οποίοι χρησιμοποιούνται. Στη συνέχεια του κεφαλαίου αυτού θα παρουσιαστούν αναλυτικότερα οι κυριότερες τεχνικές και αλγόριθμοι οι οποίοι έχουν προκύψει από την βιβλιογραφική αναζήτηση η οποία έχει γίνει για τις ανάγκες της εργασίας.

5.2. Βάσεις Δεδομένων Εικόνων (Datasets / Galleries)

Τα οπτικά χαρακτηριστικά ενός ατόμου διαφοροποιούνται ανάλογα με το αν υπάρχει διαφορά φωτισμού, διαφορετικές οπτικές γωνίες, κλίμακες και μορφολογία καμερών. Προκειμένου να αναπτυχθούν ισχυρές τεχνικές επαναταυτοποίησης, είναι σημαντικό να αποκτηθούν δεδομένα τα οποία θα χρησιμοποιηθούν για να έχουν αυξημένη αποτελεσματικότητα. Τα δεδομένα αυτά θα πρέπει να έχουν υψηλή ποιότητα για να μπορούν να συγκριθούν με τη χρήση διαφορετικών προσεγγίσεων και μεθόδων οι οποίες αναπτύσσονται και εντοπίζονται στις τεχνικές και τα σύνολα δεδομένων.

Μερικά από τα πιο βασικά στοιχεία για τη δημιουργία βάσεων δεδομένων εικόνων είναι τα εξής:

- μεταβολές στο φωτισμό
- η οπτική γωνία της κάμερας
- το περιβάλλον στο οποίο τοποθετείται (στατικό ή δυναμικό)
- τα εμπόδια τα οποία μπορεί να υπάρχουν
- ο αριθμός λήψεων που θα γίνει από κάθε κάμερα
- τα μεγέθη των εικόνων τα οποία παράγονται

Στη βιβλιογραφία υπάρχουν αρκετά διαθέσιμα σύνολα δεδομένων που έχουν δημιουργηθεί για τη δοκιμή των μοντέλων με σκοπό την καλύτερη και υψηλότερη επίτευξη της αποτελεσματικότητας των αλγορίθμων που χρησιμοποιούνται. Στην ενότητα αυτή θα παρουσιαστούν τα χαρακτηριστικά μερικών από των πιο βασικών δεδομένων εικόνων οι οποίες χρησιμοποιούνται κατά κόρον στην βιβλιογραφία και οι οποίες θα χρησιμοποιηθούν για τη σύγκριση στη παρούσα εργασία.

5.2.1. ViPER

Το σύνολο δεδομένων ViPER [8] αποτελείται από εικόνες ανθρώπων από δύο διαφορετικές προβολές κάμερας, αλλά έχει μόνο μία εικόνα για κάθε άτομο ανά φωτογραφική μηχανή. Το σύνολο δεδομένων περιέχει 632 ζεύγη εικόνων. Οι κάμερες έχουν διαφορετικές απόψεις και υπάρχουν διαφορές φωτισμού μεταξύ τους προκειμένου να υπάρχει ποικιλομορφία στο εκάστοτε δείγμα. Οι εικόνες έχουν περικοπεί και είναι σε κλίμακα 128 × 48 pixel. Η βάση δεδομένων αυτή χρησιμοποιείται κατά κόρον στη βιβλιογραφία καθώς διαθέτει διάφορα δείγματα ανθρώπων και εικόνες με διαφορετικές συνθήκες, και οι απαιτήσεις της ανίχνευσης είναι αυξημένες.

Τα δεδομένα που συλλέχθηκαν για τη δοκιμή των διαφόρων αλγορίθμων αναγνώρισης έχουν ως στόχο να συγκεντρώσουν όσες το δυνατόν περισσότερες όμποιες όψεις ζευγών εικόνων. Για να γίνει ένα ιδανικό σύνολο δεδομένων, θα πρέπει να αποτελείται από ζεύγη με ομοιόμορφη δειγματοληψία μεμονωμένων ατόμων από πολλαπλές οπτικές γωνίες. Στην εν λόγω βάση δεδομένων χρησιμοποιήθηκαν εικόνες οι οποίες έχουν ληφθεί με διάφορες γωνίες κλίσης. Στο (πίνακα 5-1) που ακολουθεί παρουσιάζονται κάποια από αυτά τα δείγματα τα οποία αντιστοιχούν σε διάφορες γωνίες λήψης.

Σχήμα 5-2 Εικόνες από τη βάση δεδομένων ViPER. [8]

Πίνακας 5-1 Κατανομή εικόνων ανάλογα με τη γωνία λήψης στο σύνολο ViPER

Γωνία λήψης εικόνων	Εικόνες στο δείγμα
45	70
90	363
135	96
180	103

5.2.2. i-LIDS

Αυτό το σύνολο δεδομένων δημιουργήθηκε το 2014 από την παρακολούθηση πολλαπλών καμερών CCTV στην αίθουσα αφίξεων ενός αεροδρομίου. Το σύνολο δεδομένων περιέχει συνολικά περίπου 43K εικόνες διαφόρων μεγεθών από 300 πεζούς που έχουν ληφθεί από δύο μη επικαλυπτόμενες κάμερες. Το σύνολο δεδομένων παρουσιάζει σημαντικές παραλλαγές φωτισμού και η οριοθέτηση των πλαισίων (b-boxes) εικόνων έχει γίνει με χειροκίνητο τρόπο. Η εικόνα που ακολουθεί παρουσιάζει ένα δείγμα εικόνων από αυτό το σύνολο δεδομένων.

Σχήμα 5-3 Εικόνες από τη βάση δεδομένων i-LIDS. [9]

5.2.3. CAVIAR4REID

Το CAVIAR4REID έχει εξαχθεί από τη βάση δεδομένων CAVIAR και ως σύνολο περιέχει εικόνες οι οποίες έχουν ληφθεί σε ένα εμπορικό κέντρο της Λισαβόνας από δύο κάμερες οι οποίες βρίσκονται σε επικάλυψη. Το σύνολο δεδομένων το οποίο δημιουργείται περιέχει εικόνες από 72 πεζούς, από τις οποίες μόνο 50 εμφανίζονται και στις δύο κάμερες, ενώ 22 προέρχονται από την ίδια κάμερα. Οι εικόνες για κάθε πεζό επιλέχθηκαν με στόχο τη μεγιστοποίηση των μεταβολών της εμφάνισης λόγω των αλλαγών στην ανάλυση και τις διαφορετικές συνθήκες φωτισμού. Το ελάχιστο μέγεθος εικόνων είναι 17×39 και το μέγιστο 72×144, αντίστοιχα. Για κάθε πεζό, επιλέχθηκαν 10 εικόνες από 2 διαφορετικές προβολές κάμερας μεγιστοποιώντας τη διακύμανση σε σχέση με τις αλλαγές της ανάλυσης, τις συνθήκες φωτισμού, τις αποκλείσεις και τις αλλαγές θέσης. Οι πολλαπλές κάμερες καθώς και οι πολλαπλές λήψεις από διαφορετικές γωνίες έχουν χρησιμοποιηθεί για λόγους ποικιλομορφίας στα δείγματα εικόνων καθώς και στα αποτελέσματα ανίχνευσης.

Σχήμα 5-4 Εικόνες από τη βάση δεδομένων CAVIAR4REID. [10]

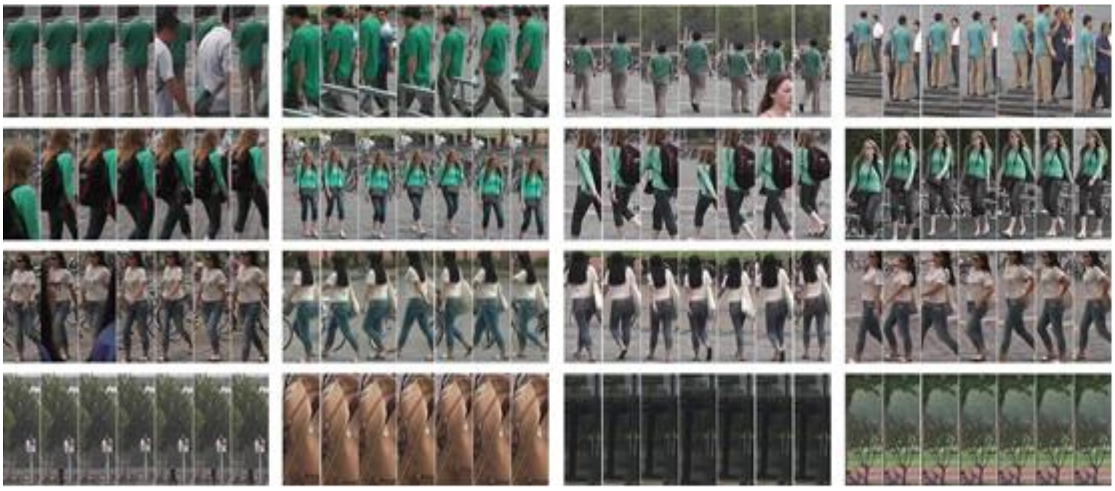
5.2.4. Market-1501

Κατά την ανάπτυξη αυτού του συνόλου δεδομένων, χρησιμοποιήθηκαν συνολικά έξι κάμερες οι οποίες τοποθετήθηκαν μπροστά σε ένα super - market Πανεπιστημιούπολης, συμπεριλαμβανομένων πέντε φωτογραφικών μηχανών HD με ανάλυση 1280 × 1080 και μιας κάμερα SD με ανάλυση 720 × 576. Οι κάμερες αυτές είχαν αλληλοεπικάλυψη, καθώς στόχος ήταν η πολύπλευρη λήψη των φωτογραφιών. Αυτό το σύνολο δεδομένων περιέχει 1.501 ταυτότητες ατόμων. Κάθε άτομο συλλαμβάνεται από τουλάχιστον δύο κάμερες, έτσι ώστε να μπορεί να πραγματοποιηθεί αναζήτηση διασταυρούμενης πληροφορίας σε κάποια άλλη κάμερα. Προκειμένου να γίνει η ανίχνευση των ατόμων αυτόματα και να επιτευχθούν καλύτερα αποτελέσματα, το σύνολο δεδομένων αυτό περικλείει έναν ανιχνευτή Deformable Part Model (DMP) [11] ο οποίος ευθυγραμμίζει την εικόνα, και παρέχει μια καλύτερη και άμεση ανίχνευση του ατόμου (bboxes). Επιπλέον, κάθε ταυτότητα μπορεί να έχει πολλές εικόνες κάτω από κάθε κάμερα. Ως εκ τούτου, κατά τη διάρκεια αναζήτησης στις κάμερες, μπορεί να βρεθούν πολλαπλά ερωτήματα και πολλαπλές απαντήσεις ως έξοδοι για το κάθε άτομο. Με αυτό το τρόπο μπορούν να αξιοποιηθούν πλήρως τα πολλαπλά ερωτήματα και οι περισσότερες διακριτικές πληροφορίες σχετικά με το άτομο το οποίο ενδιαφέρει το σύστημα. Ένα δείγμα εικόνων από το εν λόγω δείγμα δεδομένων παρουσιάζεται παρακάτω.

Σχήμα 5-5 Εικόνες από τη βάση δεδομένων Market-1501. [11]

5.2.5. MARS

Το σύνολο δεδομένων αυτό αποτελεί επέκταση του συνόλου Market-1501. Κατά τη διάρκεια της συλλογής εικόνων για το σύνολο αυτό, τοποθετήθηκαν έξι συγχρονισμένες κάμερες στην πανεπιστημιούπολη του Πανεπιστημίου Tsinghua. Το σύνολο δεδομένων MARS [12] αποτελείται από 1.261 διαφορετικά δείγματα πεζών που συλλαμβάνονται από τουλάχιστον 2 κάμερες. Για την παραγωγή του ανιχνευτή χρησιμοποιείται και στην περίπτωση αυτή ο ανιχνευτής DPM. Στη συνέχεια, χρησιμοποιείται ο αλγόριθμος GMMCP για την ομαδοποίηση ανίχνευσης της αλληλοεπικάλυψης των εικόνων, ώστε να παραχθούν τα διαδοχικά πλαίσια ανίχνευσης. Το αποτέλεσμα των διαδικασιών αυτών είναι η παραγωγή συνολικά 1.2M εικόνων.



Σχήμα 5-6 Εικόνες από τη βάση δεδομένων MARS

5.2.6. Αξιολόγηση Dataset

Σαφής τρόπος αξιολόγησης αλλά και παρουσίασης του καλύτερου dataset δεν υπάρχει. Κάθε βάση δεδομένων έχει τα δικά της χαρακτηριστικά να προσδώσει που θα χρησιμοποιηθούν αντίστοιχα από τους διάφορους αλγορίθμους προκειμένου να εξεταστεί καθένας από αυτούς και να παρουσιάσουν τα αποτελέσματα της ανίχνευσής τους. Η αξιολόγηση των βάσεων δεδομένων στην περίπτωση αυτή αφορά την παρουσίαση των συνολικών δεδομένων και την κατανόηση της διαφορετικότητας της εκάστοτε βάσης. Στον πίνακα που ακολουθεί (πίνακας 5-2) παρουσιάζονται συνοπτικά τα χαρακτηριστικά των παραπάνω βάσεων δεδομένων.

Πίνακας 5-2 Παρουσίαση χαρακτηριστικών στα διάφορα σύνολα δεδομένων

Βάση δεδομένων εικόνων	Ημερομηνία παραγωγής	Σύνολο ατόμων	Κάμερες λήψης	Σύνολο εικόνων (bboxes)	Μέθοδος ετικετοποίησης	Μέγεθος εικόνων
ViPER	2007	632	2	1264	Χειροκίνητα	128x48
CAVIAR4REID	2011	72	2	1220	Χειροκίνητα	Πολλαπλά μεγέθη
i-LIDS	2014	300	2	42.495	Χειροκίνητα	Πολλαπλά μεγέθη
Market-1501	2015	1.501	6	32.643	DPM	Πολλαπλά μεγέθη
MARS	2016	1.261	6	1,2M	DPM+GMMCP	Πολλαπλά μεγέθη

5.3. Τεχνικές Εξαγωγής Χαρακτηριστικών

Αρκετοί είναι οι αλγόριθμοι που έχουν παρουσιαστεί βιβλιογραφικά για την εξαγωγή χαρακτηριστικών από μια εικόνα. Στην παρούσα ενότητα θα παρουσιαστούν μερικές τεχνικές εξαγωγής χαρακτηριστικών εικόνων που χρησιμοποιούνται στο Re-id. Οι τεχνικές αυτές θα χρησιμοποιηθούν στο κεφάλαιο της σύγκρισης το οποίο ακολουθεί.

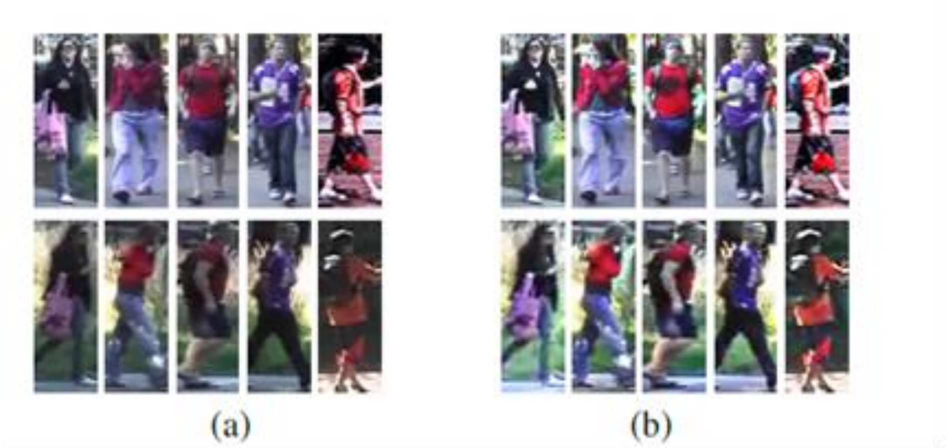
5.3.1. Local Maximal Occurrence (LOMO)

Η συγκεκριμένη τεχνική (Local maximal occurrence, LOMO) επηρεάζεται σε μεγάλο βαθμό από την άποψη των αλλαγών, οι οποίες θα πρέπει να αντιμετωπιστούν στο σχεδιασμό χαρακτηριστικών ή κατά την εκπαίδευση του ταξινομητή. Η προτεινόμενη τοπική εξαγωγή χαρακτηριστικών μέγιστης εμφάνισης είναι μια στρατηγική προς την κατεύθυνση της ισχυρής αναπαράστασης χαρακτηριστικών.

Το χαρακτηριστικό του αλγορίθμου LOMO αναλύει την οριζόντια εμφάνιση των τοπικών χαρακτηριστικών και μεγιστοποιεί την εμφάνιση για να κάνει μια σταθερή αναπαράσταση ενάντια στις αλλαγές θέσης. Εκτός αυτού, έχει διαπυστωθεί ότι η εφαρμογή του μετασχηματισμού Retinex [63] είναι χρήσιμη για την αντιμετώπιση αλλαγών φωτισμού στην επαναταυτοποίηση ατόμων. Για την εκμάθηση μιας μεροληπτικής διάκρισης, προτείνεται να αποσυντεθούν χωριστά οι εικόνες για να μειωθούν οι διαστάσεις με τη χρήση της τεχνικής τετραγωνικής ανάλυσης xQDA. Με τη χρήση αυτής εξετάζεται μια υποπεριοχή με δεδομένα τα οποία έχουν ληφθεί από αλληλεπικάλυψη. Έχει δειχθεί ότι το πρόβλημα μπορεί να διαμορφωθεί ως γενικευμένο Rayleigh Quotient, και μια λύση κλειστής μορφής μπορεί να επιτευχθεί με τη γενικευμένη αποσύνθεση ιδιοτιμών.

Το χρώμα αποτελεί ένα σημαντικό χαρακτηριστικό για την περιγραφή εικόνων προσώπου. Ωστόσο, οι συνθήκες φωτισμού στις κάμερες μπορεί να είναι πολύ διαφορετικές και οι ρυθμίσεις της κάμερας ενδέχεται να διαφέρουν από κάμερα σε κάμερα. Επομένως, τα αντιληπτά χρώματα του ίδιου ατόμου μπορεί να διαφέρουν σε μεγάλο βαθμό από τις διαφορετικές οπτικές της κάμερας. Για παράδειγμα, το Σχήμα 5-7 (α) δείχνει ορισμένες εικόνες δείγματος από τη βάση δεδομένων VIPeR. Μπορούμε να δούμε ότι οι εικόνες του ίδιου προσώπου στις δύο προβολές κάμερας έχουν μεγάλη διαφορά στον φωτισμό και την εμφάνιση χρώματος.. Το Retinex [64] προτείνεται για την προ-επεξεργασία εικόνων προσώπου γι’αυτό το σκοπό καθώς λαμβάνει την ανθρώπινη υπόσταση και την αντίληψη χρώματος. Στόχος του είναι να παράγει μια έγχρωμη εικόνα που να είναι σύμφωνη με την ανθρώπινη παρατήρηση της σκηνής. Η αποκατεστημένη εικόνα συνήθως περιέχει ζωηρές πληροφορίες χρώματος με ιδιαίτερα βελτιωμένες λεπτομέρειες σε σκιασμένες περιοχές.

Εφαρμόζοντας τον πολύ-επίπεδο αλγόριθμο Retinex σύμφωνα με το [64], ο οποίος συνδυάζει ταυτόχρονα την Retinex μικρής κλίμακας για συμπίεση δυναμικού εύρους και την ευρεία κλίμακα Retinex για την απόδοση τόνων έχει ως αποτέλεσμα, ο αλγόριθμος να χειρίζεται τη σταθερότητα του χρώματος και τη συμπίεση δυναμικού εύρους αυτόματα, επιτυγχάνοντας μια καλή προσέγγιση με την ανθρώπινη οπτική αντίληψη. Συγκεκριμένα, χρησιμοποιώντας δύο κλίμακες Retinex κεντρικού/ περιβάλλοντος, με $\sigma = 5$ και $\sigma = 20$ υπολογίζονται αυτόματα οι παραμέτροι κέρδους / αντιστάθμισης έτσι ώστε οι προκύπτουσες εντάσεις να εκτείνονται γραμμικά σε [0, 255].



Σχήμα 5-7 (a) Παράδειγμα ζευγών εικόνων από τη βάση δεδομένων VIPeR. (b) Επεξεργασμένες εικόνες με χρήση Retinex [66]

5.3.2. Bag-of-words (BoW)

Στην τεχνική αυτή χρησιμοποιείται κυρίως το μοντέλο συνόλου λέξεων (Bag-of-Words) ή BoW [39], η οποία αποτελεί μια αρκετά κοινή προσέγγιση για τον συνδυασμό των τοπικών χαρακτηριστικών στη παραγωγή μιας παγκόσμια υπογραφής στον κομμάτι του re-identification. Επιπλέον, συνδιάζεται με άλλες τεχνικές μετρικής οι οποίες χρησιμοποιούνται για την επίτευξη καλύτερων αποτελεσμάτων. Στην μέθοδο αυτή τα τοπικά χαρακτηριστικά που εξάγονται από μια εικόνα χαρτογραφούνται πρώτα σε ένα σύνολο οπτικών λέξεων και στη συνέχεια η εικόνα αντιπροσωπεύεται ως ιστόγραμμα των εμφανίσεων μιας οπτικής λέξης. Τα χαρακτηριστικά που προκύπτουν από την επεξεργασία των εικόνων, στον πραγματικό κόσμο αντιστοιχούν στις λέξεις που περιέχει ένα λεξικό. Όμως στο λεξικό οι λέξεις δεν τοποθετούνται σε τυχαία σειρά, αλλά αλφαβητικά. Ο ουσιαστικός λόγος ύπαρξης της αλφαβητικής σειρά είναι για να συγκεντρώσει τις οπτικά όμοιες λέξεις και να τις τοποθετήσει σε σειρά. Αυτή η σειρά μας επιτρέπει να γλιτώνουμε αρκετό χρόνο σε κάθε αναζήτηση για την ανάκτηση μιας λέξης, καθότι περιορίζει αρκετά το εύρος της αναζήτησης. Το μοντέλο BoW χρησιμοποιήθηκε για το re-identification, όπου οι συγγραφείς χτίζουν ομάδες περιγραφών ενσωματώνοντας τις οπτικές λέξεις σε ομόκεντρες δομές του χώρου και εμπλουτίζουν την περιγραφή ενός ατόμου από τις πληροφορίες συμφραζομένων που προέρχονται από τα επιμέρους τμήματα της εικόνας. Για να αναπαρασταθεί μια εικόνα χρησιμοποιώντας το μοντέλο BoW, χρησιμοποιείται η αναπαράστασή της ως έγγραφο. Ομοίως, οι λέξεις στις εικόνες πρέπει να οριστούν και για να επιτευχθεί αυτό, απαιτούνται συνήθως τα εξής τρία βήματα:

- ανίχνευση χαρακτηριστικών
- περιγραφή χαρακτηριστικών
- δημιουργία βιβλίου κωδικών

Μετά την ανίχνευση των χαρακτηριστικών, κάθε εικόνα αφαιρείται από πολλά τοπικά ελαττώματα, όπως είναι ο θόρυβος. Οι μέθοδοι αναπαράστασης χαρακτηριστικών ασχολούνται με τον τρόπο με τον οποίο αντιπροσωπεύεται ο θόρυβος ως αριθμητικοί φορείς. Αυτοί οι φορείς ονομάζονται περιγραφείς χαρακτηριστικών. Ένας καλός περιγραφέας πρέπει να έχει την ικανότητα να χειρίζεται την ένταση, τη περιστροφή, τη κλίμακα και συγγενείς παραλλαγές μιας εικόνας. Ένας από τους πιο διάσημους περιγραφείς είναι ο μετασχηματισμός χαρακτηριστικών κλίμακας (SIFT). Για τη δημιουργία βιβλίου κωδικών (τελικό βήμα) απαιτείται η μετατροπή των πακέτων που αντιπροσωπεύονται από τον φορέα σε "κωδικές λέξεις" το οποίο παράγει επίσης ένα "βιβλίο κωδικών". Μια λέξη κώδικα μπορεί να θεωρηθεί ως ένας εκπρόσωπος πολλών όμοιων στοιχείων. Μια απλή μέθοδος (k-means) εκτελεί την ομαδοποίηση σε όλους τους φορείς. Οι λέξεις-κλειδιά ορίζονται ως τα κέντρα κάθε ομάδας.

Παράδειγμα χρήσης της τεχνικής BoW στην [39], με τα εξής βήματα:

1.Εξαγωγή χαρακτηριστικών: Για το σκοπό αυτο χρησιμοποιείται ο περιγραφέας (color name descriptor). Δεδομένης μιας εικόνας ατόμου που δίνεται ως είσοδος με μέγεθος 128 × 64 pixels και με βήμα δειγματοληψίας 4 (δεν υπάρχει αλληλο-επικάλυψη μεταξύ των patches). Για κάθε bbox, οι περιγραφείς CN όλων των εικονοστοιχείων υπολογίζουν τα δεδομένα της εικόνας μέσα από κανονικοποίηση. Ο μέσος φορέας λαμβάνεται ως ο περιγραφέας αυτής.

2.Κωδικοποιητής: Για το Market-1501, δημιουργείτα ένα βιβλίο κωδικών για το σετ εκπαίδευσης του. Για άλλα σύνολα δεδομένων, το βιβλίο κωδικών είναι εκπαιδευμένο σχετικά με το ανεξάρτητο σύνολο δεδομένων.

3.Κβαντισμός: Δεδομένου ενός τοπικού περιγραφέα, χρησιμοποιούνται πολλαπλά στοιχεία για να βρεθούν οι εγγύτεροι γείτονες κάτω από την Ευκλείδεια απόσταση στο βιβλίο κωδικών. Πρόσφατα, το μοντέλο BoW ενισχύθηκε σημαντικά με την εισαγωγή του FLDA που κωδικοποιεί στατιστικά στοιχεία υψηλότερης τάξης τοπικών χαρακτηριστικών.

Κανόνες βελτιστοποίησης αναγνώρισης που χρησιμοποιούνται στην μέθοδο αυτή:

• Αδύναμοι γεωμετρικοί περιορισμοί (Weak Geometric Constraints):

Για την κωδικοποίηση του γεωμετρικού περιορισμού περιλαμβάνουν την εκτεταμένη μήτρα αδράνειας (ACS). Αυτή η μέθοδος είναι αποτελεσματική στην ενσωμάτωση των χωρικών περιορισμών, αλλά υποφέρει από υψηλό υπολογιστικό κόστος. Ενσωματώνοντας την ACS στο μοντέλο BoW καταφέρνουμε να χωριστεί μια εικόνα εισόδου σε M οριζόντιες λωρίδες. Στη συνέχεια, για τη λωρίδα m_i το ιστόγραμμα οπτικής λέξης αντιπροσωπεύεται ως $d_m = (d_{m1}, d_{m2}, ..., d_{mk})^T$, όπου k είναι το μέγεθος του βιβλίου κωδικών. Συνεπώς, το διάνυσμα χαρακτηριστικών για την εικόνα εισόδου υποδηλώνεται ως $f = (d_1, d_2, ..., d_M)^T$, δηλ. η σύζευξη των φορέων από όλες τις λωρίδες. Κατά την αντιστοίχιση δύο εικόνων συνοψίζεται η ομοιότητα σε όλες τις αντίστοιχες λωρίδες. Ως εκ τούτου, αποφεύγετε ο ακριβός υπολογισμός των αποστάσεων σε κάθε αναγνώριση.

• Εξομάλυνση φόντου. Οι αρνητικές επιπτώσεις του φόντου έχουν μελετηθεί εκτενώς. Σε μια ερευνητική προσέγγιση αναφέρεται η διάσπαση (ανάπτυξη μασκών) της εικόνας εισόδου σε μέρη όπως το φόντο της και το υπόλοιπο μέρος της. Η διαδικασία της δημιουργίας μιας μάσκας για κάθε εικόνα είναι χρονοβόρα και ασταθής, ωστόσο υπάρχει μια απλή λύση με τη χρήση ενός 2-D Gaussian πρότυπου εικόνας. Συγκεκριμένα, η Gaussian λειτουργία παίρνει τη μορφή του $N(\mu_x, \sigma_x, \mu_y, \sigma_y)$, όπου μ_x, μ_y είναι οι οριζόντιες και κάθετες μέσες τιμές και σ_x, σ_y είναι οι οριζόντιες και κάθετες Gaussian τυποποιημένες διακυμάνσεις, αντίστοιχα. Ορίζεται (μ_x, μ_y) το κέντρο της εικόνας και ως $(\sigma_x, \sigma_y) = (1, 1)$ για όλα τα πειράματα. Αυτή η μέθοδος υποθέτει ότι το άτομο βρίσκεται στο κέντρο μιας εικόνας, και περιβάλλεται από φόντο.

• Πολλαπλά ερωτήματα. Εμφανίζεται η χρήση πολλαπλών ερωτημάτων για να αποδωθούν καλύτερα αποτελέσματα στην αναζήτηση εικόνων και στην αναγνώριση. Επειδή λαμβάνεται υπόψη η παραλλαγή εντός της κλάσης, ο αλγόριθμος είναι πιο ανθεκτικός στις παραλλαγές των πεζών. Όταν κάθε ταυτότητα έχει πολλαπλές ερωτήσεις σε μία εικόνα κάμερα, αντί για τη στρατηγική πολλαπλών συνδυασμών αντιστοίχισης, συγχωνεύονται όλα τα στοιχεία σε ένα μόνο ερώτημα για την εξέταση της ταχύτητας.

5.3.3. Gaussian of Gaussians (GOG)

Σε αυτή την ενότητα παρουσιάζονται δύο ιεραρχικοί περιγραφείς Gaussian (HGD).Και οι δύο βασίζονται σε μια κοινή λογική που ακολουθεί δύο κίνητρα:

1)Η ιεραρχική κατανομή είναι διακριτή επειδή μπορεί να διακρίνει τις υφές με παρόμοια παγκόσμια κατανομή αλλά με διαφορετική τοπική κατανομή των χαρακτηριστικών των pixel
2)Κατά την κατασκευή του ιεραρχικού περιγραφέα, οι πληροφορίες των χαρακτηριστικών των εικονοστοιχείων στα τοπικά μέρη τείνουν να είναι σημαντικές διακριτές πληροφορίες.

Δύο ιεραρχικοί Gaussian περιγραφείς ο (GOG) και ο (ZOZ) που εξάγονται με την αλλαγή της βασικής Γκαουσιανής ενσωμάτωσης της κοινής διαδικασίας. Κάθε ένας παρέχει χαρακτηριστικά των περιοχών σε μορφή δυανίσματος για να καταστεί δυνατή εύκολη η εφαρμογή της μάθησης της συμβατικής μεθόδου μετρικής. Η απεικόνιση χαρακτηριστικών της εικόνας ενός ατόμου αποκτάται με την ανάλυση ενός μοντέλου που βασίζεται σε ένα τμήμα, το οποίο χωρίζει την εικόνα ενός ατόμου σε περιοχές G. Η κατανομή της εικόνας ενός ατόμου είναι αυθαίρετη, π.χ. η εκτίμηση που θα χρησιμοποιηθεί στα μέρη του ανθρώπινου σώματος σε αυτή την περίπτωση, οι περιοχές G ενδέχεται να έχουν ληφθεί εκ των προτέρων. Συγκεκριμένα, χρησιμοποιούνται σταθερές οριζόντιες λωρίδες για την ενίσχυση της όψης της ανισομέτρησης της εικόνας. Κάθε περιοχή, χαρακτηρίζεται από ένα εικονοστοιχείο με χαρακτηριστικά χαμηλού επιπέδου όπως το χρώμα του και η κλίση.
Για να εξαχθούν πρόσθετες πληροφορίες χρώματος σε GOG, αντικαθιστώνται οι τιμές των καναλιών RGB με τις τιμές τριών εναλλακτικών χώρων χρώματος {Lab, HSV, nRGB}. Το nRGB είναι ο κανονικοποιημένος χρωματικός χώρος, δεδομένου ότι υπάρχει πλεόνασμα σε αυτό το διάστημα, χρησιμοποιούνται μόνο {nR, nG}. Έτσι, οι διαστάσεις του χώρου χρώματος nR, nG είναι (d, m, r) = (7, 36, 703), για το ZOZ, ενώ οι διαστάσεις κάθε χώρου χρώματος {RGB, Lab, HSV} είναι (d', m', r') = (7, 28, 403) για το GOG. Επομένως, οι διαστάσεις του περιγραφέα σύντηξης για κάθε {GOG, ZOZ} είναι 3 (χρωματικοί χώροι) × G (περιοχές) × r'(dim.) + 1 (χρωματικός χώρος) × G (περιοχές) × r .

5.3.4. Gait Energy Image (GEI)

Η μέθοδος Gait Energy Image [51] παρουσίαζει μια διαδικασία προ-επεξεργασίας στις εξαγόμενες αλληλουχίες σιλουέτας μιας εικόνες. Περιλαμβάνει την κανονικοποίηση του μεγέθους (αναλογική αντιστοίχιση κάθε εικόνας σιλουέτας έτσι ώστε όλες οι σιλουέτες να έχουν το ίδιο ύψος) και την οριζόντια ευθυγράμμιση (κεντράρισμα του άνω μέρους μισού σιλουέτα σε σχέση με το οριζόντιο κέντρο). Σε μια προκατασκευασμένη ακολουθία σιλουέτας, το σήμα χρονοσειράς του κατώτερου μισού μεγέθους σιλουέτας από κάθε πλαίσιο υποδεικνύει τη συχνότητα βάσης και τις πληροφορίες φάσης. Δεδομένων των προ-επεξεργασμένων εικόνων σιλουέτας δυαδικών βημάτων στο χρόνο t σε μια ακολουθία, η εικόνα ενεργειακής απόστασης γκρίζου επιπέδου (GEI) ορίζεται ως εξής:

$$G(x,y) = \frac{1}{N}\sum_{t=1}^{N} B_t(x,y) \qquad (5.1)$$

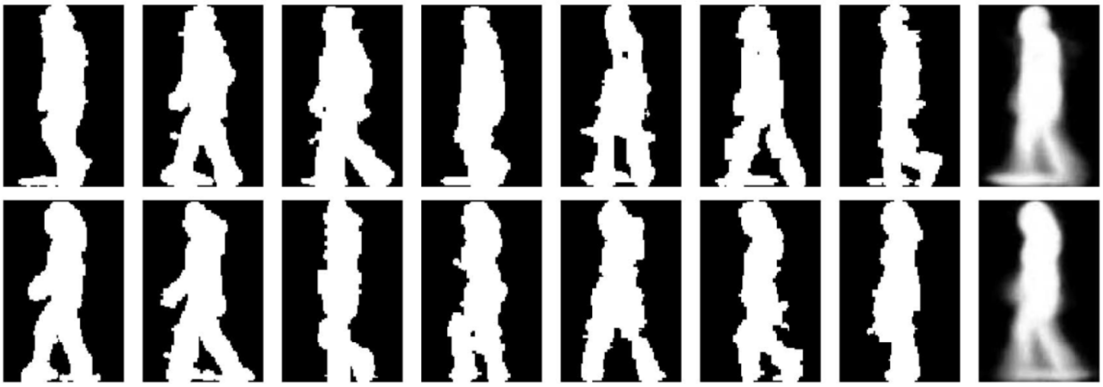
όπου N είναι ο αριθμός πλαισίων στον πλήρη κύκλο(ους) μιας αλληλουχίας σιλουέτας και t είναι ο αριθμός πλαισίου στην ακολουθία (χρονική στιγμή). Τα x και y είναι τιμές συντεταγμένων της 2D εικόνας. Το σχήμα που ακολουθεί δείχνει τις εικόνες σιλουέτας του δείγματος σε έναν κύκλο πορείας από δύο άτομα και η πιο δεξιά εικόνα είναι η αντίστοιχη GEI. Όπως αναμενόταν, η GEI αντικατοπτρίζει σημαντικά σχήματα από σιλουέτες και τις αλλαγές τους κατά τη διάρκεια του κυκλικού βηματισμού. Αναφέρεται σε αυτό ως εικόνα ενέργειας βηματισμού επειδή:

1)Η κάθε ένδειξη σιλουέτας είναι η εξομαλούμενη από το χώρο κίνησης του ανθρώπου στο διάστημα του ανθρώπινου περπατήματος τη στιγμή εκείνη.

2)Η GEI είναι η κανονικοποιημένη στο χρόνο εικόνα του ανθρώπινου περπατήματος στον πλήρη κύκλο

3)Ένα εικονοστοιχείο με τιμή υψηλότερης έντασης στο GEI σημαίνει ότι το βάδισμα του ανθρώπου συμβαίνει πιο συχνά στη θέση αυτή (δηλαδή, με υψηλότερη ενέργεια).

Οι Bobick και Davis [8] προτείνουν την εικόνα κίνησης-ενέργειας (MEI) και την εικόνα ιστορικού κίνησης (MHI) για την αναπαράσταση και αναγνώριση του ανθρώπινου τύπου κίνησης. Τόσο η MEI όσο και η MHI είναι διανυσματικές εικόνες, όπου η τιμή του διανύσματος σε κάθε εικονοστοιχείο είναι συνάρτηση των ιδιοτήτων κίνησης σε αυτή τη θέση σε μια ακολουθία εικόνων. Σε σύγκριση με το MEI και το MHI, η GEI στοχεύει σε συγκεκριμένη κανονική αναπαράσταση του ανθρώπινου βαδίσματος και το GEI χρησιμοποιείται ως πρότυπο αναγνώρισης ατομικού βαδίσματος.



Σχήμα 5-8 Παραδείγματα πλαισίων σιλουέτας σε διαφορετικές ακολουθίες βάδισης. Η εικόνα σε κάθε σειρά είναι η αντίστοιχη εικόνα ενέργειας βηματισμού (GEI).

Ο αριθμός των ακολουθιών εκπαίδευσης για κάθε άτομο είναι περιορισμένος (ένας ή περισσότεροι) σε εφαρμογές πραγματικής επιτήρησης. Αυτό καθιστά δύσκολη την αναγνώριση ατόμων υπό διάφορες συνθήκες που δεν εμφανίζονται στα δεδομένα. Μία λύση σε αυτό το πρόβλημα, είναι να μετρηθεί άμεσα η ομοιότητα μεταξύ των προτύπων εκπαίδευσης και ανίχνευσης. Ωστόσο, η άμεση αντιστοίχιση των προτύπων είναι ευαίσθητη στις παραμορφώσεις σιλουέτας, όπως οι μεταβολές κλίμακας και μετατόπισης. Η εκμάθηση στατιστικών χαρακτηριστικών μπορεί να εξάγει τις εγγενείς ιδιότητες των προτύπων εκπαίδευσης από ένα άτομο και, ως εκ τούτου, θα είναι λιγότερο ευαίσθητη σε αυτή τη στρεβλωτική σκιαγραφία. Ωστόσο, με πρότυπα βηματισμού που λαμβάνονται υπό παρόμοιες συνθήκες, τα χαρακτηριστικά μάθησης μπορεί να ξεπεράσουν τα δεδομένα εκπαίδευσης. Επομένως, για να ξεπεραστούν αυτά τα προβλήματα, δημιουργούμε δύο σύνολα προτύπων βάδισης: 1)πραγματικά πρότυπα και 2)συνθετικά πρότυπα. Τα πρότυπα πραγματικής βάσης για ένα άτομο υπολογίζονται απευθείας από κάθε κύκλο της αλληλουχίας σιλουέτας αυτού του ατόμου. Έστω $\{R_i\}$, $i = 1,..,nR$ είναι το πραγματικό σύνολο προτύπων GEI του ατόμου, όπου nR είναι ο αριθμός κύκλων που ολοκληρώνεται στην αλληλουχία της σιλουέτας. Αν και τα πραγματικά πρότυπα βηματισμού παρέχουν συμβουλές για ατομική αναγνώριση, όλα τα πρότυπα από την ίδια ακολουθία λαμβάνονται υπό τις «ίδιες» φυσικές συνθήκες. Αν αλλάξουν οι συνθήκες, οι γνωστές λειτουργίες μπορεί να μην λειτουργούν καλά για αναγνώριση. Διάφορες συνθήκες επηρεάζουν την εμφάνιση σιλουέτας από το ίδιο πρόσωπο όπως η επιφάνεια πεζοπορίας, παπούτσι, ρούχα κλπ. Η παραμόρφωση της σιλουέτας στο κάτω μέρος της, συμβαίνει στις περισσότερες συνθήκες.

Αυτό το είδος παραμόρφωσης περιλαμβάνει σκιές, έλλειψη τμημάτων του σώματος και διαδοχικές μεταβολές κλίμακας σιλουέτας. Για παράδειγμα, σιλουέτες στην επιφάνεια του γρασιδιού μπορεί να χάσουν το κάτω μέρος των ποδιών, ενώ οι σιλουέτες στην επιφάνεια του σκυροδέματος μπορεί να περιέχουν ισχυρές σκιές. Σε αυτές τις περιπτώσεις, προκύπτουν σφάλματα εξομάλυνσης μεγέθους σιλουέτας και οι σιλουέτες που αποκτήθηκαν με αυτόν τον τρόπο μπορεί να έχουν διαφορετικές κλίμακες σε σχέση με τις σιλουέτες σε άλλες επιφάνειες. Ως εκ τούτου, παράγεται μια σειρά συνθετικών προτύπων βηματισμού που είναι λιγότερο ευαίσθητες στην παραμόρφωση στο κάτω μέρος της σιλουέτας και τις μικρές μεταβολές κλίμακας σιλουέτας. Σημειώνεται ότι μια προσέγγιση ανίχνευσης κινούμενων αντικειμένων μπορεί επίσης να παρέχει πληροφορίες σχετικά με τον τύπο υλικού στον οποίο το άτομο περπατά.

5.3.5. gBiCov

Η τεχνική gBiCov [50], περιλαμβάνει τρία βήματα. Στο πρώτο βήμα, περιέχεται η εξαγωγή βιολογικών χαρακτηριστικών (BIF) [49]. Τα BIF βασίζονται στη μελέτη του ανθρώπινου οπτικού συστήματος και έχουν δείξει άριστες επιδόσεις σε διάφορα θέματα υπολογιστικής όρασης. Συγκεκριμένα, χρησιμοποιείται το στρώμα S1 (φίλτρα Gabor) και το στρώμα C1 (MAX operator) του BIF. Ενώ τα φίλτρα Gabor μπορούν να βελτιώσουν την ευρωστία με παραλλαγές φωτισμού, η μεταβλητή MAX αυξάνει την ανοχή στις μεταβολές της κλίμακας και τις μετατοπίσεις της εικόνας.

Στο δεύτερο βήμα, χρησιμοποιείται ένας περιγραφικός συντελεστής για τον υπολογισμό της ομοιότητας των χαρακτηριστικών BIF που λαμβάνονται σε γειτονικές κλίμακες. Οι περιγραφικοί χαρακτήρες μπορούν να συλλάβουν πληροφορίες σχετικά με το σχήμα, την τοποθεσία, και το χρώμα και η απόδοσή τους έχει αποδειχθεί ότι είναι καλύτερη από άλλες μεθόδους σε πολλές καταστάσεις, καθώς οι μεταβολές και οι αλλαγές φωτισμού απορροφώνται εν μέρει.

Το BIF, με βάση τη μελέτη του ανθρώπινου οπτικού συστήματος, έχει προταθεί για την αντιμετώπιση πολλών καθηκόντων μηχανικής όρασης, όπως η αναγνώριση κατηγορίας αντικειμένων, αναγνώριση προσώπου, εκτίμηση ηλικίας , για τα οποία επιτυγχάνει εξαιρετική απόδοση. Η μέθοδος αυτή [40] υπολογίζει τη συνέλιξή του με το φίλτρο Gabor ψ(z) σύμφωνα με την ακόλουθη εξίσωση:

$$G(\mu,\nu) = \mathrm{I}(z) * \psi_{\mu\nu}(z) \qquad (5.2)$$

όπου

$$\psi_{\mu\nu}(z) = \frac{\left\|\kappa_{\mu\nu}\right\|^2}{\sigma^2} e^{\left(\frac{\left\|\kappa_{\mu\nu}\right\|^2 \left\|z\right\|^2}{\mu}\right)} \left[e^{i\kappa_{\mu\nu}{}^2} - e^{\frac{-\mu^2}{\nu}} \right] \qquad (5.3)$$

Κατά τη διάρκεια του δεύτερου βήματος, κάθε εικόνα μεγέθους BIF χωρίζεται σε μικρές επικαλυπτόμενες περιοχές. Με αυτόν τον τρόπο διατηρούνται οι χωρικές πληροφορίες των εικόνων. Στη συνέχεια, κάθε περιοχή αντιπροσωπεύεται από έναν περιγραφέα συνδιασποράς. Για κάθε εικονοστοιχείο της εικόνας BIFmagnitude, υπολογίζεται ένα δισδιάστατο διάνυσμα B_i χαρακτηριστικών για την καταγραφή των στατιστικών έντασης, υφής και σχήματος.

Τα χαρακτηριστικά μεγέθους BIF μπορούν να θεωρηθούν ως χαρακτηριστικά με βάση την εμφάνιση, ενώ οι μήτρες συν-διακύμανσης μπορούν να θεωρηθούν ως περιγραφή ιδιοτήτων χαρακτηριστικών. Σε κάποιο βαθμό, τα χαρακτηριστικά μεγέθους BIF και οι πίνακες συν-διακύμανσης είναι δύο διαφορετικά επίπεδα ολόκληρης της αναπαράστασης.

Πρώτον, ο συνδυασμός φίλτρων Gabor με περιγραφικούς συντελεστές καθιστά το gBiCov πολύ ισχυρό σε παραλλαγές φωτισμού. Από την άλλη πλευρά ο περιγραφικός συντελεστής απορροφά επίσης τις αλλαγές των φωτισμών. Δεδομένου ότι ο συνδυασμός των φίλτρων Gabor και του περιγραφικού συντελεστή συνάφειας, το gBiCov μπορεί να αποδειχθεί ότι είναι ακόμα πιο ανθεκτικό στις διακυμάνσεις του φωτισμού. Δεύτερον, το gBiCov είναι επίσης πιο ανθεκτικό στις παραλλαγές υποβάθρου, δηλαδή μπορεί να επιτύχει καλές επιδόσεις (χωρίς ακριβή κατακερματισμό στο προσκήνιο / παρασκήνιο) στην ανίχνευση των τμημάτων σώματος, τα οποία είναι συχνά δύσκολες εργασίες. Σε γενικές γραμμές, οι περιοχές του συνεισφέρουν λιγότερο στην υφή, γεγονός που καθιστά τις ιδιότητές τους Gabor (και συνεπώς τους περιγραφικούς συντελεστές) σε διαφορετικές γειτονικές κλίμακες παρόμοιες. Δεδομένου ότι ο περιγραφέας του gBiCov βασίζεται στη διαφορά των περιγραφικών συντεταγμένων σε διαφορετικές κλίμακες, οι περιγραφείς gBiCov που εξάγονται από τις περιοχές υποβάθρου είναι μικρές και δεν επηρεάζουν την εξαγωγή χαρακτηριστικών.

5.4. Τεχνικές Μάθησης Μετρικής στο Person Re-id

Η χρήση οποιασδήποτε τεχνικής εξαγωγής χαρακτηριστικών που περιγράφονται στην προηγούμενη ενότητα σε συνδυασμό με την ευκλείδεια απόσταση (l2) μπορεί να χρησιμοποιηθεί για την κατάταξη των υποψηφίων εικόνων. Η ενσωμάτωση της μετρικής μάθησης (metric learning) με τη χρήση δεδομένων εκπαίδευσης οδηγεί σε ανώτερη απόδοση, η οποία είναι και ο στόχος κάθε μεθόδου δηλ. η εκμάθηση ενός νέου χώρου χαρακτηριστικών. Στην ενότητα αυτή περιγράφονται οι βασικές τεχνικές οι οποίες χρησιμοποιούνται στο κομμάτι αυτό.

5.4.1. Cross-view Quadratic Discriminant Analysis (XQDA)

Συνήθως, οι αρχικές διαστάσεις d είναι μεγάλες και ο χώρος χαμηλών διαστάσεων R_r όπου $r < d$ προτιμάται για την ταξινόμηση των μεθόδων. Στην τεχνική αυτή [65] προτείνεται να αποσυντεθούν χωριστά οι εικόνες για να μειωθούν οι διαστάσεις. Οι μέθοδοι αυτές δεν είναι βέλτιστες λόγω της μείωσης των διαστάσεων καθώς δεν θεωρούν τη μάθηση απαραίτητο στοιχείο.

Σε αυτή την μέθοδο εξετάζεται μια υποπεριοχή $W = (w_1, w_2, ..., w_r) \in R_{d \times r}$ με δεδομένα σταυρωτής προβολής (cross-view) και ταυτόχρονα για την εκπαίδευση του συστήματος χρησιμοποιείται ο υποδιαχωρισμός της διάστασης r για την εγκάρσια προβολή του μέτρου ομοιότητας. Ας υποθέσουμε ότι έχουμε μια διασταυρούμενη προβολή που ορίζεται ως {X, Z}, όπου $X = (x_1, x_2, ..., x_n) \in R_{d \times n}$ και περιέχει n δείγματα d διαστάσεων και μία άποψη, $Z = (z_1, z_2, ..., z_m) \in R_{d \times m}$ περιέχει m δείγματα στον ίδιο χώρο d διαστάσεων αλλά από άλλη οπτική γωνία. Το πρόβλημα αντιστοίχισης σταυρωτής προβολής προκύπτει σε πολλές εφαρμογές, όπως η ετερογενής αναγνώριση προσώπου και η αλλαγή ταυτότητας του αμετάβλητου ατόμου. Σημειώνεται ότι το Z είναι το ίδιο με το X στο σενάριο αντιστοίχισης μιας προβολής. Λαμβάνοντας υπόψη μια υποπεριοχή W, η συνάρτηση απόστασης στο υποδιάστημα r υπολογίζεται ως:

$$d_w(x,z)=(x-z)^T W(\Sigma_1^{`-1}-\Sigma_E^{`-1})W^T(x-z) \tag{5.4}$$

5.4.2. RankSVM

Η κατάταξη μπορεί να βασίζεται είτε στη βελτίωση, είτε στη μάθηση βάσει πυρήνα, όπως οι μηχανές υποστήριξης διανυσμάτων (SVM). Το RankBoost [109] χρησιμοποιεί ένα σύνολο αδύναμων ταξινομητών για να σχηματίσει έναν ισχυρό ταξινομητή. Ωστόσο, καθώς το πρόβλημα επαναταυτοποίησης πάσχει εγγενώς από ένα μεγάλο βαθμό αλληλεπικάλυψης χαρακτηριστικών σε ένα πολυδιάστατο χαρακτηριστικό χώρο, η επιλογή των αδύναμων ταξινομητών σε κάθε μεμονωμένη διάσταση χαρακτηριστικών, όπως θεωρείται από το [110], είναι πιθανό να οδηγήσει σε πολύ αδύναμους παράγοντες μειώνοντας την αποτελεσματικότητα της αντιστοίχησης. Αντίθετα, τα μοντέλα SVM όπως το RankSVM [111] επιδιώκουν να μάθουν μια συνάρτηση κατάταξης σε ένα χώρο υψηλότερων διαστάσεων, όπου οι σωστές και οι λάθος αντιστοιχήσεις διαχωρίζονται με μεγαλύτερη επιτυχία από τον αρχικό χώρο των χαρακτηριστικών μέσω του κόλπου του πυρήνα (kernel trick). Το RankSVM είναι επομένως δυνητικά πιο αποτελεσματικό για την αντιμετώπιση των εξαιρετικά αλληλοεπικαλυπτόμενων κατανομών χαρακτηριστικών κατά την επαναταυτοποίηση των προσώπων.

Οι Chapelle και Keerthi [112] πρότειναν ένα RankSVM (PRSVM) βασισμένο στην εκπαίδευση πρώτου βαθμού, το οποίο θα επιταχύνει την υπάρχουσα ανάκτηση εγγράφων. Αυτό το μοντέλο PRSVM χρησιμοποιείται για την αντιμετώπιση του προβλήματος επαναταυτοποίησης προσώπου και δείχνει ότι το PRSVM εξακολουθεί να υποφέρει από ένα ακόμη πρόβλημα περιορισμού της κλιμάκωσης. Συγκεκριμένα, καθώς αυξάνεται ο αριθμός των δειγμάτων κατάρτισης, ο αριθμός των αρνητικών δειγμάτων αυξάνεται μη γραμμικά, σε συνδυασμό με τις μεγάλες διαστάσεις των χαρακτηριστικών, αυτό σημαίνει ότι η κατανάλωση μνήμης μπορεί να καταστεί μη διαχειρίσιμη. Για την αντιμετώπιση αυτού του προβλήματος, προτείνεται ένα

Ensemble RankSVM που χρησιμοποιεί μια αρχή ενίσχυσης των αδυναμιών των PRSVMs για τη διατήρηση της υπολογιστικής αποδοτικότητας σε ένα χώρο με χαρακτηριστικά υψηλών διαστάσεων, ενώ ταυτόχρονα ξεπερνά το πρόβλημα περιορισμού της κλιμάκωσης των PRSVM όσον αφορά τη χρήση της μνήμης. Ένα πρόσθετο πλεονέκτημα αυτού του μοντέλου είναι ότι ενσωματώνει το βήμα ρύθμισης των παραμέτρων του PRSVM σε ένα πλαίσιο ενίσχυσης που καταργεί την ανάγκη επαναδημιουργίας συνόλων εκπαίδευσης και επικύρωσης. Ένα μοντέλο RankSVM ορίζεται στη συνέχεια ως ελαχιστοποίηση της ακόλουθης αντικειμενικής συνάρτησης:

$$\frac{1}{2}\|w\|^2 + C\sum_{s=1}^{|P|}\xi_s \qquad \text{Εξ. 5.5}$$

$$s.t.\ w^T(\hat{x}_s^+ - \hat{x}_s^-) \geq 1 - \xi_s,\ s = 1,\ldots,|P|, \xi_s \geq 0,\ s = 1,\ldots.,|P| \qquad \text{Εξ. 5.6}$$

όπου το C είναι μια θετική παράμετρος που παρουσιάζει το μέγεθος περιθωρίου έναντι του σφάλματος κατάρτισης. Ένα από τα κύρια προβλήματα χρήσης ενός SVM για την επίλυση του προβλήματος κατάταξης, είναι το δυνητικά μεγάλο μέγεθος του P. Σε προβλήματα με πολλά ερωτήματα ή / και ερωτήσεις με πολλούς συνδεόμενους φορείς παρατηρήσεως, το μέγεθος του P σημαίνει ότι ο σχηματισμός του διανύσματος, $\hat{x}_s^+ - \hat{x}_s^-$ γίνεται με συμαντική υπολογιστική πρόκληση. Ιδιαίτερα, στην περίπτωση επαναταυτοποίησης προσώπου, υποθέτοντας ότι υπάρχει ένα σετ κατάρτισης που αποτελείται από εικόνες m ατόμων σε δύο προβολές κάμερας. Το μέγεθος του P είναι ανάλογο με το m^2, επομένως αυξάνεται αρκετά γρήγορα καθώς αυξάνεται το m. Οι μέθοδοι που βασίζονται σε SVM βασίζονται επίσης στην παράμετρο C, η οποία πρέπει να είναι γνωστή πριν από την εκπαίδευση. Για να αποδοθεί ένα λογικό μοντέλο, πρέπει να χρησιμοποιηθεί η εγκάρσια επικύρωση για να συντονιστούν οι παράμετροι του μοντέλου. Αυτό το βήμα απαιτεί την ανοικοδόμηση του σετ κατάρτισης / επικύρωσης σε κάθε επανάληψη, αυξάνοντας έτσι περαιτέρω το υπολογιστικό κόστος και τη χρήση της μνήμης. Ως εκ τούτου, το RankSVM στην παραπάνω εξίσωση δεν είναι υπολογιστικά ευέλικτο στον υπολογίσμό προβλημάτων μεγάλης κλίμακας, λόγω του σημαντικού υπολογιστικού κόστους όσο και της χρήσης της μνήμης.

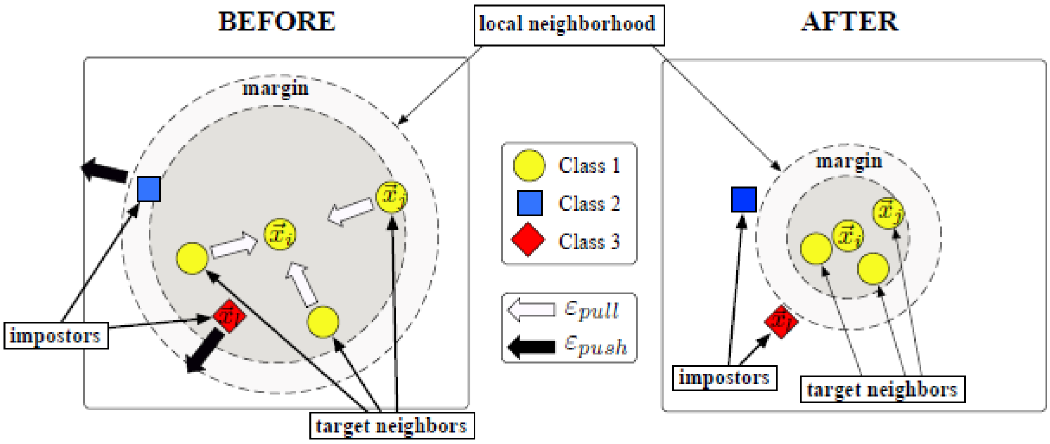
5.4.3. Large Margin Nearest Neighbor Classification (LMNN)

Η τεχνική αυτή (Large Margin Nearest Neighbor Classification, LMNN) [46] μπορεί να θεωρηθεί ως το λογικό αντιστάθμισμα των SVM. Έστω $\{(x_i,y_i)\}$ i=1,..,n δηλώνοντας ένα σετ κατάρτισης n σημασμένων παραδειγμάτων και διακριτές (αλλά όχι απαραίτητα δυαδικές) ετικέτες y_i. Χρησιμοποιείται ο δυαδικός πίνακα $y_{ij} \in \{0, 1\}$ για να υποδειχθεί το πότε οι ετικέτες y_i, y_j ταιριάζουν ή όχι. Στόχος αυτού, είναι η εκμάθηση ενός γραμμικού μετασχηματισμού ο οποίος θα χρησιμοποιηθεί για να υπολογιστούν οι τετραγωνικές αποστάσεις:

$$D(\overrightarrow{x_i}, \overrightarrow{x_j}) = \left\| L(\overrightarrow{x_i} - \overrightarrow{x_j}) \right\|^2 \qquad \text{Εξ. 5.7}$$

Εκτός από την ετικέτα κλάσης y_i, για κάθε είσοδο x_i επίσης προσδιορίζονται οι k "γειτονικοί στόχοι", δηλ. άλλες είσοδοι με την ίδια ετικέτα y_i που είναι επιθυμητό να έχουν ελάχιστη απόσταση από το x_i, όπως υπολογίζεται από την παραπάνω εξίσωση. Σε περίπτωση έλλειψης προηγούμενης γνώσης, οι γείτονες-στόχοι μπορούν απλά να αναγνωριστούν ως οι πλησιέστεροι γείτονες, που καθορίζονται από την ευκλείδεια απόσταση, που μοιράζονται την ίδια ετικέτα y_i. Αυτό έγινε και σε όλα τα πειράματα αυτής της παραπάνω εργασίας. Χρησιμοποιείται το $n_{ij} \in \{0, 1\}$ για να υποδειχθεί αν η είσοδος x_j είναι ένας γείτονας στόχος της εισόδου x_i. Όπως και ο δυαδικός πίνακας y_{ij}, ο πίνακας n_{ij} είναι σταθερός και δεν αλλάζει κατά τη διάρκεια της μάθησης.

Το σχήμα που ακολουθεί απεικονίζει την κύρια ιδέα της ταξινόμησης του LMNN. Κατά τη διάρκεια της μάθησης, τα «λάθος παραδείγματα» ωθούνται έξω από την περίμετρο που καθορίζουν οι γείτονες-στόχοι. Μετά την εκμάθηση, υπάρχει ένα πεπερασμένο περιθώριο μεταξύ των «σωστών» και των «λάθος» παραδειγμάτων. Το σχήμα δείχνει το εξιδανικευμένο σενάριο όπου τα σφάλματα ταξινόμησης kNN στον αρχικό χώρο εισόδου διορθώνονται με την εκμάθηση ενός κατάλληλου γραμμικού μετασχηματισμού.Με τη διαίσθηση και την ορολογία από την προηγούμενη ενότητα, μπορεί τώρα να κατασκευαστεί μια συνάρτηση απώλειας για την ταξινόμηση LMNN. Η συνάρτηση απώλειας αποτελείται από δύο όρους, ένας που δρα για να τραβήξει τους γείτονες-στόχους πιο κοντά και έναν άλλο που ενεργεί για να απωθήσει τα διαφορετικά επισημασμένα παραδείγματα.



Σχήμα 5-9 Σχηματική απεικόνιση της γειτονιάς μίας εισόδου πριν την εκπαίδευση (αριστερά) και μετά την εκπαίδευση (δεξιά). Η μέτρηση της απόστασης βελτιστοποιείται έτσι ώστε: (i) οι γείτονες-στόχοι να βρίσκονται σε μικρότερη ακτίνα μετά την εκπαίδευση (ii) οι διαφορετικά επισημασμένες είσοδοι να βρίσκονται εκτός της μικρότερης ακτίνας από κάποιο πεπερασμένο περιθώριο.

5.4.4. KISSME

Η μέθοδος αυτή [43] εξετάζει δύο διαδικασίες ανεξάρτητης παραγωγής για κοινότητες παρόμοιων και ανόμοιων ζευγών εικόνων. Η ανομοιογένεια ορίζεται από την ικανότητα ενός χαρακτηριστικού να ανήκει σε μια κατηγορία ή σε μια άλλη. Από στατιστικής άποψης η απόφαση εάν ένα ζεύγος (i, j) είναι ανόμοιο ή όχι μπορεί να ληφθεί με πιθανότητα δοκιμής αναλογίας. Έτσι, γίνεται η υπόθεση ότι ένα ζεύγος H0 είναι ανόμοιο με ένα άλλο, το H1:

$$\delta(x_i, x_j) = \log\left(\frac{p(x_i, x_j \mid H0)}{p(x_i, x_j \mid H1)}\right) \qquad \text{Εξ. 5.8}$$

Μια υψηλή τιμή του (x_i, x_j) σημαίνει ότι η H0 είναι επικυρωμένη. Αντίθετα, μια χαμηλή τιμή σημαίνει ότι η H0 απορρίπτεται και τα ζεύγη θεωρούνται παρόμοια. Για να υπάρχει ανεξαρτησία από την πραγματική τοποθεσία στο χώρο των χαρακτηριστικών, τοποθετείται το πρόβλημα στο χώρο των ζευγών ($x_{ij} = x_i - x_j$) με μέση τιμή μηδέν, επομένως η παραπάνω εξίσωση αλλάζει και γίνεται:

$$\delta(x_{ij}) = \log\left(\frac{p(x_{ij} \mid H0)}{p(x_{ij} \mid H1)}\right) = \log\left(\frac{p(x_{ij} \mid \theta 0)}{p(x_{ij} \mid \theta 1)}\right) \qquad \text{Εξ. 5.9}$$

Οπότε το $f(x_{ij}|\theta_1)$ είναι ένα pdf με τις παραμέτρους για την υπόθεση H1 που παρουσιάζουν ότι ένα ζεύγος (i,j) είναι παρόμοιο (y_{ij} = 1) και αντίστροφα ότι H0 για ένα ζευγάρι που είναι διαφορετικό. Υποθέτοντας μια Gaussian δομή του χώρου διαφοράς μπορούμε να ορίσουμε το πρόβλημα ως:

$$\delta(x_{ij}) = \log\left(\frac{\frac{1}{\sqrt{2\pi \mid \Sigma_{y_{ij=0}} \mid}} \exp(-1/2 x^T{}_{ij} \Sigma^{-1}{}_{y_{ij=0}} x_{ij})}{\frac{1}{\sqrt{2\pi \mid \Sigma_{y_{ij=1}} \mid}} \exp(-1/2 x^T{}_{ij} \Sigma^{-1}{}_{y_{ij=1}} x_{ij})}\right) \qquad \text{Εξ. 5.10}$$

Τα ζεύγη διαφορών x_{ij} είναι συμμετρικά. Έτσι προκύπττει μηδενική μέση τιμή. Η μέγιστη εκτίμηση πιθανότητας του Gaussian είναι ισοδύναμη με την ελαχιστοποίηση των αποστάσεων Mahalanobis σε ένα ελάχιστο τετράγωνο. Αυτό μας επιτρέπει να βρούμε τις σχετικές οδηγίες για τα δύο ανεξάρτητα σύνολα. Τέλος, λαμβάνοντας υπόψη ότι η μέτρηση απόστασης Mahalanobis αντανακλά τις ιδιότητες της δοκιμασίας του log-probability ratio, προκύπτει η τελική εξίσωση της πατακάτω μετρικής:

$$d_M^2(x_i, x_j) = (x_i - x_j)^T M (x_i - x_j) \qquad \text{Εξ. 5.11}$$

5.4.5. Local Fisher Discriminant Analysi1s (LFDA)

Ένας πίνακας περιγραφών ορίζεται ως X ≡ (x_1 | x_2 | ... | x_n), και αποτελείται από n δείγματα C διαφορετικών τάξεων με ετικέτες x_i ∈ {1, ... ,C} και φέρει διανύσματα x_i ∈ Rd. Στη συνέχεια ορίζεται ένας πίνακας μετασχηματισμού T έτσι ώστε [68]:

$$\mathbf{z}_i = T'\mathbf{x}_i \tag{5.12}$$

όπου $z_i \in R_m$ ($1 \leq m \leq d$) είναι ένας μικρότερος χώρος διαστάσεων, με διαστάσεις m. Στην ανάλυση διακρίσεων Fisher (Local Fisher Discriminant Analysis, LFDA), οι πίνακες σκέδασης μεταξύ των τάξεων και μέσα σε μια κλάση ορίζονται ως:

$$S^{(b)} = \sum_{i=1}^{N} n_i (\mu_i - \mu)(\mu_i - \mu)' \tag{5.13}$$

$$S^{(w)} = \sum_{i=1}^{N} \sum_{j:x_j=i} (\mathbf{x}_j - \mu_i)(\mathbf{x}_j - \mu_i)' \tag{5.14}$$

αντίστοιχα. Εδώ, το μ_i είναι ο μέσος όρος των δεδομένων που ανήκουν στην κλάση i και μ είναι ο μέσος όλων των δεδομένων. Στη συνέχεια, ένα πρόβλημα Eigenvalue μπορεί να διαμορφωθεί χρησιμοποιώντας $S^{(w)}$ και $S^{(b)}$ ως $S^{(b)} \varphi_i = \lambda_i S^{(w)} \varphi_i$, όπου $\{\varphi_i\}$ είναι τα ιδιοδιανύσματα που συνδέονται με τις ιδιοτιμές $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_d$. Παρόλο που η LFDA είναι μια εποπτευόμενη διαδικασία, για τις πολυτροπικές διανομές η απόδοση τείνει να είναι κακή. Ένας πιθανός λόγος για αυτό είναι ότι επεξεργάζεται τα δεδομένα στο σύνολό τους και έτσι οι τοπικές παραλλαγές στον μετασχηματισμό δεν μπορούν να προσαρμοστούν. Η προβολή διατήρησης της περιοχής (LPP) χρησιμοποιεί μια μήτρα συγγένειας n*n, η οποία περιγράφει τη συγγένεια μεταξύ διαφόρων δειγμάτων εντός των δεδομένων. Συνήθως, η τιμή συγγένειας μεταξύ δύο δειγμάτων που διαχωρίζονται από μια μικρή ευκλείδεια απόσταση είναι υψηλότερη από δύο δείγματα που διαχωρίζονται από μεγαλύτερη απόσταση. Έστω D ένα διαγωνικό πλέγμα: κάθε στοιχείο αθροίζει όλες τις τιμές συγγένειας πάνω από τις στήλες του A, έτσι ώστε D_{ii} να είναι η πηγή δεδομένων, η πιθανότητα ότι ύστερα, το μέλος στη βαθμίδα r θα είναι το σωστό στην αντιστοιχία, είναι ίσο με το χαρακτηριστικό αντιστοίχισης M(r). Η αναμενόμενη αβεβαιότητα αυτής της κατάστασης είναι ως εκ τούτου $\sum_{r=1}^{S} M(r)\log(M(r))$.

Η διαφορά μεταξύ αυτού και του αρχείου καταγραφής (S) είναι η πληροφορία που προστίθεται από την εκτίμηση ταυτότητας. Επιπλέον, είναι χρήσιμο να κανονικοποιηθεί αυτή τη μείωση της εντροπίας από την αρχική εντροπία: αυτή η «αναλογία αδυναμίας αφαίρεσης» (PUR) είναι αμετάβλητη στο μέγεθος του συνόλου δοκιμών S και η βάση του λογαρίθμου που χρησιμοποιείται για την εντροπία:

$$PUR = \frac{\log(S) - \sum_{r=1}^{S} M(r)\log(M(r))}{\log(S)} \tag{5.15}$$

Έτσι, το PUR είναι η κανονικοποιημένη μείωση της εντροπίας μεταξύ μιας τυχαίας σειράς και της κατάταξης εξόδου κάθε δεδομένης μεθόδου. Επίσης φιλοξενεί πληροφορίες από ολόκληρο το CMC, παρά από αυθαίρετες τιμές του r.

5.4.6. k - Local Fisher Discriminant Analysis (kLFDA)

Ένα μειονέκτημα της χρήσης του LFDA [47] είναι ότι απαιτεί την επίλυση ενός γενικευμένου προβλήματος ιδιοτιμής σε πολύ μεγάλες μήτρες d × d. Για παράδειγμα, στο [68] οι συγγραφείς χρησιμοποιούν διανύσματα χαρακτηριστικών με d = 22506 χαρακτηριστικά. Για να παρακαμφθεί αυτό το πρόβλημα, προτείνεται η αξιοποίηση της πλεοναστικότητας μεταξύ των χαρακτηριστικών, με την πραγματοποίηση ενός βήματος προ-επεξεργασίας όπου η κύρια ανάλυση συστατικών (PCA) χρησιμοποιείται για να μειώσει τη διαστατικότητα των δεδομένων. Εντούτοις, μια πιθανή δυσκολία εδώ είναι ότι αυτό το βήμα μείωσης της διαστολής χωρίς επίβλεψη, όταν εφαρμόζεται σε σχετικά μικρά σύνολα δεδομένων, μπορεί να οδηγήσει σε ανεπιθύμητη συμπίεση των πλέον διακριτικών χαρακτηριστικών.

Για να αποφευχθεί αυτό το πρόβλημα, προτείνεται η χρήση μιας προσέγγισης πυρήνα, με βάση τη μέθοδο που εισήχθη στο [113] στο πλαίσιο της εποπτευόμενης μείωσης των διαστάσεων. Τα οφέλη αυτής της προσέγγισης είναι διπλά: αποφεύγει την εκτίμηση της ιδιοτιμής των μεγάλων πινάκων σκέδασης και μπορεί να εκμεταλλευτεί την ευελιξία στην επιλογή του πυρήνα για να βελτιώσει την ακρίβεια ταξινόμησης. Η προτεινόμενη μέθοδος kernel LDFA (kLDFA) βρίσκει έναν πίνακα προβολής P.

$$P = \max_{P}(PS^{w}P)^{-1}P^{T}S^{b}P \tag{5.16}$$

$$S^{w} = \frac{1}{2}\varphi(X)\tilde{S}^{w}\varphi(\mathrm{X})^{T} \tag{5.17}$$

$$S^{b} = \frac{1}{2}\varphi(X)\tilde{S}^{b}\varphi(\mathrm{X})^{T} \tag{5.18}$$

$$\tilde{S}^{w} = \sum_{i,j=1}^{N} A_{i,j}^{w}(e_i - e_j)(e_i - e_j)^{T} \tag{5.19}$$

$$\tilde{S}^{b} = \sum_{i,j=1}^{N} A_{i,j}^{b}(e_i - e_j)(e_i - e_j)^{T} \tag{5.20}$$

Η προβολή του πίνακα με τα δείγματα εικόνων στον πυρήνα P, ορίζεται στον αλγόριθμο kLFDA ως:

$$Q = \max_{Q}(|\,QK\tilde{S}^{w}KQ)^{-1}QK\tilde{S}^{b}KQ \tag{5.21}$$

5.4.7. Marginal Fisher Analysis (MFA)

Η ανάλυση Fisher (Marginal Fisher Analysis, MFA) προτάθηκε στο [69] ως μια ακόμη γραφική παράσταση που ενσωματώνει τη μέθοδο μείωσης των διαστάσεων. Ομοίως με το kLDFA και το LDFA, έχει μια λύση κλειστής μορφής που δίνεται από μια γενική αποσύνθεση ιδιοτιμών. Εντούτοις, σε αντίθεση με την LDFA, ο ειδικός αντικειμενικός στόχος του επιτρέπει να μεγιστοποιήσει το διακριτό περιθώριο ακόμα και όταν η υπόθεση της Gaussian κατανομής δεν είναι αληθής. Επιπλέον, τα αποτελέσματα της [70] έδειξαν ότι τα εκπαιδευμένα δεδομένα έχουν μεγαλύτερο περιθώριο μεταξύ κατηγοριών, παρόμοιο με ένα SVM. Οι μήτρες σκέδασης για την μετρική αυτή ορίζονται ως:

$$\tilde{S}^{w} = (D^{w} - W^{w}) \tag{5.22}$$

$$\tilde{S}^{b} = (D^{b} - W^{b}) \tag{5.23}$$

$$D_{ii}^{b} = \sum_{j} W_{ij}^{b} \tag{5.24}$$

$$D_{ii}^{w} = \sum_{j} W_{ij}^{w} \tag{5.25}$$

5.5. Παρουσίαση μεθόδων

Στην ενότητα που ακολουθεί παρουσιάζονται κάποιοι μέθοδοι που χρησιμοποιούνται στην επαναταυτοποίηση ατόμου και κάνουν χρήση κάποιων από των παραπάνω τεχνικών. Οι μέθοδοι παρουσιάζονται αναλυτικά, ώστε να μπορέσουν να τεθούν σε σύγκριση τα χαρακτηριστικά τους μετέπειτα.

5.5.1. Μέθοδος 1: Multi-scale triplet convolutional neural network (CNN)

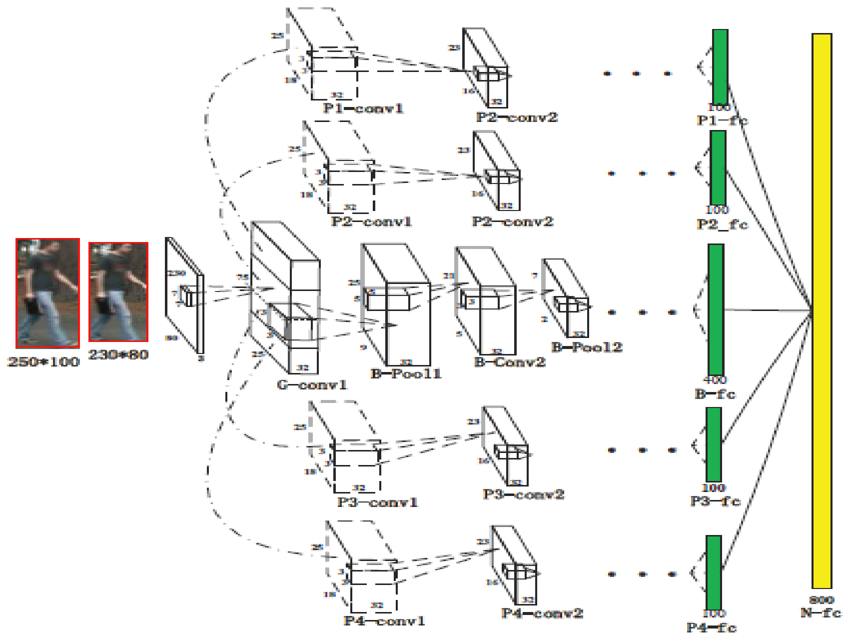
Στη μέθοδο αυτή προτείνεται μια πολυδύναμη τριπλή συνελικτική προσέγγιση νευρωνικού δικτύου (multi-scale triplet convolutional neural network) [71] που καταγράφει την οπτική εμφάνιση ενός άτομου σε διάφορες κλίμακες. Σχεδιάζεται αρχικά ένα ενοποιημένο δίκτυο πολλαπλών μεγεθών, του οποίου η αρχιτεκτονική αποτελείται από νευρικά δίκτυα, προς την κατεύθυνση της εκμάθησης ισχυρών και αποτελεσματικών χαρακτηριστικών.

Προτείνεται επίσης μια ολοκληρωμένη διαδικασία μάθησης "από άκρο σε άκρο" για να εκπαιδευτεί ο αλγόριθμος σε βαθιά χαρακτηριστικά καθώς και για τη μετρική απόστασης. Με την μέθοδο αυτή βελτιστοποιούνται οι παράμετροι του δικτύου οι οποίοι βασίζονται σε συγκριτική ομοιότητα και απώλεια μαζικών δειγμάτων για την αντιμετώπιση του προβλήματος του μικρού δείγματος εκπαίδευσης.

Τα συνεργατικά νευρωνικά δίκτυα (Convolutional Neural Networks, CNN) είναι παρόμοια με τα συνηθισμένα νευρωνικά δίκτυα, καθώς αποτελούνται από νευρώνες που έχουν μαθησιακές βαρύτητες. Το όλο δίκτυο εξακολουθεί να εκφράζει μια ενιαία λειτουργία διαφοροποιήσιμης βαθμολογίας: από τα pixels της πρώτης εικόνας στο ένα άκρο έως τα αποτελέσματα της κατηγορίας στο άλλο.

Το προτεινόμενο μοντέλο πολλαπλών καναλιών CNN αποτελείται κυρίως από τα ακόλουθα διακριτά επίπεδα: ένα στρώμα σφαιρικής συνέλιξης, ένα στρώμα συνέλιξης πλήρους σώματος, τέσσερα στρώματα συνέλιξης σώματος-μέρους, πέντε επίπεδα καναλιών πλήρους σύνδεσης και ένα στρώμα πλήρους σύνδεσης σε δίκτυο.

Όπως φαίνεται στο σχήμα που ακολουθεί, το στρώμα σφαιρικής συνέλιξης είναι το πρώτο στρώμα του CNN. Αποτελείται από 32 χάρτες χαρακτηριστικών με τον πυρήνα συνέλιξης 7 × 7 × 3 και βήμα 3 pixels. Στη συνέχεια, αυτό το στρώμα σφαιρικής συνέλιξης χωρίζεται σε τέσσερα ίσα μέρη P_i, i = {1,…,4}, και κάθε τμήμα Pi σχηματίζει το πρώτο στρώμα ενός ανεξάρτητου καναλιού μέρους-σώματος που στοχεύει στην εκπαίδευση των χαρακτηριστικών για το αντίστοιχο τμήμα του σώματος.



Σχήμα 5-10 Αρχιτεκτονική CNN δικτύου. [71]

Το πλήρες κανάλι με ολόκληρο το στρώμα σφαιρικής συνέλιξης ως το πρώτο του layer είναι επίσης καθιερωμένο για να μάθει τα παγκόσμια χαρακτηριστικά ολόκληρου του σώματος από το άτομο της εισόδου. Το τετραγωνικό κανάλι μαζί με το κανάλι πλήρους σώματος αποτελείται από πέντε ανεξάρτητα κανάλια που εκπαιδεύονται ξεχωριστά το ένα από το άλλο. Το κανάλι πλήρους σώματος διαμορφώνεται ως εξής: Το σφαιρικό στρώμα συνέλιξης φέρει το σύνολο των μέγιστων συγκεντρώσεων. Το μέγεθος του πυρήνα με τη μέγιστη συγκέντρωση είναι 3 × 3 και το στρώμα πλήρους σύνδεσης παράγει μια έξοδο 400 διαστάσεων. [71]

Τα αποτελέσματα της μεθόδου αυτής παρουσιάζονται συγκριτικά στο επόμενο κεφάλαιο το οποίο αποτελεί και τον κυρίως κορμό της εργασίας.

5.5.2. Μέθοδος 2: Αλγόριθμος με χρήση τεχνικής SDALF

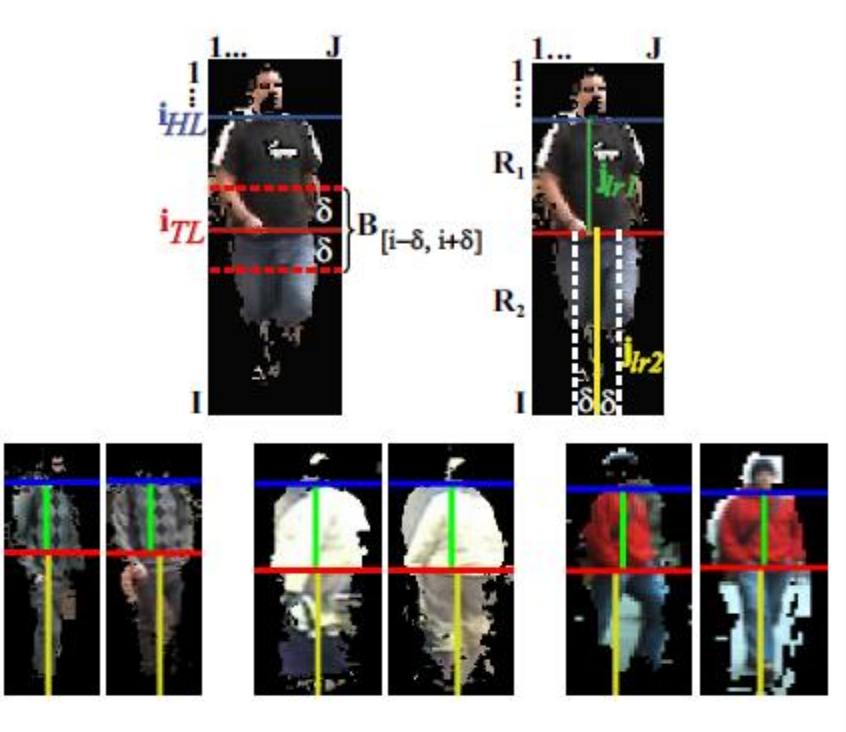
Το όνομα της προσέγγισής είναι Symmetry-Driven Accumulation of Local Features (SDALF) [52]. Η μέθοδος SDALF αποτελείται από μια περιγραφή του ανθρώπινου σώματος που βασίζεται στη συμμετρία και εμπνέεται από τη γνωστή αρχή ότι τα φυσικά αντικείμενα αποκαλύπτουν συμμετρία σε κάποια μορφή. Εφαρμόζεται τόσο στην περίπτωση που υπάρχει μία εικόνα για κάθε υποψήφιο όσο και στην περίπτωση που υπάρχουν πολλές εικόνες για κάθε άτομο (όχι απαραίτητα διαδοχικές). Απαραίτητο στοιχείο αυτών αποτελεί η συγκέντρωση των κατάλληλων τοπικών χαρακτηριστικών για την αναγνώριση: ερευνητικά, όσο μεγαλύτερος είναι ο αριθμός των εικόνων για κάθε άτομο, τόσο μεγαλύτερη είναι η εκφραστικότητα της αναγνώρισης. Για το λόγο αυτό, είναι χρήσιμο να εντοπίζονται και να χαρακτηρίζονται συμμετρίες προκειμένου να γίνεται κατανοητή η δομή των δεδομένων. Αυτός ο ισχυρισμός υποστηρίζεται έντονα από το Gestalt psychology school που θεωρεί τη συμμετρία ως θεμελιώδη αρχή της αντίληψης: τα συμμετρικά στοιχεία είναι πιθανότερο να ενσωματωθούν σε ένα συνεκτικό αντικείμενο από τις ασύμμετρες περιοχές.

Στη μέθοδο SDALF, οι αρχές ασυμμετρίας επιτρέπουν στον διαχωρισμό της εικόνας του ανθρωπίνου σώματος (κεφάλι, άνω μέρος του σώματος, κάτω σώμα). Οι συμμετρίες συμβάλλουν στην εξαγωγή χαρακτηριστικών από το πραγματικό ανθρώπινο σώμα, όπως αυτό απεικονίζεται στην εικόνα. Η ιδέα είναι ότι τα χαρακτηριστικά στον άξονα κάθετης συμμετρίας ζυγίζονται περισσότερο από αυτά που βρίσκονται μακριά από αυτό, προκειμένου να αποκτήσουν πληροφορίες από το εσωτερικό μέρος του σώματος.

Η προσέγγιση της μεθόδου αυτής είναι απλή και αποτελεσματική, καθώς βασίζεται στην εμφάνιση και την εξαγωγή τοπικών χαρακτηριστικών. Μετά από ένα βήμα προ-επεξεργασίας, επιλέγουμε τα σημαντικότερα τμήματα της φιγούρας του σώματος, υιοθετώντας αρχές συμμετρίας και ασυμμετρίας. Πρώτον, αναζητούνται δύο οριζόντιοι άξονες που απομονώνουν τρεις κύριες περιοχές σώματος, που συνήθως αντιστοιχούν στο κεφάλι, τον κορμό και τα πόδια. Στα δύο τελευταία, υπολογίζεται ένας κάθετος άξονας συμμετρίας εμφάνισης. Έπειτα, ανιχνεύονται συμπληρωματικές όψεις της εμφάνισης του ανθρώπινου σώματος σε κάθε τμήμα , υπογραμμίζοντας:

1.Το γενικό χρωματικό περιεχόμενο μέσω ιστογράμματος HSV,
2.Η μετατόπιση χρώματος ανά περιοχή, μέσω των Σταθερών Χρωματικών Περιοχών (MSCR),
3.Την παρουσία επαναλαμβανόμενων υψηλά διαμορφωμένων patch (RHSP), τα οποία υπολογίζονται μέσω μιας νέας ανάλυσης ομοιότητας.

Τα εξαγόμενα χαρακτηριστικά σταθμίζονται από την απόσταση σε σχέση με τον κατακόρυφο άξονα, έτσι ώστε να ελαχιστοποιούνται οι επιδράσεις των διακυμάνσεων των θέσεων. Η αντιστοίχιση αυτών των χαρακτηριστικών δίνει το μέτρο ομοιότητας μεταξύ των υποψηφίων.
Το SDALF είναι μια διαδικασία τριών φάσεων. Οι φάσεις ισχύουν με ελαφρώς διαφορετικό τρόπο, είτε αναφέρονται στην ενιαία, είτε στην πολλαπλή περίπτωση. Στην πρώτη φάση, εντοπίζονται άξονες ασυμμετρίας και συμμετρίας για κάθε εικόνα ατόμου το οποίο τίθεται για ανίχνευση. Αυτή η φάση προϋποθέτει την παρουσία της σιλουέτας του ατόμου, αγνοώντας το φόντο (BG). Στην περίπτωση μιας λήψης, αποκτάται μια μάσκα Ζ που περιέχει μόνο τιμές προσκηνίου (FG) για κάθε άτομο, βγάζοντας συμπεράσματα για το μοντέλο παραγωγής STEL [57]. Το μοντέλο STEL καταγράφει τη γενική δομή μιας κλάσης εικόνας ως ανάμειξη διαφόρων συνιστωσών στοιχείων, απομονώνοντας σημαντικά τμήματα που παρουσιάζουν στενές κατανομές χαρακτηριστικών. Το μοντέλο έχει προσαρμοστεί στη μέθοδο αυτή για τον διαχωρισμό FG / BG και εκπαιδεύτηκε εκ των προτέρων χρησιμοποιώντας μια βάση δεδομένων πεζών. Η κατάτμηση σε νέα δείγματα συνίσταται σε μια γρήγορη εξαγωγή συμπερασμάτων.
Στην περίπτωση πολλαπλών λήψεων, εάν γίνει επεξεργασία μιας ακολουθίας βίντεο, αρκεί μια στρατηγική αφαίρεσης BG για να ληφθεί η σιλουέτα. Για ευκολία, υποτίθεται ότι το Ζ οριοθετείται από ένα κουτί μεγέθους $I \times J$. Στη δεύτερη φάση, τα χαρακτηριστικά εξάγονται από τις εξισωμένες εικόνες FG. Στην τρίτη και τελική φάση πραγματοποιείται η αντιστοίχιση των υπογραφών.



Σχήμα 5-11 Συμμετρία με βάση την ανθρώπινη σιλουέτα. Πρώτα εξάγεται ο ασύμμετρος άξονας TL, και ο HT. Μετά, για κάθε περιοχή R_k υπολογίζονται οι συμμετρικοί άξονες j_{lrk}

Η θεωρία Gestalt θεωρεί τη συμμετρία ως θεμελιώδη αρχή της αντίληψης: τα συμμετρικά στοιχεία είναι πιθανότερο να ενσωματωθούν σε ένα συνδεδεμένο αντικείμενο, παρά σε ασύμμετρες περιοχές. Αυτό το εύρημα έχει αξιοποιηθεί σε μεγάλο βαθμό για τον χαρακτηρισμό σημαντικών τμημάτων ενός δομημένου αντικειμένου [58]. Εδώ, εφαρμόζεται αυτήν η αρχή για την εξατομίκευση σημαντικών ανθρώπινων τμημάτων που μπορούν να περιγραφούν δυναμικά. Ένας απλός τρόπος θα ήταν απλά να χρησιμοποιήσετε σταθερά διαμερίσματα του πλαισίου οριοθέτησης. Εντούτοις,

1.Δεν είναι εγγυημένο ότι το σώμα ενός ατόμου είναι καλά στο επίκεντρο του πλαισίου οριοθέτησης και

2.Διαπιστώθηκε πειραματικά ότι μια έρευνα βασισμένη στην αναζήτηση, δίνει καλύτερα αποτελέσματα, δεδομένου ότι η κατάτμηση είναι επιρρεπής σε σφάλματα.

Οι πιο βασικοί φορείς είναι δύο. Ο πρώτος είναι ο χρωματικός διμερής

$$G(i,\delta) = \sum_{B_{[i-\delta,i+\delta]}} d^2(p_i, \hat{p}_i) \tag{5.26}$$

όπου d (•,•) είναι η ευκλείδεια απόσταση που αξιολογείται μεταξύ των τιμών pixel HSV pi, ^pi, τοποθετημένων συμμετρικά σε σχέση με τον οριζόντιο άξονα στο ύψος i. Αυτή η απόσταση αθροίζεται πάνω από το $B_{[i-\delta,\ i\ +\ \delta]}$, δηλ. Η περιοχή FG που βρίσκεται στο κιβώτιο του πλάτους J και της κατακόρυφης προέκτασης [i-δ, i +δ].
Ο δεύτερος είναι ο φορέας κάλυψης του χώρου, ο οποίος υπολογίζει τη διαφορά των περιοχών FG για δύο περιοχές:

$$S(i,\delta) = \frac{1}{J\delta} \left| \mathrm{A}(\mathrm{B}_{[i-\delta,i]}) - \mathrm{A}(\mathrm{B}_{[i,i+\delta]}) \right| \tag{5.27}$$

όπου $A(B_{[i\ -\delta,\ i]})$, όπως και παραπάνω, είναι η περιοχή FG στο πλαίσιο του πλάτους J και της κάθετης επέκτασης [i - δ, i]. Ο συνδυασμός κατάλληλων C και S δίνει τους άξονες της συμμετρίας και της ασυμμετρίας. Ο κύριος άξονας x της ασυμμετρίας AxTL βρίσκεται στο ύψος iTL, που λαμβάνεται ως:

$$iTL = \arg\min_i (1 - C(i,\delta)) + S(i,\delta) \tag{5.28}$$

Αυτό διαχωρίζει περιοχές που διαφέρουν έντονα στην περιοχή και τοποθετεί AxHT μεταξύ κεφαλής και ώμων. Η αναζήτηση για iHT είναι περιορισμένη στο διάστημα [δ, iTL - δ]. Οι τιμές iHT και iTL απομονωνόμουν τρεις περιοχές R_k, k = {0, 1, 2}, που αντιστοιχούν περίπου στην κεφαλή, το σώμα και τα πόδια, αντίστοιχα. Το τμήμα κεφαλής R_0 απορρίπτεται, επειδή συχνά αποτελείται από λίγα εικονοστοιχεία και φέρει πολύ χαμηλό πληροφοριακό περιεχόμενο. Στους R_1 και R_2 υπολογίζεται ένας άξονας y συμμετρίας. Αυτό βρίσκεται σε jLR_k, (k = 1, 2), που λαμβάνεται από:

$$jLR_k = \arg\min_j C(j,\delta) + S(j,\delta) \tag{5.29}$$

5.5.3. Μέθοδος 3:Συνδιασμός Νευρωνικού δικτύου (CNN) και LSTM

Σε ένα αρκετά μεγάλο μέρος της βιβλιογραφίας, παρουσιάζεται ένα νέο πλαίσιο για την αντιμετώπιση του προβλήματος person re-identification σε εικόνες και βίντεο. Τα μοντέλα CNN έχουν αποδειχθεί αποτελεσματικά στην ταξινόμηση και έχουν προταθεί διάφορες προσεγγίσεις στην επαναταυτοποίηση ατόμου [35], [39]. Πρόσφατα, τα επαναλαμβανόμενα νευρωνικά δίκτυα, ιδιαίτερα τα LSTM, έχουν αποδειχθεί ότι είναι σταθερά και ισχυρά στη μοντελοποίηση εξαρτήσεων στις μαθησιακές εργασίες. Έτσι, συνήθως προτείνεται να συνδυαστούν με το δίκτυο LSTM για την απεικόνιση χαρακτηριστικών εικόνας και βίντεο.

Γενικά, στη μέθοδο αυτή υπάρχουν κυρίως δύο μέρη, δηλ. η αναπαράσταση χαρακτηριστικών και το υπο-δίκτυο ομοιότητας στην προτεινόμενη δομή. Το πρώτο μέρος χρησιμοποιείται για την εξαγωγή χαρακτηριστικών από την εικόνα και βίντεο, ενώ το δεύτερο μέρος χρησιμοποιείται για τη μάθηση της μετρικής.

Στο πρώτο μέρος, η μέθοδος αυτή χρησιμοποιεί ένα CNN για να εξαγάγει το χαρακτηριστικό μιας εικόνας η οποία υιοθετεί τον συνδυασμό των CNN και LSTM. Κάθε πλαίσιο βίντεο είναι το πρώτο που τίθεται προς επεξεργασία από το CNN για την παραγωγή ενός φορέα χαρακτηριστικών που αντιπροσωπεύει την εμφάνιση των προσώπων, η οποία προωθεί περαιτέρω στο LSTM προκειμένου να εκμεταλλευτεί χωρικές και χρονικές πληροφορίες μέσα σε μια αλληλουχία βίντεο. Οι έξοδοι του LSTM συνδέονται μεταξύ τους ως το διάνυσμα χαρακτηριστικών της ακολουθίας βίντεο. Τέλος, τα χαρακτηριστικά της εικόνας εισόδου και του βίντεο προωθούν την ομοιότητα στο υπο-δίκτυο για τη μάθηση.

Είναι κοινώς αποδεκτό ότι η απόδοση των deep networks οφείλεται στην ιεραρχική εξαγωγή χαρακτηριστικών που γίνεται σε πολλά επίπεδα. Επομένως για την εικόνα εισόδου χρησιμοποιείται ένα CNN για να εξαχθούν τα χαρακτηριστικά του. Όσον αφορά το βίντεο εισόδου, χρησιμοποιούνται τα CNN με την ίδια αρχιτεκτονική για τη προεπεξεργασία κάθε καρέ σε αναπαράσταση υψηλότερου επιπέδου και στη συνέχεια υιοθετείται ένα δίκτυο LSTM για περαιτέρω διερεύνηση της χωρικής και χρονικής πληροφορίας στην ακολουθία βίντεο.

Διαδικασία μεθόδου:

1) Νευρωνικό Δίκτυο (CNN): Χρησιμοποιώντας μια ιεραρχία εκπαιδευόμενου φίλτρου και λειτουργιών συγκέντρωσης χαρακτηριστικών, το CNN είναι ικανό να

αυτοματοποιήσει την εκμάθηση σύνθετων χαρακτηριστικών για κατανόηση του περιεχομένου της εικόνας και την επίτευξη ανώτερης απόδοσης στην εξαγωγή χαρακτηριστικών. Στις εξισώσεις που σημειώνονται ανφέρεται το CNN ως συνάρτηση C (•). Με δεδομένη την εικόνα εισόδου x, το CNN εξάγει μια απεικόνιση d × 1 διαστάσεων όπου ο χάρτης ενεργοποίησής του τελικού στρώματος λειτουργεί ως ο φορέας χαρακτηριστικών του x:

$$f\left(x\right)=C\left(x\right) \tag{5.30}$$

Έχοντας ως είσοδο βίντεο, χρησιμοποιείται η μέθοδος αυτή επίσης με την ίδια αρχιτεκτονική για να εξαχθούν τα χαρακτηριστικά των πλαισίων της. Το y = {y (t) | t ∈ [1, ... ,T]} είναι μια ακολουθία βίντεο, όπου y(t) είναι το frame σε χρόνο t που αποτελείται από εικόνες ολόκληρου του σώματος ενός ατόμου και T είναι το μήκος της ακολουθίας βίντεο. Κάθε πλαίσιο του y χρησιμοποιείται στο CNN για να παράγει ένα διάνυσμα ως:

$$f_y^{(t)}=C\left(y^{(t)}\right) \tag{5.31}$$

όπου t ∈ [1, ..., T]. Κάθε φορέας διαβιβάζει το frame προς τα εμπρός στο στρώμα LSTM, όπου προβάλλεται σε χαμηλότερη διάσταση στον χώρο σε συνδυασμό με πληροφορίες από προηγούμενα χρονικά βήματα. Σημειώνεται ότι οι παράμετροι των CNN είναι κοινές σε όλα τα χρονικά βήματα που σημαίνει ότι κάθε πλαίσιο εισόδου υποβάλλεται σε επεξεργασία από το ίδιο δίκτυο εξαγωγής χαρακτηριστικών.

2) Δίκτυο LSTM: Μετά την προ-επεξεργασία από το CNN, το εξαγόμενο frame του βίντεο προωθείται στα επίπεδα LSTM. Δεδομένου ότι η επανάληψη ταυτότητας εικόνων σε βίντεο περιλαμβάνει την αναγνώριση ενός ατόμου από ένα βίντεο που περιέχει μια σειρά εικόνων, τα επίπεδα LSTM μπορεί να βοηθήσουν στη βελτίωση της απόδοσης επιτρέποντας την μετάδοση πληροφοριών μεταξύ των χρονικών βημάτων. Ενσωματώνοντας τα στρώματα LSTM, στοχεύεται στην καλύτερη καταγραφή της χρονικής πληροφορίες που υπάρχει στο βίντεο. Σε κάθε βήμα, το LSTM λαμβάνει μια νέα είσοδο και παράγει μια έξοδο με βάση την τρέχουσα και πληροφορίες από τα προηγούμενα χρονοδιαγράμματα.

Κατά τη διάρκεια της εκπαίδευσης του δικτύου LSTM με τη χρήση αλγορίθμου backpropagation, οι επαναλαμβανόμενες συνδέσεις ξεδιπλώνονται για τη δημιουργία ενός deep network προώθησης. Το δίκτυο LSTM μπορεί να απομνημονεύει εύκολα τις αλληλεξαρτήσεις μεγάλου χρονικού διαστήματος στα διαδοχικά δεδομένα και την εκμάθηση της χρονικής εξάρτησης η οποία μπορεί να μετατραπεί εύκολα στη χωρική περιοχή.

Χρησιμοποιούνται δύο στρώματα LSTM για το βίντεο και την ενσωμάτωση πληροφοριών. Ένα επίπεδο είναι ένα στρώμα κωδικοποιητή, και το άλλο είναι στρώμα αποκωδικοποιητή. Έστω ότι το μήκος ενός βίντεο είναι T, τότε το δίκτυο LSTM θα εξάγει ένα χάρτη μιας διάστασης 400 × T. Στη συνέχεια, εφαρμόζετε ένας συνθετικός πυρήνας 1 × T για το χαρακτηριστικό χάρτη και, τέλος, να ληφθεί το διάνυσμα χαρακτηριστικών 400 διαστάσεων του βίντεο.

3)Υπο-δύκτιο ομοιότητας: Το τρίτο κομμάτι της μεθόδου, αφορά την ομοιότητα των εικόνων και τον έλεγχο ταυτοποίησης. Στην μέθοδο αυτή χρησιμοποιείται ένα 400 × 801 πλήρως συνδεδεμένο στρώμα για την κατασκευή της ομοιότητας του υπο-δικτύου. Το στοιχείο που σχετίζεται με το χαρακτηριστικό εισόδου είναι η διάσταση των 801 f(x).Το στοιχείο που σχετίζεται με το χαρακτηριστικό εισόδου f(y), είναι η διάσταση των 400, αντίστοιχα.

4)Μεγέθυνση δεδομένων: Για μια δεδομένη ακολουθία, η ίδια αύξηση εφαρμόζεται σε όλα τα πλαίσια. Κατά τη διάρκεια της δοκιμαστικής φάσης, εφαρμόζεται αύξηση των δεδομένων και η ομοιότητα βαθμολογείται μεταξύ των ακολουθιών που υπολογίζονται κατά μέσο όρο σε όλες τις συνθήκες αύξησης.

5.5.4. Μέθοδος 4: Supervised Smoothed Manifold (SSM)

Η μέθοδος Supervised Smooted Manifold (SSM) [72] συγκρίνει την τιμή ομοιότητας μεταξύ δύο ζευγών περιπτώσεων, ώστε να αντικατοπτρίζει την ομοιότητα και διαφοροποίηση των χαρακτηριστικών των εικόνων.Επιπλέον, το SSM προσαρμόζεται ειδικά για την επαναταυτοποίηση του ατόμου, και διαθέτει τρία πλεονεκτήματα:

• Εποπτεία: Αντί να εξετάζει κάθε περίπτωση ξεχωριστά, προτείνεται να μελετηθεί η ομοιότητα των στιγμιότυπων (σε ζεύγη). Με αυτόν τον τρόπο, η SSM μπορεί να επωφεληθεί από την εποπτεία σε ατομικούς περιορισμούς.

• Αποτελεσματικότητα: Ξεπερνά τον περιορισμό της υψηλής πολυπλοκότητας. Συνεπώς, πραγματοποιείται η μάθηση μόνο με περιπτώσεις βάσεων δεδομένων, και μπορεί να εφαρμοστεί σε σενάρια σχετικά με την επαναταυτοποίηση προσώπων μεγάλης κλίμακας.

• Γενίκευση: Η μέθοδος SSM επικεντρώνεται στη μάθηση μεταξύ περιπτώσεων. Ως εκ τούτου, η μέθοδος αυτή μπορεί να θεωρηθεί ως διαδικασία μετα-επεξεργασίας των εικόνων για την αύξηση της ακρίβειας ταυτοποίησης των εικόνων.

5.5.5. Μέθοδος 5: Spatial Transformer Networks (STN)

Η μέθοδος αυτή (Spatial Transformer Networks, STN) παρέχει σαφείς χωρικές μετασχηματιστικές δυνατότητες, ώστε να γίνει τόσο η επεξεργασία των εικόνων όσο και το ταιριασμά αυτών στην παραγωγή του αποτελέσματος. Διαθέτει 3 καθοριστικές ιδιότητες που την καθιστούν πολύ ελκυστική.

• Εύκολος μετασχηματισμός: Τα STNs μπορούν να εισαχθούν σε υπάρχουσες αρχιτεκτονικές με σχετικά μικρή επεξεργασία.

• Διαφοροποιησιμότητα: Τα STNs μπορούν να εκπαιδευτούν με εύκολο τρόπο επιτρέποντας την εκπαίδευση από άκρο σε άκρο.

• Δυναμικότητα: Τα STNs εκτελούν ενεργό χωρικό μετασχηματισμό με χάρτη χαρακτηριστικών για κάθε δείγμα εισόδου σε σύγκριση με το στρώμα συγκέντρωσης το οποίο δρα ταυτόσημα για όλα τα δείγματα εισόδου.

Το STNs περιλαμβάνουν δύο στοιχεία, τον χωρικό εντοπισμό δίκτυο για την εκπαίδευση των παραμέτρων και την παραγωγή του δικτύου για να δοκιμαστεί η εικόνα εισόδου χρησιμοποιώντας μια παρεμβολή εικόνα-πυρήνα. Κατά την εφαρμογή του STN, ο πυρήνας διηλεκτρικής παρεμβολής υιοθετείται για να δοκιμάσει την εικόνα εισόδου. Και οι τέσσερις μετασχηματισμοί χρησιμοποιούν τις παραμέτρους $\theta = [s_x, t_x, s_y, t_y]$, όπου s_x και s_y είναι ο μετασχηματισμός οριζόντιας και κατακόρυφης κλίμακας και οι παράμετροι t_x και t_y είναι οι οριζόντιες και οι κάθετες παράμετροι μετάφρασης. Το ύψος και το πλάτος της εικόνας κανονικοποιούνται στο [-1, 1].

Η εξίσωση η οποία χρησιμοποιείται είναι η εξής:

$$\begin{pmatrix} x_i^{in} \\ y_i^{in} \end{pmatrix} = \begin{bmatrix} s_x & 0 & t_x \\ 0 & s_y & t_y \end{bmatrix} \begin{pmatrix} x_i^{out} \\ y_i^{out} \\ 1 \end{pmatrix} \tag{5.32}$$

Τέλος η απώλεια για τις παραμέτρους μετασχηματισμού στο δίκτυο εντοπισμού περιγράφεται ως εξής:

$$L_{loc} = L_{cen} + \xi_1 L_{pos} + \xi_2 L_{in} \tag{5.33}$$

όπου ξ_1 και ξ_2 είναι υπερπαραμετρικά και οι τιμές τους συνήθως στα πειράματα είναι ίσες με 1.

ΚΕΦΑΛΑΙΟ 6. ΣΥΓΚΡΙΣΗ ΑΛΓΟΡΙΘΜΩΝ ΚΑΙ ΑΠΟΤΕΛΕΣΜΑΤΩΝ

Το παρόν κεφάλαιο αποτελεί την κορύφωση της εργασίας αυτής η οποία αφορά την σύγκριση βασικών αλγορίθμων και τεχνικών οι οποίοι χρησιμοποιούνται στην διαδικασία του Person Re-identification. Σημειώνεται ότι για τη σύγκριση των αποτελεσμάτων τους χρησιμοποιήθηκαν οι ίδιες βάσεις δεδομένων εικόνων προκειμένου να μπορεί να γίνει άμεσα η σύγκριση τους και για να μπορέσουν να βγούν κάποια ασφαλή συμπεράσματα. Οι πειραματικές διαδικασίες αφορούν την εκτέλεση ενός ερωτήματος (singleQ) για την ταυτοποίηση ενός ατόμου και πολλαπλών ερωτημάτων (multiQ), προκειμένου να μπορέσει να αξιολογηθεί καλύτερα η εκάστοτε τεχνική - μέθοδος.

6.1. Παρουσίαση Πειραματικής Διαδικασίας

Προκειμένου να μπορέσει κάποιος να αναγνωρίσει ποιος αλγόριθμος είναι πιο αποδοτικός και να καταφέρει να κάνει μια ορθή και ολοκληρωμένη σύγκριση, θα πρέπει αρχικά να γνωρίζει τα στοιχεία του εκάστοτε αλγορίθμου και στη συνέχεια να θέσει τα στοιχεία αυτά προς σύγκριση. Για τους περιγραφείς χαρακτηριστικών (features) των εικόνων παρουσιάζονται 5 τεχνικές, και αντίστοιχα για τις μετρικές (metrics) παρουσιάζονται 7 τεχνικές. Στη παρουσίαση και τη συγκριτική ανάλυση η οποία θα ακολουθήσει θα παρουσιαστούν συγκριτικά στοιχεία των παραπάνω

τεχνικών προκειμένου να επιβεβαιωθεί ποια τεχνική πετυχαίνει καλύτερα αποτελέσματα ανίχνευσης σε κάθε περίπτωση και επομένως είναι πιο αποδοτική. Παράλληλα, υπάρχει μια ακόμη ενότητα στο κεφάλαιο αυτό, στην οποία θα παρουσιαστεί μια συγκριτική ανάλυση κάποιων μεθόδων καθώς και κάποια από τα στοιχεία τα οποία χρησιμοποιούν κάθε μια από αυτές ξεχωριστά.
Η πειραματική διαδικασία η οποία θα ακολουθηθεί στη παρούσα εργασία είναι απαραίτητο να έχει βιβλιογραφικές αναφορές και να μπορεί να συγκρίνει τους αλγορίθμους σε όμοιο μεταξύ τους επίπεδο. Το παρόν κεφάλαιο αποτελεί το κυρίως κομμάτι της παρούσας εργασίας καθώς θα παρουσιαστεί η σύγκριση των μεθόδων και τεχνικών που έχουν αναφερθεί, καθώς και τα αποτελέσματα της σύγκρισης αυτής. Σημειώνεται ότι η σύγκριση γίνεται με βάση τη χρησιμότητα, τη χρηστικότητα, αλλά και τα αποτελέσματα ανίχνευσης - αναγνώρισης τα οποία σημειώνονται σε κάθε μέθοδο - τεχνική.

6.2. Συγκριτικά Αποτελέσματα Τεχνικών

Στη συνέχεια παρουσιάζεται η ανάλυση των αποτελεσμάτων και των πειραματικών διαδικασιών των τεχνικών, οι οποίες περιγράφηκαν στο προηγούμενο κεφάλαιο, πάνω στις γνωστές βάσεις δεδομένων εικόνων προκειμένου να αξιολογηθούν οι αποδόσεις της εκάστοτε τεχνικής.

Πίνακας 6-1 Αποτελέσματα BoW σε Market-1501 και ViPER.

Μέθοδος	Market-1501		ViPER	
	Rank-1	mAP	Rank-1	mAP
BoW	9.04	3.26	7.82	11.44
+Geo	21.23	8.46	15.47	19.85
+Geo+Gauss	34.38	14.10	21.74	26.55

Η μέθοδος BoW (βασικής γραμμής) παράγει σχετικά χαμηλής ακρίβειας αποτελέσματα: ακρίβεια βαθμού (rank-1) = 9,04% και 7,82% στα σύνολα δεδομένων

Market-1501 και VIPeR, αντίστοιχα. Στη συνέχεια αφού ενσωματώνονται οι γεωμετρικοί περιορισμοί με αντιστοίχιση λωρίδων, παρατηρείται κάποια βελτίωση στην ακρίβεια των αποτελεσμάτων. Στο Market-1501, για παράδειγμα, το mAP αυξάνεται από 3,26% σε 8,46% (+ 5,20%) και η ακρίβεια της τάξης-1 (rank-1), αυξάνεται από 9,04% σε 21,23% (+ 12,19%). Επίσης είναι σαφές ότι η μάσκα Gauss λειτουργεί το ίδιο καλά και για τα δύο σύνολα δεδομένων εικόνων. Παρατηρείται επίσης, αύξηση κατά 5,64% στο mAP στο σύνολο δεδομένων Market-1501. Επομένως, συμπεραίνουμε ότι οι εικόνες οι οποίες χρησιμοποιούνται σε αυτό το σύνολο δεδομένων έχουν καλύτερα αποτελέσματα, καθώς οι πεζοί και τα άτομα γενικότερα παρουσιάζονται στο κέντρο της εικόνας, έχουν πιο καθαρά χαρακτηριστικά και γενικότερα παρουσιάζουν μειωμένο θόρυβο.

Στη συνέχεια, δοκιμάζονται πολλαπλά ερωτήματα στα σύνολα δεδομένων, όπου κάθε ταυτότητα ερωτήματος έχει πολλαπλά bboxes. Τα αποτελέσματα υποδεικνύουν ότι η χρήση πολλαπλών ερωτημάτων βελτιώνει περαιτέρω την ακρίβεια αναγνώρισης. Η βελτίωση είναι πιο εμφανής στο σύνολο δεδομένων Market-1501, όπου οι εικόνες των ερωτημάτων υιοθετούν μια πιο διαφοροποιημένη εμφάνιση σε σχέση με αυτό του ViPER. Επιπλέον, η αναζήτηση πολλαπλών ερωτήσεων με βάση το μέγιστο είναι ελαφρώς ανώτερη από τη μέση συγκέντρωση, πιθανώς επειδή η μέγιστη συγκέντρωση δίνει μεγαλύτερα βάρη στα σπάνια αλλά σημαντικά χαρακτηριστικά και βελτιώνει την ανάκληση αυτών στην εικόνα.

Πίνακας 6-2 Παρουσίαση αποτελεσμάτων για τη χρήση διαφόρων τεχνικών στο σύνολο Market-1501

Μέθοδος	Market-1501	
	r-1	mAP
gBiCov	8.28	2.23
LOMO	26.07	7.75
BoW	34.38	14.10
+LMNN	34.00	15.66
+KISSME	39.61	17.73
BoW(MultiQ)	42.64	19.47

Η μέθοδος BoW παράγει σχετικά καλή απόδοση στο σύνολο εικόνων Market-1501, κυρίως επειδή τα δύο ζεύγη φωτογραφικών καμερών μοιράζονται περισσότερη αλληλ-επικάλυψη. Επιπλέον, η χρήση καμερών 720 × 576 SD καταγράφει ξεχωριστό υπόβαθρο σε σχέση με άλλες κάμερες HD. Το Market-1501 συλλέχθηκε από συνολικά έξι κάμερες, και φέρει εικόνες ατόμων από πολλαπλές γωνίες αλλά και με διαφορετικό θόρυβο, φωτεινότητα και γενικότερα χαρακτηριστικά, δηλαδή κάθε ταυτότητα ατόμου καταγράφηκε από έξι κάμερες το πολύ και δύο κάμερες τουλάχιστον. Προκειμένου να γίνει ένα σωστό πείραμα το οποίο θα μπορέσει να χρησιμοποιηθεί για την διαδικασία των πειραμάτων και της αξιολόγησης το σύνολο αυτό χωρίζεται σε δύο μέρη. Ένα μέρος έχει 750 ταυτότητες ως εκπαιδευτικό σετ και το άλλο έχει 751 ταυτότητες ως σύνολο δοκιμών, στις οποίες υπάρχουν 3.363 εικόνες ερωτήσεων που ανήκουν στις 750 ταυτότητες και κάθε ερώτημα έχει 14.8 εικόνες οι οποίες θα χρησιμοποιηθούν για την επαλήθευση των αποτελεσμάτων.

Ωστόσο, πέρα από τα χαρακτηριστικά και τη σύγκριση της μεθόδου σε δύο διαφορετικά σύνολα δεδομένων, σημασία έχει να παρουσιαστεί και η σύγκριση με κάποια άλλη μέθοδο - τεχνική (π.χ μέθοδο CNN) στα παραπάνω σύνολα δεδομένων, προκειμένου να μπορέσει να σημειωθεί η μέθοδος με την μεγαλύτερη ακρίβεια, αλλά και να παρουσιαστεί το καλύτερο σύνολο δεδομένων.

Το εκπαιδευμένο μοντέλο Caffe χρησιμοποιήθηκε για την προετοιμασία των βαρών του CNN. Τα δύο δίκτυα εκπαιδεύτηκαν από το μηδέν. Αρχικά εκτελέστηκε η μέθοδο της στοχαστικής κλίσης (SGD) με ρυθμό εκμάθησης 0.001 και την αποσύνθεση βάρους 0.0005. Οι παράμετροι μ, C ρυθμίζονται στα 0.002 και 0.008 αντιστοίχως στα πειράματα. Κάθε παρτίδα περιέχει 160 τριπλέτες δείγματος. Το σύνολο του πλαισίου εκπαιδεύτηκε για 150.000 επαναλήψεις συνολικά. Στο σημείο αυτό γίνεται η σύκριση με τον παραπάνω αλγόριθμο στη βάση δεδομένων Market-501.

Συγκρίνοντας τις μεθόδους BoW και CNN σημειώθηκαν τα πειραματικά αποτελέσματα που παρουσιάζονται στη συνέχεια (πίνακας 6-3). Μπορεί να παρατηρηθεί ότι η προσέγγισή του CNN επιτυγχάνει την καλύτερη απόδοση μεταξύ των δύο μεθόδων. Λαμβάνοντας υπόψη μια ενιαία εικόνα ερωτήματος, η προσέγγισή αυτή επιτυγχάνει ποσοστά αναγνώρισης κατά 45,1% και βελτιώνει κατά 26,0% την BoW (singleQ). Δεδομένων των πολλαπλών ερωτηματολογίων, η προσέγγισή αυτή ξεπερνά την υπάρχουσα μέθοδο BOW (MutiQ) με σημαντική βελτίωση της απόδοσης όσον αφορά το ποσοστό αναγνώρισης (rank-1). Μπορεί επίσης να παρατηρηθεί ότι η απόδοση της προσέγγισής αυτής με δίκτυο πολλαπλών επιπέδων λειτουργεί καλύτερα από ότι με ένα μόνο δίκτυο. Αυτό καταδεικνύει την αποτελεσματικότητα του συνδυασμού των βαθών και ρηχών δικτύων κατά την ανίχνευση και την αναγνώριση του προσώπου. Βέβαια στο σημείο αυτό θα πρέπει να σημειωθεί και η κρισιμότητα του συνόλου δεδομένων το οποίο χρησιμοποιήθηκε, καθώς είναι απαραίτητο στοιχείο στην διαδικασία του Re-identification.

Πίνακας 6-1 Παρουσίαση αποτελεσμάτων των μεθόδων BoW και CNN.

Μέθοδος	**Rank-1**
BoW (singleQ)	35.8
BoW (multiQ)	44.4
CNN (singleQ)	45.1
CNN (multiQ)	55.4

Πίνακας 6-2 Σύγκριση μεθόδων με τη χρήση απλών (single) και πολλαπλών (multi) ερωτημάτων (Quiries).

Μέθοδος	**SingleQ**		**MultiQ**	
	Rank-1	mAP	Rank-1	mAP
BoW	44.42	20.76	-	-
CNN	45.58	26.11	56.59	32.26

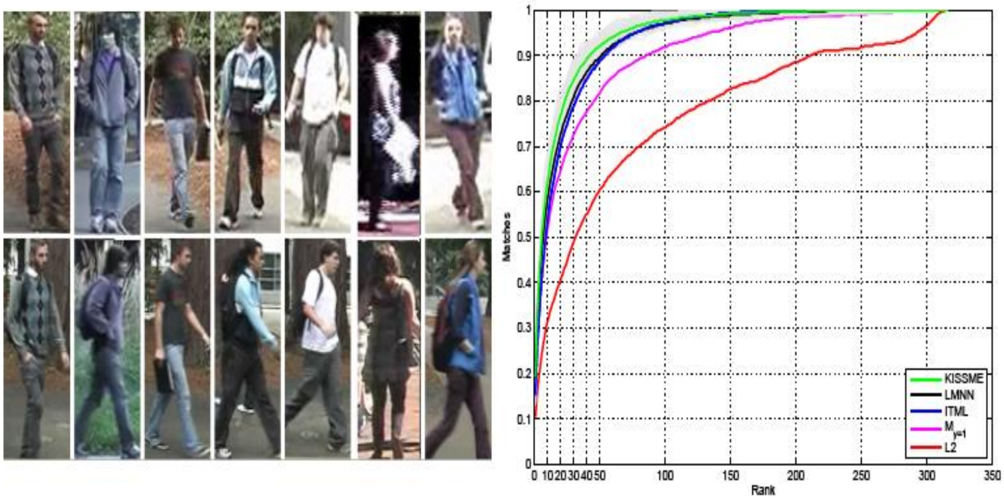
Πίνακας 6-5 Παρουσίαση επιπλέον αποτελεσμάτων για τη χρήση διαφόρων τεχνικών στο σύνολο Market-1501

Μέθοδος	Market -1501		
	r-1	r-10	mAP
gBiCov	23.4	57.2	10.6
LOMO	40.0	73.4	21.0
ZOZ	63.5	87.2	38.3
GOG	66.5	89.0	41.2

Μια ενδιαφέρουσα παρατήρηση της βάσης δεδομένων ViPER είναι ότι παρέχονται χαρακτηριστικά «υψηλού επιπέδου» που περιγράφουν την παρουσία ή την απουσία των οπτικών χαρακτηριστικών του προσώπου. Η εμφάνιση κωδικοποιείται αυτόματα σε ονομαστικά χαρακτηριστικά όπως φύλο, ηλικία, μαλλιά κ.λπ. ή «προσομοιώσεις» που σχετίζονται με την ομοιότητα του προσώπου. Αυτή η έμμεση περιγραφή αποδίδει ιδιότητες, όπως η ευρωστία σε παραλλαγές εικόνας σε σύγκριση με χαρακτηριστικά χαμηλούεπιπέδου.Επιπλέον, προσφέρει έναν συμπληρωματικό τύπο χαρακτηριστικών για την αξιολόγηση της απόδοσης των τεχνικών μετρικής απόστασης. Στην εικόνα που ακολουθεί (Σχήμα 6-1) παρουσιάζονται οι καμπύλες CMC για LMNN, SVM και KISSME αντίστοιχα. Όπως φαίνεται, η KISSME ξεπερνά σε ποσοστό την LMNN, ενώ τα αποτελέσματα ταιριάζουν με τις τελευταίες επιδόσεις της μεθόδου που βασίζεται στο SVM και προτείνεται από τους Kumar et al. [42]. Το LMNN παρέχει παρόμοια αποτελέσματα με τον αλγόριθμό KISSME ενώ είναι τάξεις μεγέθους βραδύτερος στην εκπαίδευση (πίνακας 6-6). Αυτό κάνει τον αλγόριθμο μη αποδοτικό για χρήση μεγάλης κλίμακας.

Πίνακας 6-6 Μέσος χρόνος εκπαίδευσης για KISSME και LMNN στο ViPER.

Μέθοδος	Σύνολο δεδομένων
KISSME	0,01s
LMNN	27.56s



Σχήμα 6-1 Καμπύλες CMC για LMNN, SVM και KISSME στο σύνολο ViPER.

Στη συνέχεια παρουσιάζονται συνοπτικά αποτελέσματα τεχνικών εξαγωγής χαρακτηριστκών (features) και μάθησης μετρικής (metric) στο σύνολο ViPER. Από τις τεχνικές feature παρατηρούμε ότι η τεχνική GOG πετυχαίνει τα καλύτερα αποτελέσματα r-1: 50.9%, r-5: 80.8% , r-20: 94.6%. Αντίστοιχα από τις τεχνικές metric οι τεχνικές kLFDA και MFA πετυχαίνουν τα καλύτερα αποτελέσματα τα οποία είναι αρκετά κοντά r-1: 32.3%, r-5: 65.8% , r-20: 90.9%. και r-1: 33.2%, r-5: 66.0% , r-20: 90.6% αντίστοιχα.

Πίνακας 6-7 Παρουσίαση αποτελεσμάτων για τη χρήση διαφόρων τεχνικών στο σύνολο ViPER.

Μέθοδος	ViPER		
rank	r-1	r-5	r-20
gBiCov	22.9	49.8	77.6
LOMO	41.2	71.2	91.3
ZOZ	48.9	76.5	93.0
GOG	50.9	80.8	94.6
LMNN	6.2	19.6	52.2
KISSME	19.6	-	62.60
kLFDA	32.3	65.8	90.9
MFA	33.2	66.0	90.6

6.3. Αποτελέσματα Μεθόδων

Στον πίνακα που ακολουθεί συγκρίνονται τα αποτελέσματά διαφόρων μεθόδων. Στο σύνολο δεδομένων iLIDS συγκρίνονται οι περιγραφικοί δείκτες, δηλαδή, GEI, CNN και BoW και οι μετρικής μάθησης, όπως είναι οι XQDA [24] και Kissme [13]. Παρατηρούμε ότι ο περιγραφέας του CNN είναι ανώτερος από αυτές τις μεθόδους, επιτυγχάνοντας την ποσοστό ακρίβειας ίσο με rank-1 = 53,0% στο iLIDS. Συγκρίνοντας με τα πρόσφατα έργα ανακατασκευής βίντεο, η πιο γνωστή τιμή για τη μεταβλητή αυτή είναι το 58% στο σύνολο iLIDS.

Πίνακας 6-8 Αποτελέσματα μεθόδων στο iLIDS σε rank-1, 5, 20.

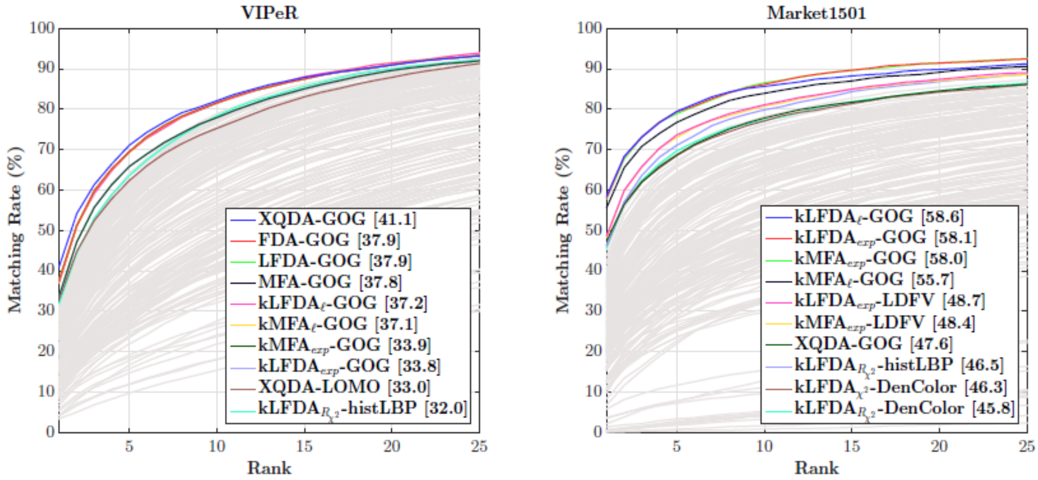
Μέθοδος	iLIDS		
rank	r-1	r-5	r-20
BoW+XQDA	14.0	32.2	59.5
GEI+KISSME	10.3	30.5	31.5
CNN +KISSME	48.8	75.6	92.6
CNN+XQDA	53.0	81.4	95.1

Στη συνέχεια παρουσιάζονται αποτελέσματα για διάφορες μεθοδολογίες στο σύνολο δεδομένων MARS κάνωντας χρήση των τεχνικών gBiCov, LOMO, BoW και CNN συνδιάζοντάς τες με τεχνικές μάθησης μετρικής όπως είναι η KISSME και η XQDA πετυχαίνοντας καλύτερα αποτελέσματα σε κάθε περίπτωση.Στο σύνολο αυτό επιτυγχάνεται ακρίβεια ίση με 68,3% και mAP = 49,3%. Η παρατήρησή μας είναι ότι μεγάλος αριθμός πεζών παρουσιάζουν παρόμοια κίνηση, ένα φαινόμενο που είναι λιγότερο συχνό σε μικρά σύνολα δεδομένων. Επομένως, η χρήση των χαρακτηριστικών κίνησης τείνει να παρατηρείται σε πεζούς με παρόμοιες δραστηριότητες. Αντίθετα, η ενσωμάτωση του CNN είναι αποτελεσματική στην αντιμετώπιση της εικόνας με όποιες διακυμάνσεις γίνουν και αποδίδει εξαιρετικά αποτελέσματα σε χαρακτηριστικά κίνησης.

Πίνακας 6-9 Αποτελέσματα μεθόδων στο MARS σε mAP και rank-1, 5, 20.

Μέθοδος	MARS			
	r-1	r-5	r-20	mAP
gBiCov+XQDA	9.2	19.8	33.5	3.7
LOMO+XQDA	30.7	46.6	60.9	16.4
BoW+KISSME	30.6	46.2	59.2	15.5
CNN+XQDA	65.3	82.0	89.0	47.6
CNN+KISSME	65.0	81.1	88.9	45.6
CNN +MQ	68.3	82.6	89.4	49.3

Οι καμπύλες CMC για δύο σύνολα δεδομένων εικόνων (ViPer, Market-1501) πολλαπλών λήψεων παρουσιάζονται στη παρούσα ενότητα (Σχήμα 6-2). Η καμπύλη CMC είναι μια γραφική αναπαράσταση του ρυθμού επαναπροσδιορισμού των k κλάσεων από τη μετρική rank-k. Όπως μπορεί να φανεί από τις καμπύλες CMC, η «εξάπλωση» στην απόδοση των αλγορίθμων με τη πάροδο των χρόνων και τη παρουσίαση όλο και πιο σύγχρονων τεχνικών για κάθε σύνολο δεδομένων είναι τεράστια, υποδεικνύοντας την πρόοδο που έχει σημειωθεί από την κοινότητα που ασχολείται με το Person Re-id τα τελευταία χρόνια σε σχέση με την περασμένη δεκαετία. Στην εικόνα αυτή παρουσιάζονται συνοπτικά οι δέκα καλύτερες επιδόσεις στα σύνολα (ViPer, Market-1501) που έχουν εφαρμοστεί στα παραπάνω σύνολα δεδομένων αναφέροντας τη χρήση των τεχνικών που έχουν εφαρμστεί σε συνδυασμό σε κάθε σύνολο δεδομένων. Σημειώνετε ότι η GOG πετυχαίνει τα καλύτερα αποτελέσματα από τις τεχνικές εξαγωγής χαρακτηριστικών, όπου σε συνδυασμό μαζί με τεχνικές μάθησης μετρικής όπως η XQDA και η kLDFA πετυχαίνει τις καλύτερες επιδόσεις σε κάθε περίπτωση. Στην περίπτωση του συνόλου ViPER η GOG με τη χρήση της XQDA πετυχαίνει τη καλύτερη επίδοση με μια μέση απόδοση 41,1%, ενώ για το σύνολο Market-1501 με τη χρήση της kLFDA πετυχαίνει τη καλύτερη επίδοση με μέση απόδοση 58,6%,



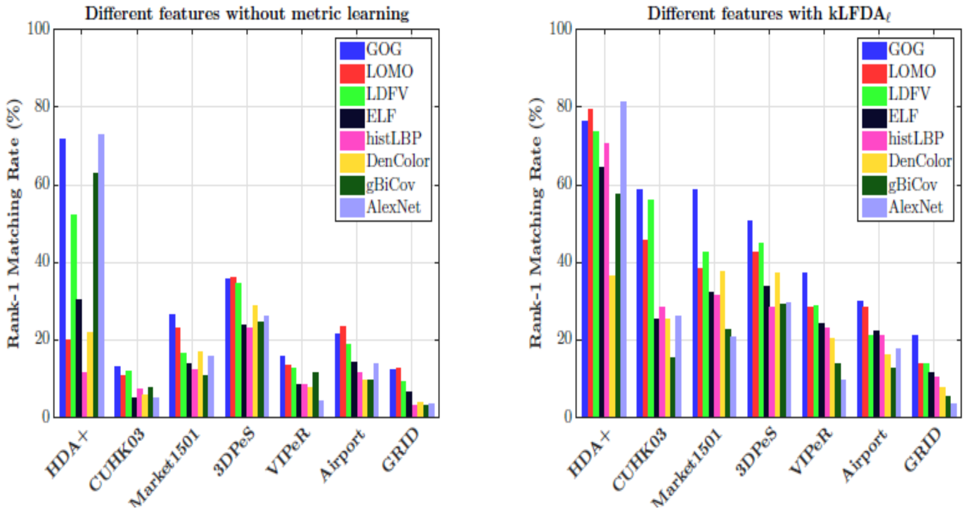
Σχήμα 6-2 Καμπύλες CMC με τις δέκα καλύτερες επιδόσεις στα σύνολα δεδομένων VIPeR και Market-1501.

Η επαναταυτοποίηση ατόμων περιλαμβάνει δύο βασικές διαδικασίες: 1)την εξαγωγή χαρακτηριστικών και 2) τη μάθηση της μετρκής. Αρχικά, σημειώνεται ότι η GOG πετυχαίνει τα καλύτερα αποτελέσματα με τις καλύτερες επιδόσεις στην εξαγωγή χαρακτηριστικών. Για να επιβεβαιωθεί αυτή η παρατήρηση, μελετάται η χρήση του GOG σε σύγκριση με άλλους αλγόριθμους εξαγωγής χαρακτηριστικών τόσο με τη παρουσία όσο και την απουσία οποιασδήποτε μετρικής μάθησης (Σχήμα 6-3). Ο GOG πετυχαίνει την καλύτερη επίδοση σε 3 από τα 7 σύνολα δεδομένων χωρίς τη χρήση κάποιας μετρικής ενώ παρουσιάζει αποτελέσματα πολύ κοντά σε απόδοση με τις καλύτερες επιδόσεις των υπολοίπων 4 συνόλων δεδομένων. Επίσης με τη χρήση κάποιας μετρικής και συγκεκριμένα της τεχνικής kLFDA η GOG πετυχαίνει τις καλύτερες επιδόσεις στα 6 από τα 7 σύνολα δεδομένων. Αυτά τα πειράματα δείχνουν σαφώς ότι η GOG είναι πράγματι η τεχνική εξαγωγής χαρακτηριστικών με τις καλύτερες επιδόσεις. Αυτό συμβαίνει επειδή το χρώμα και η υφή είναι οι πιο περιγραφικές πτυχές της εικόνας ενός ατόμου και η GOG περιγράφει το παγκόσμιο χρώμα και τις διανομές υφής χρησιμοποιώντας τοπικές Gaussian τεχνικές στα χαρακτηριστικά επιπέδου pixel. Ένας άλλος σημαντικός λόγος για την επιτυχία του GOG είναι η ιεραρχική μοντελοποίηση του τοπικού χρώματος και της υφής. Αυτό αποτελεί ένα κρίσιμο βήμα, γιατί συνήθως τα ρούχα ενός ατόμου αποτελούνται από τοπικά μέρη, καθένα από τα οποία έχει τοπικά χαρακτηριστικά και ιδιότητες.

Πίνακας 6-10 Αποτελέσματα τεχνικών με τα καλύτερα αποτελέσματα ανά dataset.

Dataset	Μέθοδος	r-1	r-5	r-10
ViPER	GOG+XQDA	41.1	71.1	82.1
iLIDS	GOG+KISSME	75.7	90.1	93.6
CAVIAR4REID	GOG+KISSME	55.6	79.6	95.6
Market-1501	GOG+kLFDA	58.6	79.4	85.7

Εξετάζοντας τη μέση απόδοση-1 συνολικά και στα 7 σύνολα δεδομένων, η kLFDA πετυχαίνει μία απόδοση 47,5%, ενώ η XQDA 45,3% αντίστοιχα, υποδεικνύοντας ότι η kLFDA είναι οριακά η μετρική με τις καλύτερες επιδόσεις καθώς η διαφορά δεν είναι στατιστικά σημαντική. Από την παραπάνω έρευνα μπορούμε να συμπεράνουμε τα εξής: τόσο η kLFDA όσο και η XQDA αποτελούν ισχυρές τεχνικές μετρικής με τη GOG-kLFDA να πετυχαίνει στο σύνολο τα καλύτερα αποτελέσματα. Είναι ενδιαφέρον να σημειώσουμε επίσης ότι τόσο η kLFDA όσο και η XQDA περιλαμβάνουν την επίλυση της γενικευμένης αποσύνθεσης ιδιοτιμών.



Σχήμα 6-3 Συγκριτικά αποτελέσματα τεχνικών σε διάφορα datasets.

ΚΕΦΑΛΑΙΟ 7. ΣΥΜΠΕΡΑΣΜΑΤΑ ΚΑΙ ΜΕΛΛΟΝΤΙΚΗ ΕΡΕΥΝΑ

Η επαναταυτοποίηση ενός ατόμου αποτελεί μια αρκετά δύσκολη διαδικασία με ευρεία εφαρμογή σε πολλούς τομείς και με μεγάλη ερευνητική δραστηριότητα τα τελευταία χρόνια. Τα μοντέλα Re-ID και οι τεχνικές αναγνώρισης παρότι έχουν προχωρήσει πολύ, εξακολουθούν να είναι πολύ περιορισμένες ως προς την εφαρμογή τους και τη χρήση σε προβλήματα πραγματικού κόσμου. Το προφανές επόμενο βήμα στην ανάπτυξη τέτοιων μοντέλων είναι η ενσωμάτωση βιομετρικών σημείων σε αυτά. Τα μοντέλα πρέπει να σχεδιάζονται έχοντας κατά νου την πολυπλοκότητα της διαδικασίας κατά την εξαγωγή των χαρακτηριστικών καθώς θέματα προκλήσεων που εμφανίζονται κατά τη διαδικασία αυτή όπως θέματα αποθήκευσης. Τα ιεραρχικά μοντέλα μπορούν να χρησιμοποιηθούν για να μειωθεί σημαντικά ο χώρο αναζήτησης μέσα σε ένα σύνολο εικόνων. Αυτό θα διευκολύνει πολύ τα ζητήματα κλιμακώσεως καθώς και τη μείωση της απόδοσης αναγνώρισης.

Η εστίαση της παρούσας εργασίας Re-ID είναι προσανατολισμένη σε ισχυρούς περιγραφείς και αποτελεσματικά συστήματα αντιστοίχισης, ωστόσο το ζήτημα της επεκτασιμότητας είναι αυτό που συχνά αγνοείται. Η επεκτασιμότητα αναφέρεται στην ικανότητα του συστήματος να προσαρμόζεται σε ρεαλιστικές συνθήκες, διατηρώντας παράλληλα την απόδοση του. Λόγω των ζητημάτων επεκτασιμότητας χρειάζονται περαιτέρω έρευνα για την αντιμετώπιση των προβλημάτων αυτών. Σε εφαρμογές πραγματικού κόσμου το μέγεθος μιας γκαλερί είναι αρκετά μεγάλο και συνεχώς αυξάνεται. Οι συνήθεις τεχνικές ταξινόμησης που βασίζονται στην ομοιότητα δεν αποδίδουν και επομένως πρέπει να διερευνηθούν αποτελεσματικά συστήματα εκτίμησης. Επίσης καθώς το μέγεθος μιας γκαλερί αλλάζει, προστίθενται νέα μοντέλα που βασίζονται στη μάθηση των τεχνικών re-id. Η βελτιστοποίηση επομένως πρέπει να εξεταστεί για να ενσωματωθεί η μεταβλητότητα στη συλλογή και να διατηρήσουν την απόδοσή τους.

Οι περιγραφείς είναι συχνά πολύπλοκοι, υψηλών διαστάσεων και δαπανηροί για χρήση επομένως επηρεάζουν τη πολυπλοκότητα του συστήματος. Τα μεγάλα μεγέθη των γκαλερί και των μοντέλων υψηλής διάστασης απαιτούν και μεγάλες ποσότητες χώρου αποθήκευσης και υπολογιστικών πόρων για την αποτελεσματική ανάλυση των δεδομένων. Οι αυτοματοποιημένες αναλύσεις βίντεο μπορούν να απλουστευθούν με επικοινωνία μεταξύ των καμερών, εντούτοις, συστήματα με έντονη απαίτηση για χώρο αποθήκευσης και υπολογιστικούς πόρους δεν μπορούν να κλιμακωθούν εύκολα. Ζητήματα όπως η ποιότητα του μοντέλου, η πολυπλοκότητα της σχεδίασης και ο συνδυασμός των καμερών απαιτούν περαιτέρω διερεύνηση..

Σημειώνεται επίσης ότι μέσα από την παρούσα εργασία είδαμε ότι η τεχνική GOG πετυχαίνει τη καλύτερη απόδοση μεταξύ των μεθόδων εξαγωγής χαρακτηριστικών.Η μοντελοποίηση των τοπικών χαρακτηριστικών χρώματος και υφής είναι κρίσιμη διαδικασία για έναν ισχυρό περιγραφέα. Ενώ αυτή η τεχνική βασίζεται στη τοπική συνδιακύμανση των χαρακτηριστικών που έχουν προηγουμένως κατασκευαστεί, ανιχνευτές μεταβλητών χαρακτηριστικών μάθησης μπορεί να οδηγήσουν σε περαιτέρω κέρδη απόδοσης. Για τη μάθηση της μετρικής είδαμε ότι τόσο η kLFDA όσο και η XQDA παρουσίασαν τα καλύτερα αποτελέσματα, δεδομένου ότι και οι δύο αυτές τεχνικές χρησιμοποιούν κριτήρια τύπου Fisher. Ωστόσο, η LOMO παουσιάζει σημαντικά αποτελέσματα με την χρήση της μέγιστης τοπικής συγκέντρωσης σε σύγκριση με τεχνικές που χρησιμοποιούν αλληλουχίες τοπικών χαρακτηριστικών. Σαφώς, μια τέτοια τεχνική η οποία χρησιμοποιείται ήδη ευρέως από προσεγγίσεις με χρήση CNN δικτύων είναι σημαντική και μπορεί να αποτελέσει χρήσιμο εργαλείο στο πλαίσιο της επανεξέτασης..

Στα πλαίσια δεδομένων πολλαπλών λήψεων η ανάπτυξη τεχνικών εξαγωγής χαρακτηριστικών ή μάθησης που εκμεταλλεύονται ειδικά τη χρονική δομή των δεδομένων [28], [84], [85] αποτελούν μια φυσική επέκταση των μεθόδων που χρησιμοποιούν μόνο χωρικές πληροφορίες εικόνας. Ενώ οι περισσότερες υπάρχουσες μέθοδοι αντιμετωπίζουν αυτά τα δύο θέματα ξεχωριστά, η ανάπτυξη μιας ενιαίας τεχνικής μπορεί ενδεχομένως να οδηγήσει σε περαιτέρω κέρδη απόδοσης.

Ως μελλοντική επέκταση της εργασίας αυτής είναι ο έλεγχος όλων των παραπάνω στοιχείων που έχουν αναφερθεί και σημειωθεί τόσο σε θεωρητικό όσο και σε πειραματικό επίπεδο. Στόχος μας είναι να αξιολογήσουμε όλα τα παραπάνω στοιχεία της εργασίας αυτής, ώστε να μπορέσουμε μέσα από ελέγχους να επεκτείνουμε τις ήδη υπάρχουσες μεθόδους και τεχνικές ώστε να μπορέσουμε να οδηγηθούμε στο σχεδιασμό πιο αποτελεσματικών συστημάτων Re-id.

ΑΝΑΦΟΡΕΣ

[1] D. Gray and H. Tao. Viewpoint invariant pedestrian recognition with an ensemble of localized features. In European Conference on Computer Vision, 2008.

[2]. Liong, Venice Erin, Jiwen Lu, and Yongxin Ge. "Regularized local metric learning for person re-identification." Pattern Recognition Letters 68 (2015): 288-296.

[3]. Hamdoun, Omar, et al. "Person re-identification in multi-camera system by signature based on interest point descriptors collected on short video sequences." Distributed Smart Cameras, 2008. ICDSC 2008. Second ACM/IEEE International Conference on. IEEE, 2008.

[4]. Camelia βιβλιοθήκη επεξεργασίας εικόνων, HTTP://CAMELLIA.SOURCEFORGE.NET

[5]. Beis 1. and Lowe D., "Shape indexing using approximate nearest-neighbour search in high-dimensional spaces", In Proc. 1997 IEEE Conf. on Computer Vision and Pattern Recognition, pages 1000-1006, Puerto Rico, 1997.

[6]. Paisitkriangkrai, S., Shen, C., & van den Hengel, A. (2015). Learning to rank in person re-identification with metric ensembles. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 1846-1855).

[7]. C. H. Teo, A. Smola, S. V. N. Vishwanathan, and Q. V. Le. A scalable modular convex solver for regularized risk minimization. In Proc. of Int. Conf. on Knowl. Disc. and Data Mining, 2007.

[8]. D. Gray, S. Brennan, H. Tao, Evaluating appearancemodels for recognition, reacquisition, and tracking, IEEE International Workshop on Performance Evaluation of Tracking and Surveillance, 2007.

[9]. U.H. Office, i-Lids Multiple Camera Tracking Scenario Definition, 2007.

[10]. Cheng, Dong Seon, et al. "Custom Pictorial Structures for Re-identification." Bmvc. Vol. 2. No. 5. 2011.

[11]. Felzenszwalb, P.F., Girshick, R.B., McAllester, D., Ramanan, D.: Object detection with discriminatively trained part-based models. Pattern Analysis and Machine Intelligence, IEEE Transactions on 32(9), 1627{1645 (2010).

[12]. Zheng, Liang, et al. "Mars: A video benchmark for large-scale person re-identification." European Conference on Computer Vision. Springer International Publishing, 2016.

[13]. Kostinger, M., Hirzer, M.,Wohlhart, P., Roth, P.M., Bischof, H.: Large scale metric learning from equivalence constraints. In: CVPR. pp. 2288{2295 (2012).

[14]. Hastie, T., Tibshirani, R., & Friedman, J. (2009). Overview of supervised learning. In The elements of statistical learning (pp. 9-41). Springer New York.

[16]. Fisher, D. H., Pazzani, M. J., & Langley, P. (Eds.). (2014). Concept formation: Knowledge and experience in unsupervised learning. Morgan Kaufmann.

[17]. Chapelle, O., Scholkopf, B., & Zien, A. (2009). Semi-supervised learning (chapelle, o. et al., eds.; 2006)[book reviews]. IEEE Transactions on Neural Networks, 20(3), 542-542.

[18]. Murphy, K. P. (2012). Machine learning: a probabilistic perspective.

[19]. Phil Simon (March 18, 2013). Too Big to Ignore: The Business Case for Big Data. Wiley, σελ. 89. ISBN 978-1-118-63817-0.

[20]. Vidanapathirana, M., Sudasingha, I., Kanchana, P., Vidanapathirana, J., & Perera, I. Open Set Person Re-Identification Framework on Closed Set Re-Id Systems.

[21]. Aichert, A. (2008, January). Feature extraction techniques. In Camp Medical Seminar (Vol. 13, pp. 325-33).

[22]. Li, W.,Wang, X.: Locally aligned feature transforms across views. In: CVPR (2013)

[23]. Lindeberg, T. (2012). Scale invariant feature transform. Scholarpedia, 7(5), 10491.

[24]. Calonder, M., Lepetit, V., Strecha, C., & Fua, P. (2010). Brief: Binary robust independent elementary features. Computer Vision–ECCV 2010, 778-792.

[25]. Pele, O., Werman, M. (2010). The quadratic-chi histogram distance family. European Conference on Computer Vision, Crete, Greece.

[26]. Wang, Z., Fan, B., & Wu, F. (2011, November). Local intensity order pattern for feature description. In Computer Vision (ICCV), 2011 IEEE International Conference on (pp. 603-610). IEEE.

[27].HTTPS://EN.WIKIPEDIA.ORG/WIKI/FEATURE_DETECTION_(COMPUTER_VISION)

[28]. Dollar, P., Wojek, C., Schiele, B., & Perona, P. (2012). Pedestrian detection: An evaluation of the state of the art. IEEE transactions on pattern analysis and machine intelligence, 34(4), 743-761.

[29]. Pearson, K. (1900). On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. Philosophical Magazine, Series 5 50 (302): 157–175.

[30]. Welling, M. (2005). Fisher linear discriminant analysis. Department of Computer Science, University of Toronto, 3(1).

[31]. Holland, S. M. (2008). Principal components analysis (PCA). Department of Geology, University of Georgia, Athens, GA, 30602-2501.

[32]. X. Wang, "Intelligent multi-camera video surveillance: A review," Pattern recognition letters, vol. 34, no. 1, pp. 3–19, 2013.

[33]. T. Huang and S. Russell, "Object identification in a bayesian context," in IJCAI, vol. 97, 1997, pp. 1276–1282.

[34]. Madden, C., Cheng, E.D., Piccardi, M.: Tracking people across disjoint camera views by an illumination-tolerant appearance representation. Mach. Vis. Appl. 18(3), 233–247 (2007).

[35]. Farenzena, M., Bazzani, L., Perina, A., Murino, V., Cristani, M.: Person re-identification by symmetry-driven accumulation of local features. In: IEEE Conference Computer Vision and Pattern Recognition, pp. 2360–2367 (2010).

[36]. Prosser, B., Zheng,W., Gong, S., Xiang, T.: Person re-identification by support vector ranking.In: British Machine Vision Conference, pp. 21.1–21.11 (2010)

[37]. Bak, S., Corvee, E., Brémond, F., Thonnat, M.: Person re-identification using spatial covariance regions of human body parts. In: IEEE International Conference on AdvancedVideo and Signal Based Surveillance, pp. 435–440 (2010).

[38]. W. Zajdel, Z. Zivkovic, and B. Krose, "Keeping track of humans: Have i seen this person before?" in Proceedings of the 2005 IEEE International Conference on Robotics and Automation. IEEE, 2005, pp. 2081–2086.

[39]. L. Zheng, L. Shen, L. Tian, S.Wang, J.Wang, and Q. Tian, "Scalable person re-identification: A benchmark," in Proceedings of the IEEE International Conference on Computer Vision, 2015, pp. 1116–1124.

[40]. Gong, S., Cristani, M., Loy, C. C., & Hospedales, T. M. (2014). The re-identification challenge. In Person re-identification (pp. 1-20). Springer, London.

[41]. Hamdoun, O., Moutarde, F., Stanciulescu, B., Steux, B., 2008. Person re-identification in multi-camera system by signature based on interest point descriptors

collected on short video sequences. In: IEEE Conf. on Distributed Smart Cameras. pp. 1–6.

[42]. N. Kumar, A. C. Berg, P. N. Belhumeur, and S. K. Nayar. Attribute and Simile Classifiers for Face Verification. In Proc. IEEE Intern. Conf. on Computer Vision, 2009. 2, 4, 5, 6

[43]. Koestinger, M., Hirzer, M., Wohlhart, P., Roth, P. M., & Bischof, H. (2012, June). Large scale metric learning from equivalence constraints. In Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on (pp. 2288-2295). IEEE.

[44]. Arth, C., Leistner, C. and Bishof, H. (2007). Object reacquisition and tracking in large-scale smart camera networks. IEEE International Conference on Distributed Smart Cameras, Vienna, Austria.

[45]. Javed, O., Shafique, K., Rasheed, Z. and Shah, M. (2008). Modeling inter-camera space-time and appearance relationships for tracking across non-overlapping views. Computer Vision and Image Understanding, Vol. 109, pp. 146 –162.

[46]. Weinberger, K. Q., & Saul, L. K. (2009). Distance metric learning for large margin nearest neighbor classification. Journal of Machine Learning Research, 10(Feb), 207-244.

[47]. Xiong, F., Gou, M., Camps, O., & Sznaier, M. (2014, September). Person re-identification using kernel-based metric learning methods. In European conference on computer vision (pp. 1-16). Springer, Cham.

[48]. Trivedi, M. M., Gandhi, T. L. and Huang K. S. (2005). Distributed interactive video arrays for event capture and enhanced situational awareness. IEEE Intelligent Systems, Vol. 20, pp. 58-66.

[49]. M. Riesenhuber, T. Poggio, Hierarchical models of object recognition in cortex, Nat. Neurosci. 2 (11) (1999) 1019–1025.

[50]. Ma, B., Su, Y., & Jurie, F. (2014). Covariance descriptor based on bio-inspired features for person re-identification and face verification. Image and Vision Computing, 32(6-7), 379-390.

[51]. Man, J., & Bhanu, B. (2006). Individual recognition using gait energy image. IEEE transactions on pattern analysis and machine intelligence, 28(2), 316-322.

[52]. Bazzani, L., Cristani, M., & Murino, V. (2014). Sdalf: modeling human appearance with symmetry-driven accumulation of local features. In Person re-identification (pp. 43-69). Springer, London.

[53]. Bauml, M., Stiefelhagen, R. (2011). Evaluation of local features for person reidentification in image sequences. IEEE Int. Conf. on Advanced Video and Signal Based Surveillance, Klagenfurt, Austria.

[54]. Bak, S., Corvee, E., Bremond, F., Thonnat, M. (2010a). Person re-identification using haar-based and dcd-based signature. IEEE Int. Conf. on Advanced Video and Signal Based Surveillance, Boston, USA.

[55]. Bak, S., Corvee, E., Bremond, F., Thonnat, M. (2010b) Person re-identification using spatial covariance regions of human body parts. IEEE Int. Conf. on Advanced Video and Signal Based Surveillance, Boston, USA.

[56]. Cai, Y., Pietikainen, M. (2010). Person re-identification based on global color context. Int. Conf. on Computer vision, November 08-09, Queenstown, New Zealand.

[57]. N. Jojic, A. Perina, M. Cristani, V. Murino, and B. Frey. Stel component analysis: Modeling spatial correlations in image class structure. In CVPR, pages 2044–2051, 2009.

[58]. K. L. M. Cho. Bilateral symmetry detection and segmentation via symmetry-growing. In BMVC, 2009.

[59]. D. Gray, S. Brennan, and H. Tao. Evaluating appearance models for recongnition, reacquisition and tracking. In PETS, 2007.

[60]. D. Gray and H. Tao. Viewpoint invariant pedestrian recognition with an ensamble of localized features. In ECCV, pages 262–275,2008.

[61]. U. H. Office. i-LIDS multiple camera tracking scenario definition, 2008.

[62]. W. Zheng, S. Gong, and T. Xiang. Associating groups of people. In BMVC, 2009.

[63]. E. H. Land and J. McCann. Lightness and retinex theory. JOSA, 61(1):1–11, 1971.

[64]. D. J. Jobson, Z.-U. Rahman, and G. A. Woodell. A multiscale retinex for bridging the gap between color images and the human observation of scenes. Image Processing, IEE Transactions on, 6(7):965–976, 1997.

[65]. Liao, S., Hu, Y., Zhu, X., & Li, S. Z. (2015). Person re-identification by local maximal occurrence representation and metric learning. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 2197-2206).

[66]. S. Gong, M. Cristani, S. Yan, and C. C. Loy. Person Re- Identification. Springer, 2014.

[67]. Ma, B., Su, Y., & Jurie, F. (2012, October). Local descriptors encoded by fisher vectors for person re-identification. In European Conference on Computer Vision (pp. 413-422). Springer, Berlin, Heidelberg.

[68]. Pedagadi, S., Orwell, J., Velastin, S., & Boghossian, B. (2013, June). Local fisher discriminant analysis for pedestrian re-identification. In Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on (pp. 3318-3325). IEEE.

[69]. D. Xu, S. Yan, L. Zhang, Z. Liu, and H. Zhang, "Coupled Subspaces Analysis," Microsoft Research Technical Report, MSRTR- 2004-106, Oct. 2004.

[70]. S. Yan, D. Xu, Q. Yang, L. Zhang, X. Tang, and H. Zhang, "Discriminant Analysis with Tensor Representation," Proc. Int'l Conf. Computer Vision and Pattern Recognition, vol. 1, pp. 526-532, 2005.

[71]. Cheng, D., Gong, Y., Zhou, S., Wang, J., & Zheng, N. (2016). Person re-identification by multi-channel parts-based cnn with improved triplet loss function. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 1335-1344).

[72]. Bai, S., Bai, X., & Tian, Q. (2017, July). Scalable person re-identification on supervised smoothed manifold. In CVPR (Vol. 6, p. 7).

[73]. D. A. Reynolds, "Gaussian mixture models," in Encyclopedia of Biometrics (S. Z. Li and A. K. Jain, eds.), pp. 659–663, Springer US, 2009.

[75]. OpenCV Documentation – Optical Flow, HTTP://DOCS.OPENCV.ORG/MASTER/D7/D8B/TUTORIAL_PY_LUCAS_KANADE.HTML#GSC.TAB=0, ΙΟΎΝΙΟΣ 2016.

[77]. Roger Y. Tsai. A Versatile Camera Calibration Techniaue for High-Accuracy 3D Machine Vision XXII AIVELA Annual Meeting IOP Publishing Journal of Physics: Conference Series 658 (2015) 012009 doi:10.1088/1742-6596/658/1/012009 6 Metrology Using Off-the-shelf TV Cameras and Lenses. IEEE, Journal of Robotics and Automation, Vol. RA-3, August 1987.

[78]. Wang Qi, Fu Li, Liu Zhenzhong. Review on Camera Calibration. Shenyang Institute of Aeronautical Engineering, Shenyang 110136, China.

[79]. Cattaneo, C., Mainetti, G., & Sala, R. (2015). The Importance of Camera Calibration and Distortion Correction to Obtain Measurements with Video Surveillance Systems. In Journal of Physics: Conference Series (Vol. 658, No. 1, p. 012009). IOP Publishing.

[80]. Pearson, K. (1900). The grammar of science. Рипол Классик.

[81]. N. Gheissari, T. B. Sebastian, and R. Hartley, "Person reidentifi- cation using spatiotemporal appearance," in 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06), vol. 2. IEEE, 2006, pp. 1528–1535.

[82]. L. Bazzani, M. Cristani, A. Perina, M. Farenzena, and V. Murino, "Multiple-shot person re-identification by hpe signature," in Pattern Recognition (ICPR), 2010 20th International Conference on. IEEE, 2010, pp. 1413–1416.

[83]. M. Farenzena, L. Bazzani, A. Perina, V. Murino, and M. Cristani, “Person re-identification by symmetry-driven accumulation of local features,” in Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on. IEEE, 2010, pp. 2360–2367.

[84]. A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classifi- cation with deep convolutional neural networks,” in Advances in Neural Information Processing Systems, 2012, pp. 1097–1105.

[85]. D. Yi, Z. Lei, S. Liao, S. Z. Li et al., “Deep metric learning for person re-identification.” in ICPR, vol. 2014, 2014, pp. 34–39.

[86]. W. Li, R. Zhao, T. Xiao, and X. Wang, “Deepreid: Deep filter pairing neural network for person re-identification,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 152–159.

[87]. J. Bromley, J. W. Bentz, L. Bottou, I. Guyon, Y. LeCun, C. Moore, E. Sackinger, and R. Shah, “Signature verification using a siamese ¨ time delay neural network,” International Journal of Pattern Recognition and Artificial Intelligence, vol. 7, no. 04, pp. 669–688, 1993.

[88]. Y. Xu, B. Ma, R. Huang, and L. Lin, “Person search in a scene by jointly modeling people commonness and person uniqueness,” in Proceedings of the 22nd ACM international conference on Multimedia. ACM, 2014, pp. 937–940.

[89]. T. Xiao, H. Li, W. Ouyang, and X. Wang, “Learning deep feature representations with domain guided dropout for person reidentification,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016.

[90]. L. Zheng, H. Zhang, S. Sun, M. Chandraker, and Q. Tian, “Person re-identification in the wild,” arXiv preprint arXiv:1604.02531, 2016.

[91]. T. Xiao, S. Li, B. Wang, L. Lin, and X. Wang, “End-to-end deep learning for person search,” arXiv preprint arXiv:1604.01850, 2016.

[92]. L. Zheng, Z. Bie, Y. Sun, J. Wang, C. Su, S. Wang, and Q. Tian, “Mars: A video benchmark for large-scale person re-identification,” in European Conference on Computer Vision, 2016.

[93]. C. Su, F. Yang, S. Zhang, Q. Tian, L. S. Davis, and W. Gao, “Multitask learning with low rank attribute embedding for person reidentification,” in Proceedings of the IEEE International Conference on Computer Vision, 2015, pp. 3739–3747

[94]. C. Su, S. Zhang, J. Xing, W. Gao, and Q. Tian, “Deep attributes driven multi-camera person re-identification,” in European Conference on Computer Vision, 2016.

[95]. H. Wang, S. Gong, and T. Xiang, “Unsupervised learning of generative topic saliency for person re-identification,” 2014.

[96]. X. Wang, W.-S. Zheng, X. Li, and J. Zhang, “Cross-scenario transfer person re-identification,” 2015.

[97]. A. J. Ma, J. Li, P. C. Yuen, and P. Li, “Cross-domain person reidentification using domain adaptation ranking svms,” IEEE Transactions on Image Processing, vol. 24, no. 5, pp. 1599–1613, 2015.

[98]. P. Peng, T. Xiang, Y. Wang, M. Pontil, S. Gong, T. Huang, and Y. Tian, “Unsupervised cross-dataset transfer learning for person re-identification,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016.

[99]. C. Liu, C. Change Loy, S. Gong, and G. Wang, “Pop: Person reidentification post-rank optimisation,” in Proceedings of the IEEE International Conference on Computer Vision, 2013, pp. 441–448.

[100]. H. Wang, S. Gong, X. Zhu, and T. Xiang, “Human-in-the-loop person re-identification,” in European Conference on Computer Vision, 2016.

[101]. N. Martinel, A. Das, C. Micheloni, and A. K. Roy-Chowdhury, “Temporal model adaptation for person re-identification,” in European Conference on Computer Vision, 2016.

[102]. L. Zheng, S. Wang, L. Tian, F. He, Z. Liu, and Q. Tian, “Queryadaptive late fusion for image search and person re-identification,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 1741–1750.

[103]. S. Paisitkriangkrai, C. Shen, and A. van den Hengel, “Learning to rank in person re-identification with metric ensembles,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 1846–1855.

[104]. J. Garcia, N. Martinel, C. Micheloni, and A. Gardel, “Person re-identification ranking optimisation by discriminant context information analysis,” in ICCV, 2015.

[105]. Q. Leng, R. Hu, C. Liang, Y. Wang, and J. Chen, “Person reidentification with content and context re-ranking,” Multimedia Tools and Applications, vol. 74, no. 17, pp. 6989–7014, 2015.

[106]. D. Qin, S. Gammeter, L. Bossard, T. Quack, and L. Van Gool, “Hello neighbor: Accurate object retrieval with k-reciprocal nearest neighbors,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2011, pp. 777–784.

[107]. W.-S. Zheng, S. Gong, and T. Xiang, “Towards open-world person re-identification by one-shot group-based verification,” IEEE transactions on pattern analysis and machine intelligence, vol. 38, no. 3, pp. 591–606, 2016.

[108]. S. Liao, Z. Mo, J. Zhu, Y. Hu, and S. Z. Li, “Open-set person re-identification,” arXiv preprint arXiv:1408.0872, 2014.

[109]. Y. Freund, R. Iyer, R. E. Schapire, and Y. Singer. An efficient boosting algorithm for combining preferences. Journal of Machine Learning Research, 4:933–969, 2003.

[110]. T. Joachims. Optimizing search engines using clickthrough data. In KDD ’02: Proceedings of the eighth ACMSIGKDD international conference on Knowledge discovery and data mining, pages 133–142, New York, NY, USA, 2002. ACM. ISBN 1-58113-567-X.

[111]. B. Prosser, S. Gong, and T. Xiang. Multi-camera matching under illumination change over time. In ECCV Workshop on Multi-camera and Multi-modal Sensor Fusion Algorithms and Applications, Marseille, France, 2008.

[112]. O. Chapelle and S. Keerthi. Efficient algorithms for ranking with svms. Information Retrieval, 18, 2010. to appear.

[113]. Sugiyama, M.: Local fisher discriminant analysis for supervised dimensionality reduction. In: 23rd International Conference on Machine Learning, pp. 905–912. ACM (2006)