Εντοπισμός και Ομαδοποίηση Προσώπων σε Βίντεο

Χαράλαμπος Παπαδόπουλος

Διπλωματική Εργασία

Επιβλέπων: Αριστείδης Λύκας

Ιωάννινα, Οκτώβριος 2019

ΤΜΗΜΑ ΜΗΧ. Η/Υ & ΠΛΗΡΟΦΟΡΙΚΗΣ
ΠΑΝΕΠΙΣΤΗΜΙΟ ΙΩΑΝΝΙΝΩΝ

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING
UNIVERSITY OF IOANNINA

Ευχαριστίες

Είμαι ευγνώμων σε όλους όσους με υποστήριξαν σε αυτό το εγχείρημα και τον καθηγητή Αριστείδη Λύκα, ο οποίος επέβλεψε αυτή την εργασία και με καθοδήγησε μέχρι την ολοκλήρωση της. Οι συμβουλές του υπήρξαν χρήσιμες και θα ήθελα να τον ευχαριστήσω επειδή με βοήθησε να κατανοήσω εις βάθος το θέμα.

Οκτώβριος 2019

Χαράλαμπος Παπαδόπουλος

Περίληψη

Ο αλγόριθμος που αναπτύχθηκε στοχεύει στην ομαδοποίηση προσώπων από δεδομένο βίντεο, βασισμένο στο μοντέλο αναγνώρισης προσώπων FaceNet.

Το πρώτο στάδιο της μεθόδου αποτελείται από την εξαγωγή συγκεκριμένων εικονοπλαισίων από το βίντεο, τον εντοπισμό όλων των προσώπων που εμπεριέχονται σε κάθε ένα από τα εικονοπλαίσια και τον υπολογισμό των χαρακτηριστικών διανυσμάτων των προσώπων χρησιμοποιώντας το προ-εκπαιδευμένο βαθύ νευρωνικό δίκτυο FaceNet.

Έπειτα, η μέθοδος ομαδοποιεί τα διανύσματα προσώπου χρησιμοποιώντας τη ρουτίνα K-means. Το κύριο μειονέκτημα του K-means είναι ότι ο αριθμός των ομάδων, K, πρέπει να προσφέρεται ως παράμετρος. Σε αυτή την εργασία θα παρουσιάσουμε μια απλή μέθοδο μέτρησης εγκυρότητας βασισμένη στις μετρήσεις αποστάσεων εσωτερικά και μεταξύ της ομάδας, η οποία χρησιμοποιείται για τον αυτόματο καθορισμό του αριθμού των ομάδων.

Τέλος, στο τελευταίο στάδιο η μέθοδος δίνει έμφαση στα παρεκκλίνοντα πρόσωπα και τα μεταχειρίζεται με ειδικό τρόπο.

Λέξεις Κλειδιά: Βίντεο, Εντοπισμός Προσώπου, FaceNet, Ομαδοποίηση, K-means, Silhouette

Abstract

Given a video as input the presented algorithm aims at clustering faces, based on the FaceNet face recognition model.

The first stage of this method consists of extracting specific frames from the video, locating all faces contained in each frame and calculating the feature vectors from those faces using FaceNet pretrained deep neural network.

Afterwards, our method clusters the face vectors using K-means. The main disadvantage of the K-means algorithm is that the number of clusters, K, must be supplied as a parameter. In this paper we use a simple validity measure based on the intra-cluster and inter-cluster distance measures, which allows the number of clusters to be determined automatically.

In the final stage, the method puts emphasis on outlier faces and treats them in a special manner.

Keywords: Video, Face Detection, FaceNet, Clustering, K-means, Silhouette

Περιεχόμενα

Κεφάλαιο 1. Εισαγωγή

1.1 Σκοπός της Εργασίας

Ο βασικός στόχος της εργασίας, δοθέντος ενός βίντεο, είναι να πραγματοποιηθεί μία σύνοψη των προσώπων που υπάρχουν σε αυτό, χρησιμοποιώντας τον αλγόριθμο FaceNet [1]. Δηλαδή, η εργασία έχει ως σκοπό την εμφάνιση όλων των προσώπων που υπάρχουν σε ένα βίντεο.

Ο αλγόριθμος FaceNet αποτελεί τη βάση αυτής της εργασίας επειδή είναι υπεύθυνος για την μοντελοποίηση των προσώπων στον διανυσματικό χώρο. Για κάθε πρόσωπο παράγει διανύσματα χαρακτηριστικών (feature vectors) τα οποία χρησιμοποιούμε για την ομαδοποίηση των όμοιων προσώπων.

Επιπρόσθετα, αποτελεί στόχος η επίτευξη ορθής ομαδοποίησης χωρίς παρεκκλίνοντα πρόσωπα (outliers). Η πρόσφατη πρόοδος της τεχνολογίας αναγνώρισης προσώπων οδήγησε σε εντυπωσιακά αποτελέσματα. Καθιστά δυνατό τον εντοπισμό προσώπων ακόμα και σε πραγματικό χρόνο. Στην έρευνα τους οι Zhang και λοιποί υποστηρίζουν ότι [2] μπορεί να επιτύχει πολύ γρήγορη ταχύτητα στην ανίχνευση και ευθυγράμμιση του προσώπου. Όμως, οι ψευδοθετικοί (false-positive) εντοπισμοί είναι ακόμα μια σημαντική πρόκληση. Στόχος είναι η διαχείριση αυτών των εντοπισμών αλλά και άλλων παρεκκλίνων προσώπων που μπορεί να παραχθούν.

Τέλος, σκοπεύουμε στην δημιουργία γραφικού περιβάλλοντος (GUI) για την δημιουργία φόρμας εισαγωγής των παραμέτρων εισόδου και για την εμφάνιση των αποτελεσμάτων, έτσι ώστε να είναι φιλικό προς τον τελικό χρήστη.

Με αυτό το λογισμικό μπορεί να μπει και η βάση για την δημιουργία ευρετηρίου προσώπων. Εκτελώντας τον αλγόριθμο σε πολλά βίντεο, δημιουργείται μια βάση δεδομένων προσώπων τα οποία ο χρήστης θα μπορούσε να αναζητήσει χωρίς να χρειάζεται να γνωρίζει το όνομα του προσώπου.

Συμπερασματικά, ο στόχος αυτής της εργασίας είναι να αναπτυχθεί ένα πλήρες πρόγραμμα με φιλική προς το χρήστη γραφική διεπαφή (GUI), σχεδιασμένο ώστε να συνοψίζει τα πρόσωπα από το βίντεο αξιόπιστα και σε ικανοποιητικό χρόνο.

Κεφάλαιο 2. Βασικές Μέθοδοι

2.1 Εντοπισμός Προσώπου (μέθοδος MTCNN)

Ο αλγόριθμος MTCNN [2] υλοποιεί κοινή ανίχνευση προσώπων και ευθυγράμμιση με κλιμακώμενα συνελικτικά δίκτυα πολλαπλών λειτουργιών (Multi-task Cascaded Convolutional Networks). Στη συνέχεια, θα παρουσιάσουμε επιγραμματικά τη λειτουργία του.

Αναφορικά με την εικόνα 2.1 υλοποιείται σωλήνωση (pipeline) αλληλεπικαλυπτόμενων δομών με σύνολο τριών σταδίων επεξεργασίας. Αρχικά, γίνονται αλλαγές στο μέγεθος της εικόνας έτσι ώστε να σχηματιστεί μια πυραμίδα από εικόνες διαφορετικών μεγεθών, οι οποίες μπαίνουν ως είσοδος στο πρώτο στάδιο.

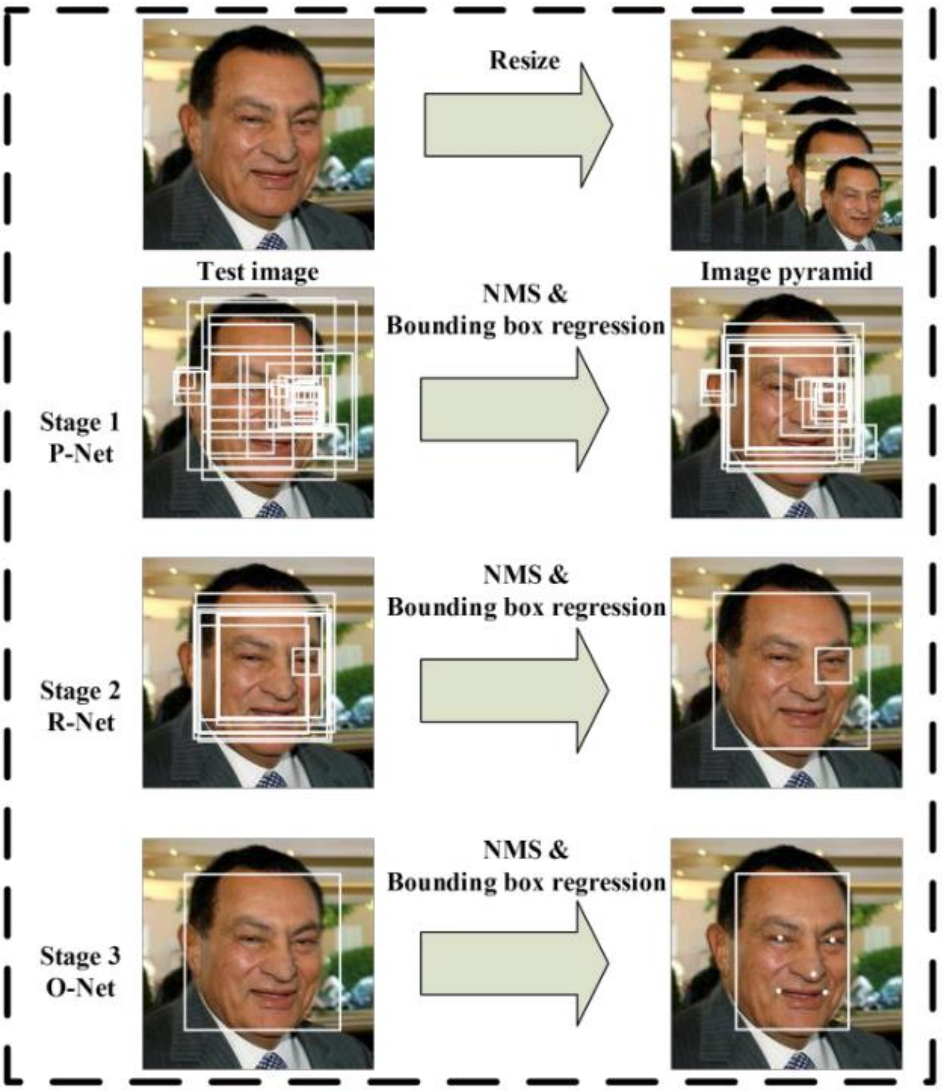
Στο **πρώτο στάδιο** χρησιμοποιείται ένα συνελικτικό δίκτυο, που ονομάζεται Proposal Network (P-Net), το οποίο προτείνει πολλά υποψήφια παράθυρα στα οποία αργότερα θα γίνει αποκοπή ώστε να απομονωθεί το πρόσωπο από την υπόλοιπη εικόνα. Στη συνέχεια, τα παράθυρα βαθμονομούνται με βάση τους εκτιμώμενους συντελεστές οπισθοδρόμησης (bounding box regression). Μετά από αυτό, χρησιμοποιείται η μη-μέγιστη καταστολή (NMS) για τη συγχώνευση παραθύρων τα οποία έχουν παρόμοιες συντεταγμένες.

Στο **δεύτερο στάδιο**, τα υποψήφια παράθυρα τροφοδοτούνται σε ένα άλλο βαθύ συνελικτικό δίκτυο (DCNN) το οποίο ονομάζεται Refine Network (R-Net). Το συγκεκριμένο δίκτυο ουσιαστικά είναι ένα φίλτρο το οποίο απορρίπτει υποψήφια παράθυρα, επειδή δεν ανιχνεύονται ως πρόσωπα. Μετά, επαναλαμβάνεται η ρουτίνα bounding box regression και η NMS.

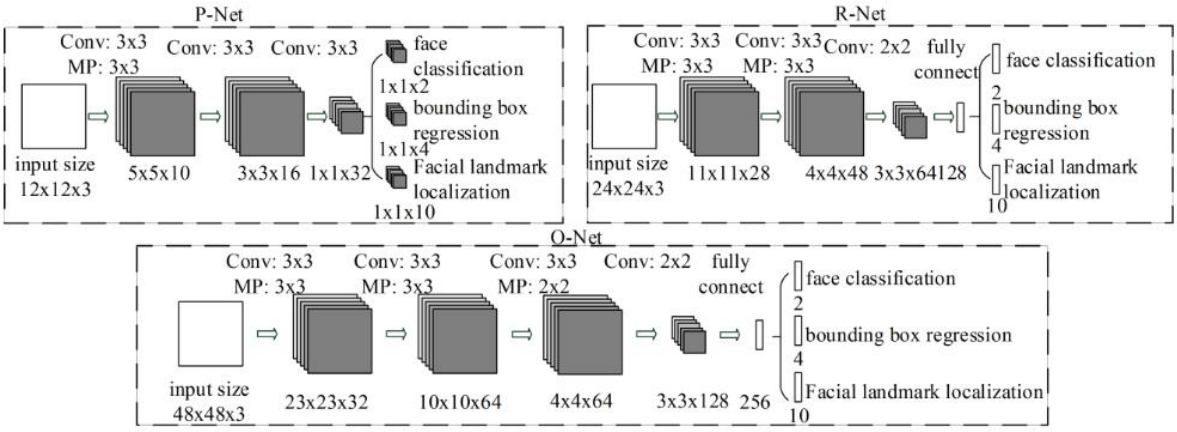
Στο **τρίτο** και τελευταίο **στάδιο**, το οποίο ονομάζεται Output Network (O-Net), είναι παρόμοιο με το δεύτερο, αλλά επιπρόσθετα ανιχνεύει κάποια χαρακτηριστικά του προσώπου. Τέτοια χαρακτηριστικά είναι:

- αριστερό και δεξί μάτι,
- μύτη,
- αριστερή και δεξιά γωνία στόματος.

Στην εικόνα 2.2 φαίνεται ακόμα πιο αναλυτικά η αρχιτεκτονική του αλγορίθμου.



Εικόνα 2.1. Υλοποίηση της σωλήνωσης στο MTCNN για την ανίχνευση προσώπου.



Εικόνα 2.2. Οι αρχιτεκτονικές των P-Net, R-Net και O-Net, όπου "MP" η διαδικασία μείωσης διαστάσεων max pooling και "Conv" σημαίνει συνέλιξη.

2.2 FaceNet

Ο αλγόριθμος FaceNet για κάθε πρόσωπο παράγει διανύσματα χαρακτηριστικών (feature vectors) τα οποία χρησιμοποιούμε για την ομαδοποίηση προσώπων. Επί της ουσίας, ένα πρόσωπο αποτελείται από διανύσματα τα οποία εκφράζουν αποστάσεις σημείων πάνω στο πρόσωπο. Για παράδειγμα, μία τέτοια απόσταση θα μπορούσε να είναι το μέγεθος του ματιού ή της μύτης. Στη συνέχεια αναλύεται η λειτουργία του.

Ο αλγόριθμος FaceNet μαθαίνει απευθείας μια απεικόνιση από εικόνες προσώπου σε ένα συμπαγές Ευκλείδειο χώρο, όπου οι αποστάσεις αντιστοιχούν άμεσα σε ένα μέτρο της ομοιότητας προσώπου. Μόλις δημιουργηθεί αυτός ο χώρος οι εργασίες όπως η αναγνώριση προσώπου, η επαλήθευση και η ομαδοποίηση μπορούν εύκολα να υλοποιηθούν, χρησιμοποιώντας τυπικές τεχνικές με τα δεδομένα που παράγει ο FaceNet ως διανύσματα χαρακτηριστικών (feature vectors) [1].

Για την κατασκευή του χρησιμοποιείται βαθύ συνελικτικό νευρωνικό δίκτυο (DCNN) και η αρχιτεκτονική του είναι βασισμένη σε δίκτυο τύπου Inception [3]. Το CNN εκπαιδεύεται με Stochastic Gradient Descent (SGD) με backpropagation [4] και AdaGrad [5]. Με την μέθοδο AdaGrad (adaptive gradient algorithm) αναγνωρίζεται δυναμικά η γεωμετρία των δεδομένων που παρατηρήθηκαν σε προηγούμενες επαναλήψεις ώστε να

προταθεί μία πιο ενημερωμένη και ανεξάρτητη κλίση (gradient). Το μοντέλο χρησιμοποιεί τη συνάρτηση ReLU ως μη γραμμική συνάρτηση ενεργοποίησης. Σχηματικά η δομή του μοντέλου απεικονίζεται στην εικόνα 2.3.

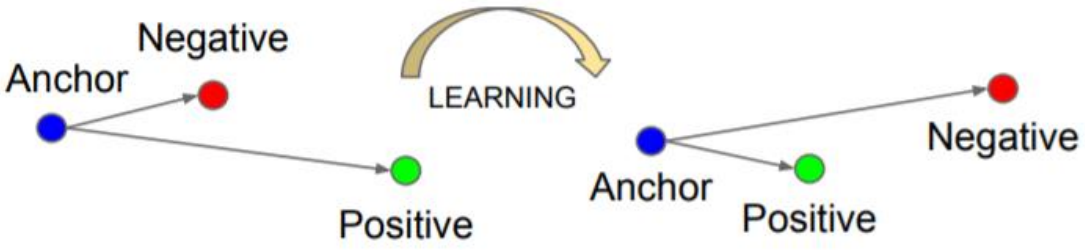
Στο επόμενο στάδιο της επεξεργασίας, τα δεδομένα κανονικοποιούνται με την L2 νόρμα και από τα διανύσματα που προκύπτουν, εφαρμόζονται ως είσοδος στη μέθοδο ελαχιστοποίησης τριπλέτας (Triplet Loss).

Η μέθοδος Triplet Loss ελαχιστοποιεί την απόσταση μεταξύ μιας άγκυρας (σημείο αναφοράς) και μιας θετικής εικόνας, που και οι δύο έχουν την ίδια κατηγορία και μεγιστοποιεί την απόσταση μεταξύ της άγκυρας και μίας αρνητικής οντότητας που είναι διαφορετικής κατηγορίας. Εδώ θέλουμε να διασφαλίσουμε ότι μια εικόνα (άγκυρα) ενός συγκεκριμένου ατόμου είναι πιο κοντά με όλες τις άλλες εικόνες (θετική) του ίδιου ατόμου, από οποιαδήποτε άλλη εικόνα (αρνητική) άλλου ατόμου. Σχηματικά φαίνεται στην εικόνα 2.4.

Επειδή, όμως, η συνθήκη της Triplet Loss ικανοποιείται πολύ συχνά και πολλές φορές χωρίς συνεισφορά στην εκπαίδευση, επιλέχθηκαν μόνο οι σημαντικότερες τριπλέτες με σκοπό την μείωση του υπολογιστικού κόστους και της ταχύτερης σύγκλισης.

Με αυτό τον τρόπο, το μοντέλο είναι έτοιμο να εκπαιδευτεί και μόλις συγκλίνει μπορούμε να του δώσουμε ένα πρόσωπο στο οποίο δεν έχει εκπαιδευτεί και να μας δώσει μία αναπαράστασή του (embedding). Δηλαδή, για κάθε εικόνα προσώπου παράγεται ένα διάνυσμα αναπαράστασης διάστασης 128, το οποίο μπορούμε να χρησιμοποιήσουμε για να κάνουμε ομαδοποίηση.

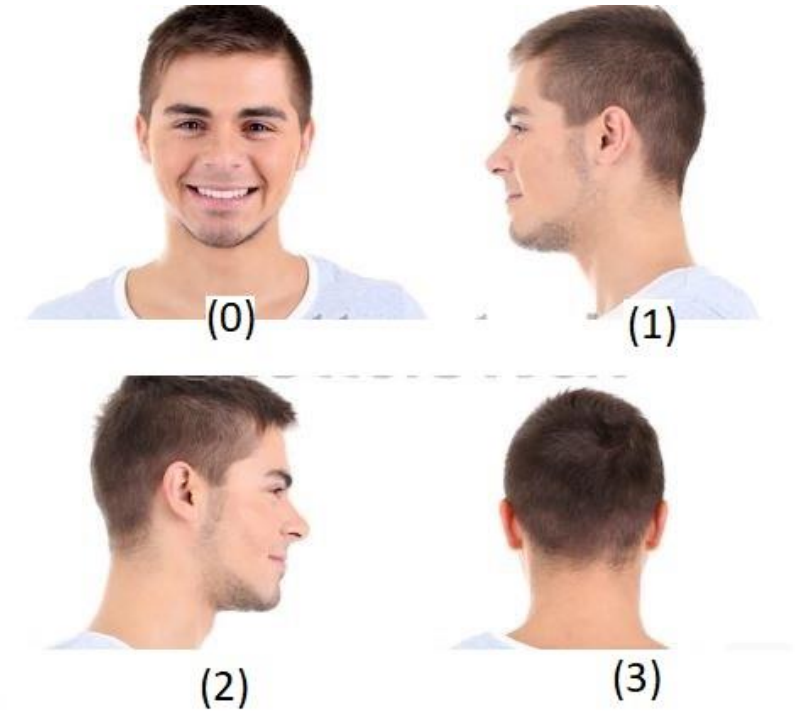
Εικόνα 2.3. Δομή του μοντέλου FaceNet.



Εικόνα 2.4. Triplet loss

Για να αξιολογήσουμε το FaceNet μπορούμε να συγκρίνουμε τα διανύσματα από εικόνες ίδιου προσώπου. Για κάθε πρόσωπο υπολογίζουμε την ευκλείδεια απόσταση από τα υπόλοιπα και έτσι δημιουργείται ο πίνακας απόστασης. Στην εικόνα 2.5 εμφανίζεται το ίδιο πρόσωπο φωτογραφισμένο από τέσσερις πλευρές και στον πίνακα 2.1 δείχνουμε τον πίνακα απόστασης. Από το συγκεκριμένο παράδειγμα μπορούμε να πούμε ότι υπάρχουν σημαντικές αποκλίσεις στις αποστάσεις και, άρα, αποτυγχάνει να αποφασίσει ότι πρόκειται για το ίδιο πρόσωπο.

Πιο συγκεκριμένα, η μπροστινή εικόνα απέχει περίπου 1 από τις προφίλ, όπως φαίνεται στον πίνακα 2.1, η οποία είναι αρκετά μεγάλη απόσταση.



Εικόνα 2.5. Παράδειγμα προσώπου φωτογραφισμένο από διαφορετικές γωνίες.

Distance matrix

	0	1	2
0	0.0000	1.0225	1.1222
1	1.0225	0.0000	0.9753
2	1.1222	0.9753	0.0000

Πίνακας 2.1. Πίνακας απόστασης για το πρόσωπο της εικ. 2.5.

Σε ένα άλλο παράδειγμα το πρόσωπο εμφανίζεται γυρισμένο, αλλά όχι προφίλ. Για αυτό το λόγο υπάρχουν πιο «κοντινές» αποστάσεις. Δηλαδή, θα μπορούσαμε να συμπεράνουμε μέσα από το FaceNet ότι πρόκειται για το ίδιο πρόσωπο.

(0) *(1)*

(2)

Εικόνα 2.7. Δεύτερο παράδειγμα προσώπου φωτογραφισμένο από διαφορετικές γωνίες.

Distance matrix

	0	1	2
0	0.0000	0.6157	0.7289
1	0.6157	0.0000	0.4747
2	0.7289	0.4747	0.0000

Πίνακας 2.2. Πίνακας απόστασης για το πρόσωπο της εικ. 2.7.

2.3 K-means

Ο αλγόριθμος k-means είναι μια δημοφιλής μέθοδος για προβλήματα ομαδοποίησης. Ο αλγόριθμος k-means στοχεύει στη διαίρεση n δειγμάτων σε k ομάδες. Το κριτήριο της ομαδοποίησης είναι το δείγμα να ανήκει στην ομάδα με το πλησιέστερο κέντρο, το οποίο χρησιμεύει ως αντιπρόσωπος της ομάδας. Αυτό έχει ως αποτέλεσμα τον διαχωρισμό του χώρου δεδομένων σε κύτταρα Voronoi (εικόνα 2.5).

Ο αλγόριθμος παρουσιάζεται συχνά ως αναθέσεις δειγμάτων στη πλησιέστερη ομάδα με βάση την (Ευκλείδεια) απόσταση. Αρχικά, επιλέγονται τυχαία κεντροειδή, ένα για κάθε ομάδα. Μετά, ο αλγόριθμος εναλλάσσεται μεταξύ δύο βημάτων. Στο πρώτο, αναθέτουμε τα δείγματα στο κοντινότερο κέντρο. Στο δεύτερο, ενημερώνουμε τα κέντρα ως τη μέση τιμή των δειγμάτων της εκάστοτε ομάδας. Επαναλαμβάνεται αυτή η εναλλαγή βημάτων μέχρι τα κεντροειδή να μην μετακινούνται σημαντικά. Η διαδικασία φαίνεται παραστατικά στο σχήμα 2.6. Παρακάτω δίνεται ο ορισμός στη φορμαλιστική του μορφή.

Με δεδομένο ένα αρχικό σύνολο k κέντρων: $m_1^{(1)}$,..., $m_k^{(1)}$, ο αλγόριθμος προχωράει εναλλασσόμενος μεταξύ δύο βημάτων:

Βήμα ανάθεσης: Αντιστοιχίζουμε κάθε δείγμα στην ομάδα της οποίας ο μέσος όρος έχει την ελάχιστη τετραγωνική ευκλείδεια απόσταση, αυτό διαισθητικά σημαίνει ο "πλησιέστερος" μέσος όρος. (Μαθηματικά, αυτό σημαίνει τον χωρισμό των παρατηρήσεων σύμφωνα με το διάγραμμα Voronoi που παράγεται από τα μέσα).

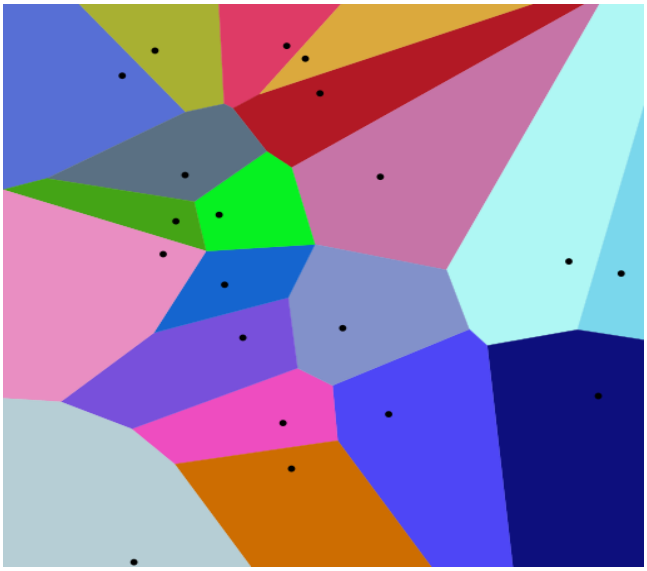
$$S_i^{(t)} = \{x_p : \|x_p - m_i^{(t)}\|^2 \leq \|x_p - m_j^{(t)}\|^2 \ \forall j, 1 \leq j \leq k\},$$

όπου x_p το p-οστό δείγμα, $S_i^{(t)}$ η ομάδα στην οποία θα ανήκει και t ο αριθμός επανάληψης.

Βήμα ενημέρωσης: Υπολογίζουμε τα νέα κέντρα (centroids) $m_i^{(t)}$ των δειγμάτων στις νέες ομάδες.

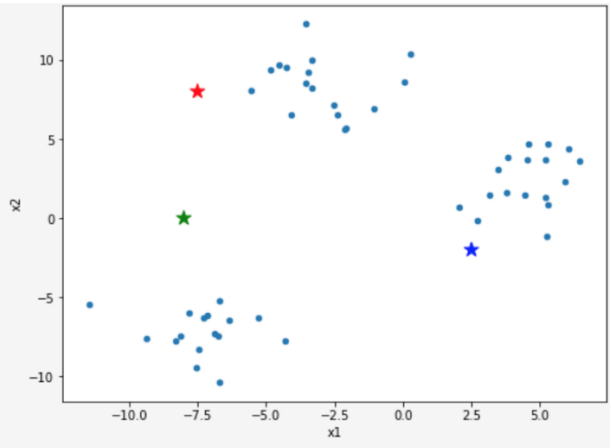
$$m_i^{(t+1)} = \frac{1}{\left|S_i^{(t)}\right|} \sum_{x_j \in S_i^{(t)}} x_j$$

Ο αλγόριθμος θα συγκλίνει όταν οι αναθέσεις δεν αλλάζουν πλέον. Αξίζει να σημειωθεί, όμως, ότι αλγόριθμος δεν εγγυάται την εύρεση βέλτιστης ομαδοποίησης [6].

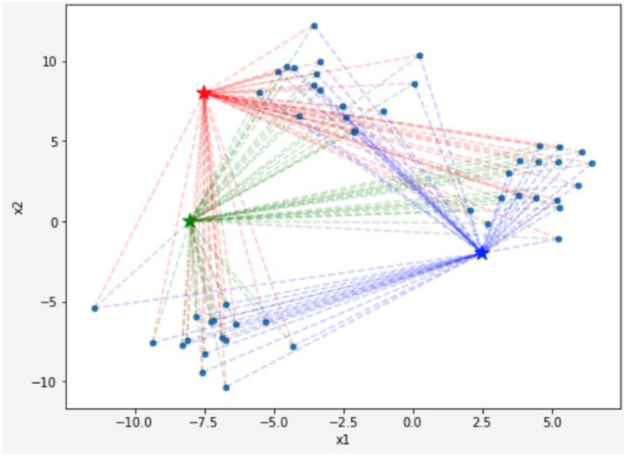


Εικόνα 2.5. Παράδειγμα κυττάρων Voronoi. Οι τελείες είναι τα κέντρα κάθε ομάδας που διαχωρίζονται με διαφορετικά χρώματα.

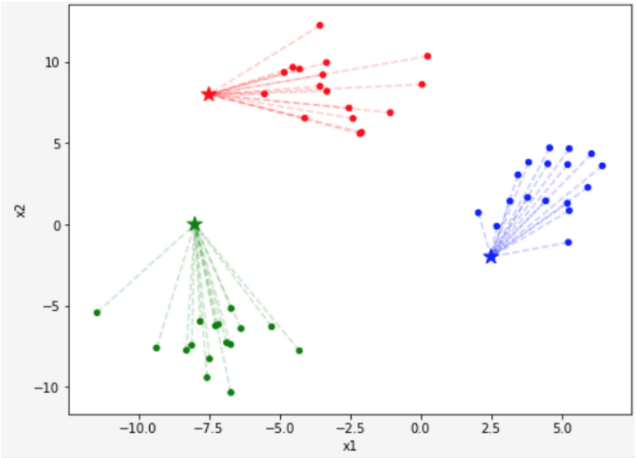
Εικόνα 2.6. Σχηματική αναπαράσταση του K-Means:



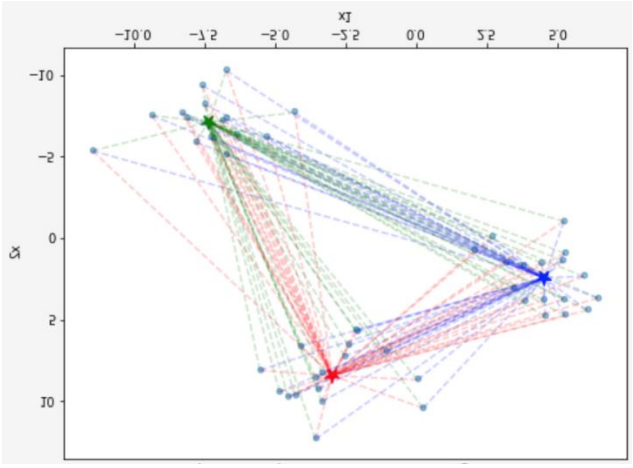
Εικόνα 2.6.1. Πρώτο βήμα. Αρχικοποίηση κεντροειδών (Centroids)



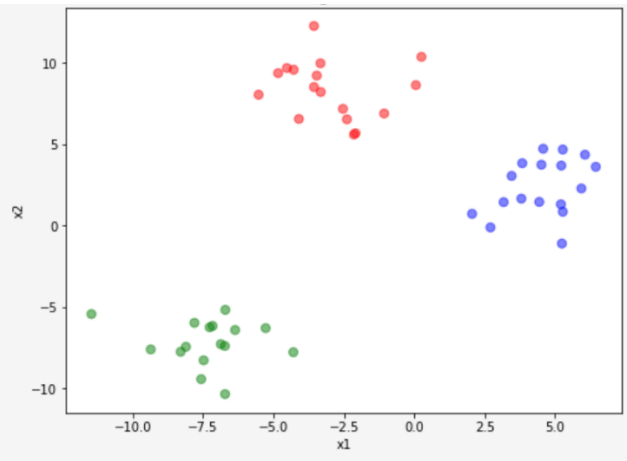
Εικόνα 2.6.2. Δεύτερο βήμα. Ευκλείδειος υπολογισμός αποστάσεων



Εικόνα 2.6.3. Τρίτο βήμα. Ομαδοποίηση με βάση το κοντινότερο κεντροειδές.



Εικόνα 2.6.4. Επανάληψη των προηγούμενων βημάτων μέχρι να υπάρξει σύγκλιση.



Εικόνα 2.6.5. Έξοδος του K-Means.

2.4 Αξιολόγηση ομαδοποίησης με τη μέθοδο Silhouette

Η ανάλυση Silhouette είναι χρήσιμη για την αξιολόγηση ομαδοποίησης σε προβλήματα μάθησης χωρίς επίβλεψη, δηλαδή, σε προβλήματα όπου δεν υπάρχουν βασικές ετικέτες αληθείας (truth labels). Η αξιολόγηση, σε αυτή τη περίπτωση, πρέπει να γίνει χρησιμοποιώντας το ίδιο το μοντέλο. Ο συντελεστής Silhouette είναι ένα παράδειγμα μιας τέτοιας αξιολόγησης, όπου ένα υψηλότερο σκορ συντελεστή Silhouette σχετίζεται με ένα μοντέλο με καλύτερα καθορισμένες ομάδες.

Ο συντελεστής Silhouette ορίζεται για κάθε δείγμα και αποτελείται από δύο βαθμολογίες:

a: Τη μέση απόσταση μεταξύ ενός δείγματος από όλα τα άλλα σημεία της ίδιας ομάδας.

b: Τη μέση απόσταση μεταξύ ενός δείγματος από όλα τα άλλα σημεία της επόμενης πλησιέστερης ομάδας.

Ο συντελεστής *Silhouette για ένα μόνο δείγμα* δίνεται ως εξής:

$$s = \frac{b - a}{\max(a, b)}$$

Όπου a,b οι αποστάσεις που περιγράφηκαν παραπάνω.

Η καλύτερη τιμή είναι 1 και η χειρότερη τιμή είναι -1. Οι τιμές κοντά στο 0 δείχνουν αλληλεπικαλυπτόμενες ομάδες. Οι αρνητικές τιμές υποδεικνύουν γενικά ότι ένα δείγμα έχει αντιστοιχηθεί σε λάθος ομάδα. Ο συντελεστής Silhouette για ένα σύνολο δειγμάτων δίνεται ως *μέσος όρος* του συντελεστή Silhouette για κάθε δείγμα.

Στο πλαίσιο αυτής της εργασίας είναι χρήσιμο για δυο λόγους. Πρώτων, το χρησιμοποιούμε για την εύρεση του αριθμού διαφορετικών προσώπων χρησιμοποιώντας τον μέσο όρο του συντελεστή Silhouette (βλ. ενότητα 3.3). Δεύτερων, είναι χρήσιμο για την ανίχνευση των παρεκκλινόντων προσώπων (outliers) (βλ. ενότητα 3.4).

Κεφάλαιο 3. Υλοποίηση

3.1 TensorFlow

Η βιβλιοθήκη TensorFlow είναι μια πλατφόρμα ανοιχτού κώδικα σχεδιασμένη για μηχανική μάθηση (ML). Έχει ένα ολοκληρωμένο, ευέλικτο οικοσύστημα εργαλείων, βιβλιοθηκών και κοινοτικών πόρων που επιτρέπει στους ερευνητές να αναπτύξουν αποδοτικά την τεχνολογία ML και στους προγραμματιστές να κατασκευάσουν και να αναπτύξουν εύκολα εφαρμογές ML [7]. Αρχικά αναπτύχθηκε για να εκτελεί μεγάλους αριθμητικούς υπολογισμούς. Έχει έναν ταχύτερο χρόνο μεταγλώττισης (compiling time) από άλλες βιβλιοθήκες βαθιάς μάθησης, όπως τις Keras και Torch και έχει API υψηλού επιπέδου ώστε, για παράδειγμα, να εξοικονομούμε χρόνο από τη συγγραφή γνωστών αλγόριθμων ML και τη διαμόρφωση των επιμέρους νευρώνων.

Η TensorFlow, όπως υποδεικνύει και το όνομα, είναι ένα πλαίσιο για τον ορισμό και την εκτέλεση υπολογισμών με τη χρήση τανυστών (tensors). Ένας τανυστής είναι μια γενίκευση των διανυσμάτων και των μητρών σε δυνητικά υψηλότερες διαστάσεις. Εσωτερικά, το TensorFlow αντιπροσωπεύει τους τανυστές ως ν-διαστάσεων πίνακες από βασικούς τύπους δεδομένων.

Ο τρόπος με τον οποίο προγραμματίζουμε σε TensorFlow βασίζεται στα γραφήματα ροής (Dataflow), τα οποία έχουν κόμβους (νευρώνες, λειτουργίες) και ακμές (τανυστές). Παρέχει έναν πρακτικό τρόπο για την ανάπτυξη γραφημάτων και καθιστά εύκολη την απεικόνισή τους με το εργαλείο TensorBoard, το οποίο ήταν δύσκολο έργο πριν. Ο μηχανισμός εκτέλεσης γίνεται με μια συνεδρία. Αυτό σημαίνει ότι μετά τη δημιουργία των γραφημάτων, δημιουργούμε μια περίοδο συνεργίας (TensorFlow session) για να τρέξουμε τμήματα των γραφημάτων.

Είναι το πιο σημαντικό εργαλείο για αυτό το εγχείρημα, δεδομένου ότι το FaceNet υλοποιείται στη TensorFlow της Python από τον D. Sandberg. Χρησιμοποιείται στο πρώτο στάδιο της προσέγγισής μας, όπως αναφέρθηκε προηγουμένως, για τον εντοπισμό και την ευθυγράμμιση του προσώπου χρησιμοποιώντας το MTCNN [2] και στη μοντελοποίηση των εικονοστοιχείων από τα πρόσωπα σε διανύσματα (embeddings), με την εκπαίδευση του μοντέλου Inception Resnet v1 χρησιμοποιώντας την μέθοδο ελαχιστοποίησης τριπλέτας (Triplet Loss) [1].

3.2 Εντοπισμός Προσώπου

3.2.1 Προσπέλαση του Βίντεο

Αρχικά, ο χρήστης, μέσα από τη γραμμή εντολών ή τη λειτουργεία του GUI, διαλέγει ένα αρχείο βίντεο και πληκτρολογεί τη συχνότητα που επιθυμεί να διαβάζει το πρόγραμμα τα εικονοπλαίσια. Πιο αναλυτικά, έστω ότι fi (frame interval) είναι η συχνότητα που επέλεξε ο χρήστης. Ο αλγόριθμος διαβάζει κάθε εικονοπλαίσιο του βίντεο, αλλά αποθηκεύει μόνο ένα ανά fi, σύμφωνα με τον τύπο:

(τρέχον πλαίσιο) *modulo* fi

3.2.2 Εντοπισμός και Ευθυγράμμιση Προσώπου

Έπειτα, εκτελείται ο εντοπιστής προσώπων MTCNN [2] ο οποίος δέχεται επαναληπτικά κάθε ένα από τα αποθηκευμένα εικονοπλαίσια που παράχθηκαν κατά τη προσπέλαση. Αν στο εικονοπλαίσιο δεν εντοπιστεί κανένα πρόσωπο, τότε διαγράφεται αυτό το εικονοπλαίσιο.

Αφού ολοκληρωθεί η εκτέλεση του MTCNN, εισάγουμε μια επιπρόσθετη προεπεξεργασία στα εντοπισμένα πρόσωπα του εικονοπλαισίου. Η πρώτη επεξεργασία είναι μια αλλαγή στο μέγεθος της εικόνας προσώπου, ώστε όλα τα δείγματα να έχουν ίδιο μέγεθος μεταξύ τους. Το νέο μέγεθος είναι 160x160 pixels. Αυτό αποσκοπεί στο να τρέξει ο αλγόριθμος του FaceNet ανεπηρέαστος από την κλίμακα των εικόνων. Η δεύτερη επεξεργασία κανονικοποιεί την εικόνα ώστε να έχει μηδενική μέση τιμή και μοναδιαία διακύμανση. Αυτό γίνεται με τον παρακάτω τύπο:

$$\frac{x - mean}{\max\left(std, \frac{1}{\sqrt{N}}\right)},$$

όπου mean είναι ο μέσος όρος όλων των τιμών στην εικόνα,
std είναι η *τυπική απόκλιση* των τιμών της εικόνας και

N ο αριθμός των εικονοστοιχείων (pixels) της κάθε εικόνας.

Αυτή η λειτουργία εξομαλύνει τις εικόνες ώστε να διευκολύνει και να επιταχύνει την εκπαίδευση του FaceNet.

Στην εικόνα 3.1 απεικονίζεται αυτή η επεξεργασία που περιγράφηκε.

Εικόνα 3.1. (α) Έξοδος του MTCNN, (β) αλλαγή μεγέθους σε 160x160 pixels, (γ) εξομάλυνση εικόνας.

3.3 FaceNet

3.3.1 Προεκπαιδευμένο Μοντέλο

Στο επόμενο βήμα, χρησιμοποιούμε ένα προεκπαιδευμένο FaceNet. Επιλέξαμε το μοντέλο 20170512-110547 το οποίο εκπαιδεύτηκε με τα δεδομένα MS-Celeb-1M [8] . Δοκιμάστηκε στα δεδομένα LFW [9] όπου επέτυχε ακρίβεια 0,992 (πίνακας 3.1).

Model name	LFW accuracy	Training dataset	Architecture
20170511-185253	0.987	CASIA-WebFace	Inception ResNet v1
20170512-110547	0.992	MS-Celeb-1M	Inception ResNet v1

Πίνακας 3.1. Λεπτομέρειες από δύο παγωμένους γράφους.

3.3.2 Εκτέλεση του Μοντέλου

Στη συνέχεια, αφού φορτωθεί επιτυχώς το προεκπαιδευμένο μοντέλο, μπαίνουν ως είσοδος σε αυτό οι εικόνες προσώπων οι οποίες είναι η έξοδος της προεπεξεργασίας και επιστρέφονται οι υπολογισμένες αναπαραστάσεις (embeddings).

3.4 K-Means και Silhouette

Για να ομαδοποιήσουμε με ακρίβεια τα πρόσωπα, αρχικά υποθέτουμε ότι τα διαφορετικά πρόσωπα του βίντεο θα είναι από 2 έως 12 και εκτελούμε τον αλγόριθμο K-Means για αυτό το πλήθος ομάδων. Δηλαδή, εκτελούμε 2-Means, 3-Means έως 12-Means. Άρα συνεπάγεται ότι για να δουλέψει σωστά ο αλγόριθμος πρέπει να υπάρχουν στο βίντεο πάνω από ένα εντοπίσιμα πρόσωπα με άνω όριο το δώδεκα.

Για κάθε εκτέλεση του K-Means, αρχικοποιούμε σε τυχαία θέση τα κεντροειδή. Έχουμε τόσα κεντροειδή όσα είναι και οι ομάδες. Μετά εκτελούμε τον K-Means, μέχρι ο αλγόριθμος να συγκλίνει. Επαναλαμβάνουμε αυτή τη διαδικασία 10 φορές, αλλά κάθε φορά, κατά το βήμα της αρχικοποίησης, διαλέγουμε τυχαία διαφορετικά κεντροειδή. Από αυτές τις 10 επαναλήψεις επιλέγουμε ως έξοδο του K-Means την λύση που έχει το μικρότερο σφάλμα ομαδοποίησης. Ως σφάλμα ομαδοποίησης ορίζουμε το άθροισμα των Ευκλείδειων αποστάσεων των δειγμάτων εσωτερικά της ομάδας.

Επίσης, μετά από κάθε εκτέλεση του K-Means υπολογίζουμε τον συντελεστή Silhouette. Δηλαδή, υπολογίζουμε και αποθηκεύουμε για κάθε δείγμα τον συντελεστή Silhouette και ύστερα αποθηκεύουμε τον μέσο όρο για όλα τα δείγματα. Έτσι, κάθε εκτέλεση του K-Means αντιπροσωπεύεται από έναν αριθμό Silhouette. Από αυτές τις 10 εκτελέσεις της ρουτίνας K-means ({2-12}-Means) αποθηκεύουμε την ομαδοποίηση με το

μεγαλύτερο silhouette score. Πχ. Στην εικόνα 3.2. καταλαβαίνουμε ότι στο βίντεο είναι πιο πιθανό να εμφανίζονται μόνο δύο άτομα γιατί ο αριθμός 0.396 είναι πιο κοντά στο 1 από τους υπόλοιπους. Όσο πιο κοντά είναι στο 1, σημαίνει ότι η ομαδοποίηση είναι καλύτερη.

Σε αυτό το στάδιο έχουμε δημιουργήσει ένα αρχικό σύστημα ομαδοποίησης προσώπων ενός βίντεο. Όμως, εξετάζοντας τα αποτελέσματα σε διάφορα βίντεο, συμπεραίνουμε ότι υπάρχουν πολλά πρόσωπα τα οποία είναι ταξινομημένα λάθος. Για αυτόν το λόγο, στην επόμενη ενότητα περιγράφουμε έναν τρόπο βελτίωσης των αποτελεσμάτων.

```
For n_clusters = 2 The average silhouette_score is : 0.39678037
For n_clusters = 3 The average silhouette_score is : 0.3441345
For n_clusters = 4 The average silhouette_score is : 0.31663883
For n_clusters = 5 The average silhouette_score is : 0.22027192
For n_clusters = 6 The average silhouette_score is : 0.23158517
For n_clusters = 7 The average silhouette_score is : 0.1604214
For n_clusters = 8 The average silhouette_score is : 0.17040136
For n_clusters = 9 The average silhouette_score is : 0.1762308
For n_clusters = 10 The average silhouette_score is : 0.12946133
For n_clusters = 11 The average silhouette_score is : 0.1314255
For n_clusters = 12 The average silhouette_score is : 0.14780599
best number of clusters:  2
```

Εικόνα 3.2. Παράδειγμα τρόπου επιλογής του ορθότερου πλήθος ομάδων από τις 10 εκτελέσεις του K-Means, *με βάση τον αριθμό Silhouette.*

3.5 Διαχείριση των παρεκκλινόντων Προσώπων (Outliers)

Ο αλγόριθμος που έχουμε δημιουργήσει μέχρι αυτό το σημείο εμφανίζει λάθη ομαδοποίησης εξαιτίας διαφόρων παραγόντων. Θα αναλύσουμε μερικά από αυτά τα προβλήματα και μετά θα περιγράψουμε τρόπους αντιμετώπισης.

3.5.1 Παρουσίαση Προβλημάτων

Όπως προαναφέρθηκε στην υποενότητα 3.1.1, ο χρήστης επιλέγει τη συχνότητα διαβάσματος εικονοπλαισίων. Εάν είναι σποραδική αυτή η δειγματοληψία, τότε υπάρχει σημαντική πιθανότητα ορισμένα πρόσωπα να έχουν εμφανιστεί μόνο μία φορά

στα δεδομένα μας. Αυτό το μοτίβο συμβαίνει σε περιπτώσεις όπως είναι οι συνεντεύξεις και τα τηλεπαιχνίδια όπου η κάμερα μερικές φορές κάνει κοντινό πλάνο σε άτομα του ακροατήριου για μικρό χρονικό διάστημα. Έτσι, υπάρχουν πρόσωπα με μόνο ένα ή δύο εικονοπλαίσια.

Αυτό για αλγορίθμους ταξινόμησης και ομαδοποίησης είναι αρνητικό επειδή για τόσο περιορισμένα δεδομένα ενός προσώπου θα είναι αδύνατο να ταξινομηθούν σε δική τους, ξεχωριστή ομάδα. Συγκεκριμένα στη δική μας προσέγγιση, ένα τέτοιο πρόσωπο θα ταξινομηθεί σε μία κοντινή ομάδα, αντί να γίνει ξεχωριστή ομάδα, επειδή ο μέσος όρος του συντελεστή Silhouette σε αυτή την ομάδα θα επιβαρυνθεί από αυτό το παρεκκλίνων πρόσωπο κατά αμελητέα ποσότητα. Άρα, ο συνολικός μέσος όρος θα είναι πάλι αρκετά κοντά στο 1 ώστε ο αριθμός των ομάδων να θεωρηθεί βέλτιστος.

Ένα άλλο πρόβλημα είναι η περίπτωση λάθος εντοπισμού προσώπου (false detection). Δηλαδή, υπάρχει μία μικρή πιθανότητα να εντοπιστεί πρόσωπο σε σημείο που δεν υπάρχει στην πραγματικότητα (παράδειγμα στην εικόνα 3.3). Αυτό συμβαίνει επειδή ο MTCNN επιτυγχάνει ακρίβεια εντοπισμού προσώπου 95.4% [2] , η οποία είναι αρκετά υψηλή, όμως αφήνει περιθώριο για λάθη.

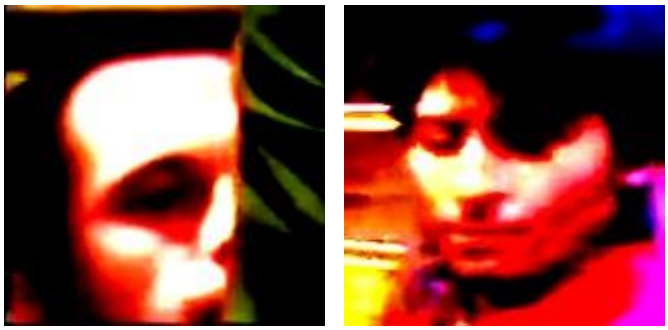
Επίσης, είναι πιθανό να εντοπιστεί πρόσωπο του οποίου η εικόνα είναι χαμηλής ποιότητας. Αυτό μπορεί να οφείλεται επειδή το πρόσωπο είναι εκτός εστίασης (out of focus), ή επειδή είναι στο παρασκήνιο του πλάνου, ή ακόμα μπορεί να εμφανίζεται μερικώς κρυμμένο από άλλο αντικείμενο (παράδειγμα στην εικόνα 3.4).

Ένα ακόμα πρόβλημα, το οποίο συμβαίνει και τακτικά, είναι η διαφορετική ομαδοποίηση προσώπου πλαγίας θέσης (προφίλ) σε σχέση με το ίδιο πρόσωπο εμπρόσθιας θέσης. Όπως είναι φυσικό, ο K-means αδυνατεί να διακρίνει ότι το πρόσωπο πλαγίας θέσης είναι το ίδιο με το άλλο, λόγω σημαντικών αποκλίσεων των μεταξύ τους αναπαραστάσεων (embeddings), που επιστρέφει το FaceNet.

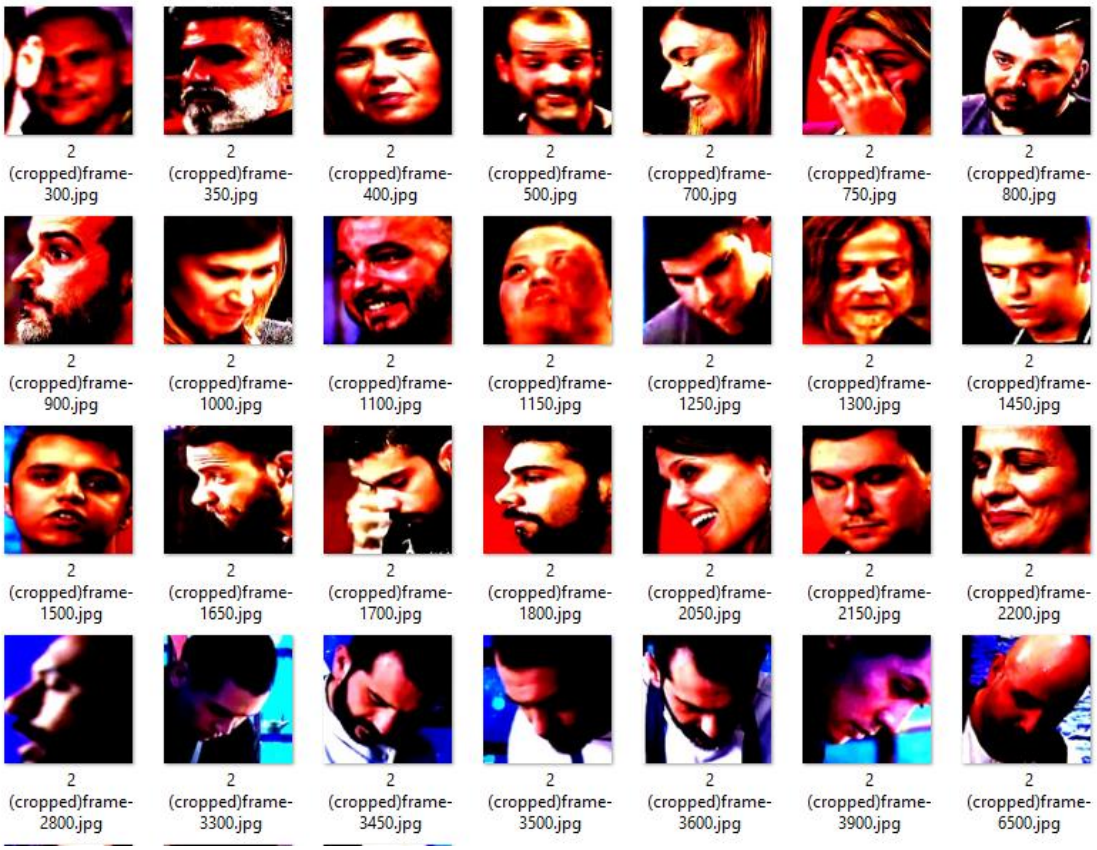
Τέλος, παρατηρήσαμε ότι υπάρχουν ομάδες με προβληματικά εικονοπλαίσια σαν αυτά που περιγράφηκαν, τα οποία κανένα δε μοιάζει με τα υπόλοιπα της ομάδας τους. Ο λόγος που συμβαίνει αυτό είναι επειδή ο αλγόριθμος FaceNet δεν επεξεργάζεται σωστά τα χαρακτηριστικά του προσώπου, λόγω έλλειψης ποιότητας. Έτσι ο K-means αδυνατεί

να τα εντάξει στις σωστές ομάδες, οπότε δημιουργεί μία ομάδα που περιέχει τις προβληματικές εικόνες. Παράδειγμα τέτοιας ομάδας φαίνεται στην εικόνα 3.5.

Εικόνα 3.3. Παραδείγματα λάθος εντοπισμού προσώπου (false detection)



Εικόνα 3.4. Παραδείγματα χαμηλής ποιότητας δείγματος. Κρυμμένο πρόσωπο αριστερά. Εκτός εστίασης δεξιά.



Εικόνα 3.5. Παράδειγμα προβληματικής ομάδας. Ομάδα γεμάτη με 'παρεκκλίνοντα' πρόσωπα.

3.5.2 Τρόποι Αντιμετώπισης

Έχοντας αποθηκεύσει τους αριθμούς Silhouette από κάθε εικόνα από το προηγούμενο στάδιο, θα τους χρησιμοποιήσουμε για να αξιολογήσουμε τα πρόσωπα ως παρεκκλίνοντα (outliers). Έστω ένας αριθμός κατωφλίωσης T. Αν ο αριθμός silhouette είναι μικρότερος από το κατώφλι T, τότε λέμε ότι το πρόσωπο είναι παρεκκλίνον. Μία καλή τιμή του αριθμού T, μετά από πολλές δοκιμές, είδαμε ότι είναι το 0.1.

Έχοντας μοντελοποιήσει το τι θεωρούμε ως παρεκκλίνοντα πρόσωπα, μένει μόνο να τα ξεχωρίσουμε από τα υπόλοιπα. Στη γενική περίπτωση, αυτά μεταφέρονται σε έναν ξεχωριστό φάκελο τον οποίο ονομάζουμε 'outliers'.

Αν όμως συμβεί η περίπτωση όπου συσσωρεύονται πολλά 'outliers' σε μία ομάδα, τότε θα λέγαμε ότι σε αυτή την ομάδα κανένα πρόσωπο δεν είναι σωστά ομαδοποιημένο. Με

την έννοια ότι στη συγκεκριμένη ομάδα κανένα πρόσωπο δε μοιάζει με τα υπόλοιπα. Άρα, υπάρχει ανάγκη να λάβουμε διαφορετική προσέγγιση από αυτήν της γενικής περίπτωσης. Δηλαδή, θα εμπλουτίσουμε τα δεδομένα μας με περισσότερα εικονοπλαίσια, τα οποία είναι γειτονικά στα παρεκκλίνοντα και μετά θα επαναλάβουμε τη διαδικασία της ομαδοποίησης.

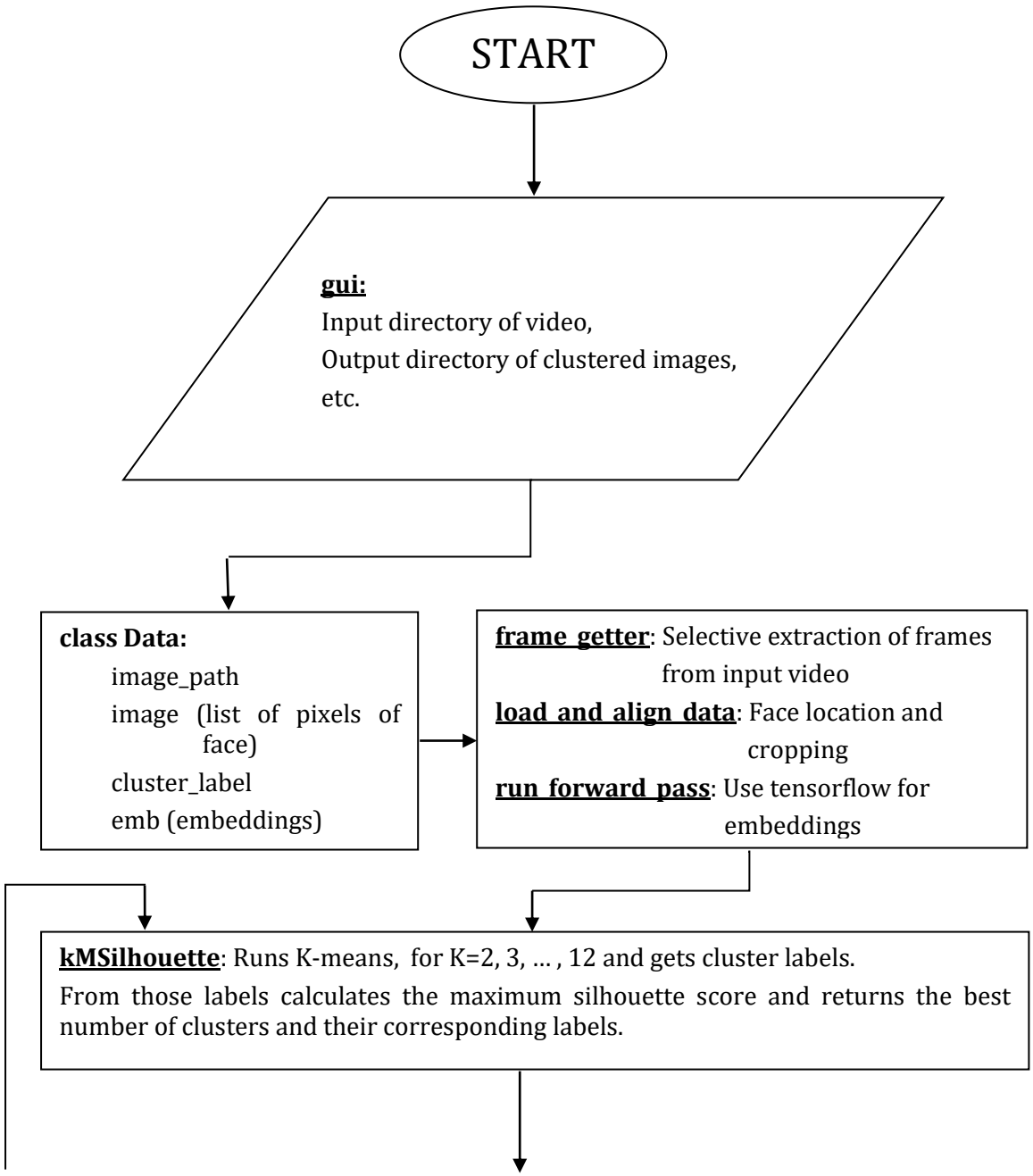
Εάν, λοιπόν, τα 'outliers' είναι πάνω από το 40% του συνόλου δειγμάτων μιας ομάδας, τότε για κάθε 'outlier' επιλέγουμε από το βίντεο 4 επιπλέον εικονοπλαίσια, τα δύο εκ των οποίων είναι πριν από το κάθε 'outlier' και τα άλλα δύο μετά από αυτό. Έπειτα, για κάθε καινούριο εικονοπλαίσιο επαναλαμβάνονται οι ρουτίνες που περιγράφηκαν στις ενότητες 3.1 και 3.2 δηλαδή, ο εντοπισμός των προσώπων και ο υπολογισμός των 'embeddings'. Εν συνεχεία, τα καινούρια δεδομένα συνενώνονται με τα υπόλοιπα και επαναλαμβάνονται οι ρουτίνες της ενότητας 3.3 (K-means και Silhouette).

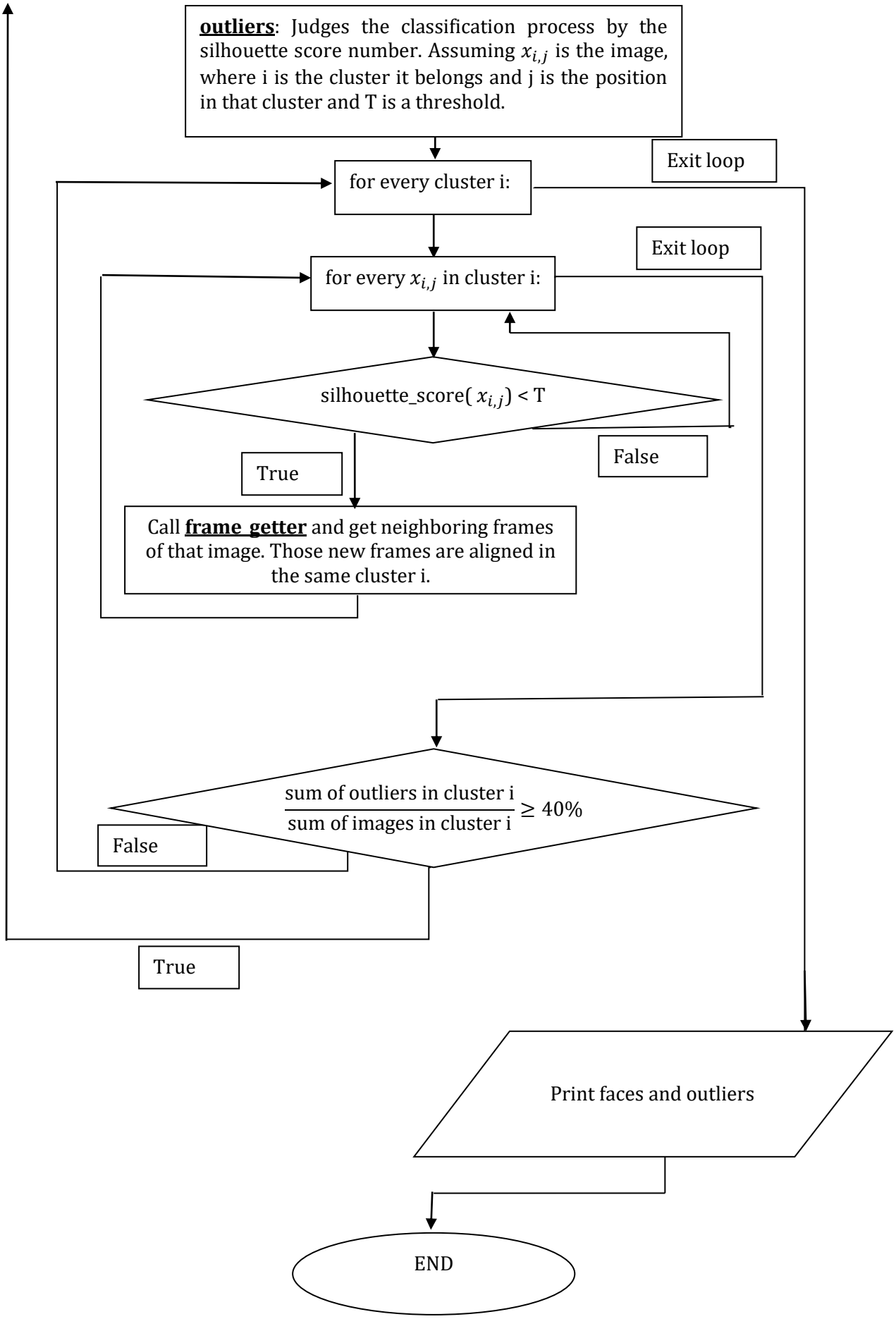
Εάν πάλι υπάρχει ομάδα με πολλά 'outliers' επαναλαμβάνεται η παραπάνω διαδικασία με όριο επαναλήψεων τρεις φορές. Εάν και μετά το όριο των επαναλήψεων εξακολουθεί να υπάρχει τέτοια ομάδα, τότε τη μεταφέρουμε στο σύνολό της στον φάκελο outliers.

Η χρήση γειτονικών εικονοπλαισίων λύνει το πρώτο πρόβλημα που περιγράψαμε στην ενότητα 3.4.1. Αν, για παράδειγμα, στα δεδομένα μας έχουμε ένα άτομο με μόνο ένα δείγμα και πάρουμε εικονοπλαίσια κοντά σε αυτό που το εντοπίσαμε είναι πολύ πιθανό στην επόμενη επανάληψη να δημιουργηθεί δική του ομάδα.

Για τα προβλήματα λάθος εντοπισμού προσώπου και προσώπου χαμηλής ποιότητας, όμως, η μόνη αντιμετώπιση είναι η μεταφορά τους στον φάκελο outliers. Την ευθύνη για αυτό το πρόβλημα την έχει ο αλγόριθμος MTCNN. Δηλαδή, αυτά τα πρόσωπα δεν θα έπρεπε να ανιχνευθούν αρχικά. Για αυτόν τον λόγο περιοριζόμαστε στον απλό διαχωρισμό αυτών των δειγμάτων από το υπόλοιπο σύνολο.

3.6 Κώδικας σε Διάγραμμα Ροής (Flowchart)

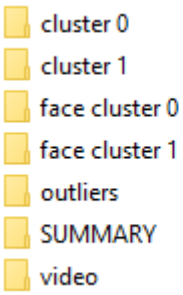



outliers: Judges the classification process by the silhouette score number. Assuming $x_{i,j}$ is the image, where i is the cluster it belongs and j is the position in that cluster and T is a threshold.
for every cluster i:
Exit loop
for every $x_{i,j}$ in cluster i:
Exit loop
silhouette_score($x_{i,j}$) < T
False
True
Call **frame_getter** and get neighboring frames of that image. Those new frames are aligned in the same cluster i.
$\frac{\text{sum of outliers in cluster i}}{\text{sum of images in cluster i}} \geq 40\%$
False
True
Print faces and outliers
END

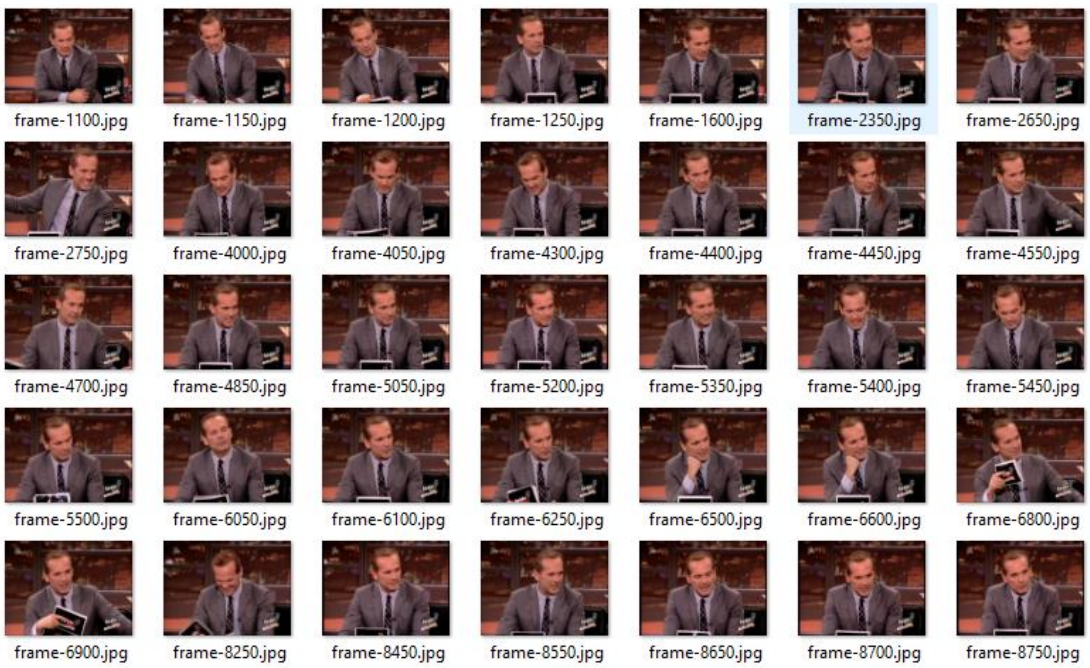
3.7 Έξοδος των Αποτελεσμάτων

Στην τελική φάση της υλοποίησης δημιουργούνται φάκελοι με τις ομάδες των εικόνων, ώστε να μπορούμε να δούμε την τελική ομαδοποίηση. Για κάθε ομάδα δημιουργείται ένας φάκελος με τα εικονοπλαίσια στα οποία ανήκουν τα πρόσωπα της ομάδας τον οποίο ονομάζουμε ‘cluster i’ , όπου i είναι το εκάστοτε cluster. Επίσης, δημιουργείται ένας φάκελος ο οποίος περιέχει τα επεξεργασμένα πρόσωπα της ομάδας τον οποίο ονομάζουμε ‘face cluster i’. Επιπλέον, δημιουργείται φάκελος με το όνομα ‘outliers’ ο οποίος συμπεριλαμβάνει όλα τα παρεκκλίνοντα πρόσωπα και τα εικονοπλαίσια στα οποία ανήκουν. Τέλος, δημιουργείται ένας φάκελος με το όνομα ‘SUMMARY’ ο οποίος περιέχει το πιο αντιπροσωπευτικό πρόσωπο της κάθε ομάδας. Αυτός ο φάκελος θα χρησιμοποιηθεί από την γραφική διεπαφή ώστε να οπτικοποιήσει τα αποτελέσματα της ομαδοποίησης.

Τα αντιπροσωπευτικά πρόσωπα του φακέλου ‘SUMMARY’ υπολογίζονται από τα κεντροειδή του αλγορίθμου K-Means. Πιο συγκεκριμένα, όταν συγκλίνει ο K-Means αποθηκεύονται τα κεντροειδή. Όμως, επειδή δεν είναι απαραίτητο το κάθε κεντροειδές να ταυτίζεται με κάποιο δείγμα, υπολογίζουμε το δείγμα το οποίο είναι πιο κοντά στο εκάστοτε κεντροειδές και έπειτα βρίσκουμε την εικόνα και το πρόσωπο στο οποίο ανήκει. Οι φάκελοι παρουσιάζονται στις εικόνες 3.5.1-4.



Εικόνα 3.5.1. Παράδειγμα φακέλων με τις ομάδες (clusters).



Εικόνα 3.5.2. Παράδειγμα εικονοπλαισίων του cluster 0.

Εικόνα 3.5.3. Παράδειγμα φακέλου 'outliers'.

Εικόνα 3.5.4. Παράδειγμα αντιπροσωπευτικού προσώπου του cluster 0 από τον φάκελο 'SUMMARY'.

Κεφάλαιο 4. Τρόποι Εκτέλεσης και Γραφική Διεπαφή Χρήστη (GUI)

Ο κώδικας της εργασίας προσφέρεται μέσω της πλατφόρμας GitHub [10]. Μέσα στον φάκελο 'video-face-clusters' βρίσκεται το αρχείο 'requirements.txt' το οποίο περιέχει τις απαραίτητες βιβλιοθήκες που πρέπει να εγκατασταθούν για την εκτέλεση του προγράμματος.

Για να είναι εύχρηστο το πρόγραμμα, υλοποιήθηκαν δύο τρόποι εκτέλεσης. Ο πρώτος είναι μέσα από τις παραμέτρους της γραμμής εντολών και ο δεύτερος μέσω GUI. Για την σωστή εκτέλεσή του πρέπει μέσω του τερματικού να είμαστε στον κατάλογο 'video-face-clusters'.

4.1 Εκτέλεση από γραμμή εντολών

Ο πιο άμεσος τρόπος εκτέλεσης είναι από το τερματικό πληκτρολογώντας:

```
python src/compare.py
```

και στην συνέχεια τις παραμέτρους όπως φαίνεται στην εικόνα 4.1.

Ένα παράδειγμα εκτέλεσης είναι το εξής:

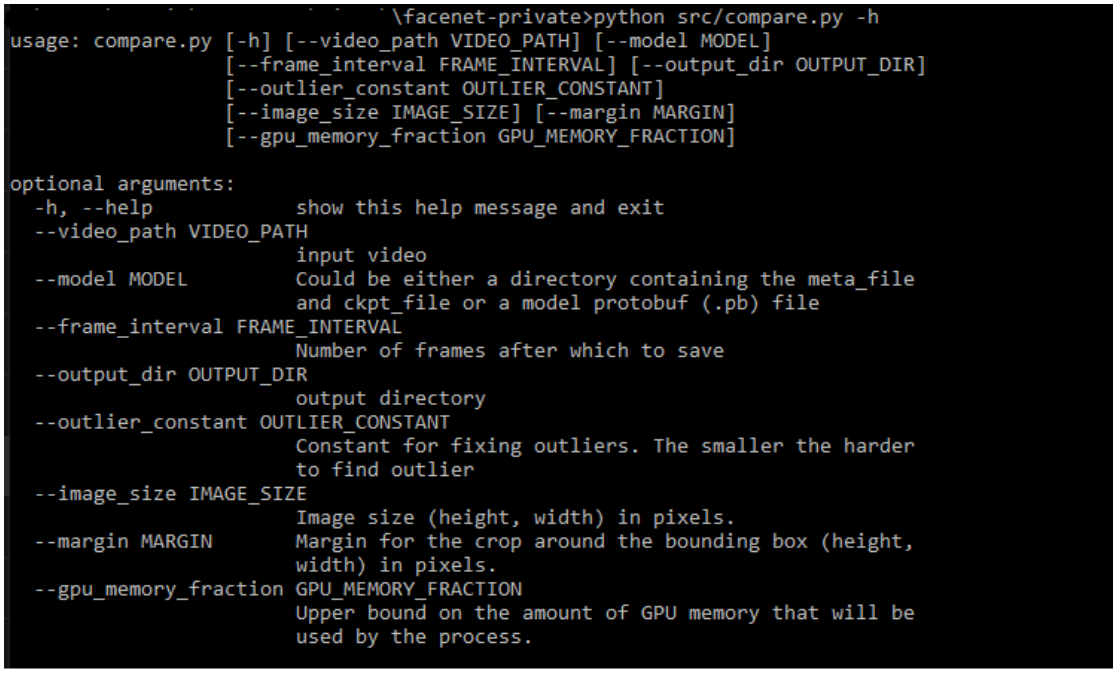
```
python src/compare.py --video_path
"Cristiano Ronaldo.mp4"    -- frame_interval 200
```

Λεπτομέρειες για τις παραμέτρους δίνονται αν εκτελέσουμε την εντολή:

```
python src/compare.py -h
```

με την προϋπόθεση ότι βρισκόμαστε στον κατάλογο ‘video-face-clusters’.

Αυτός ο τρόπος εκτέλεσης είναι ο πιο γρήγορος και υποστηρίζει ενσωμάτωση σε άλλα προγράμματα και ‘scripts’.



Εικόνα 4.1. Οι παράμετροι του προγράμματος αναλυτικά.

4.2 Εκτέλεση με χρήση του GUI

Ενώ είναι δυνατόν να εκτελεστεί το πρόγραμμα απευθείας από την γραμμή εντολών, υλοποιήσαμε ένα γραφικό περιβάλλον ώστε να είναι πιο φιλικό για τον τελικό χρήστη.

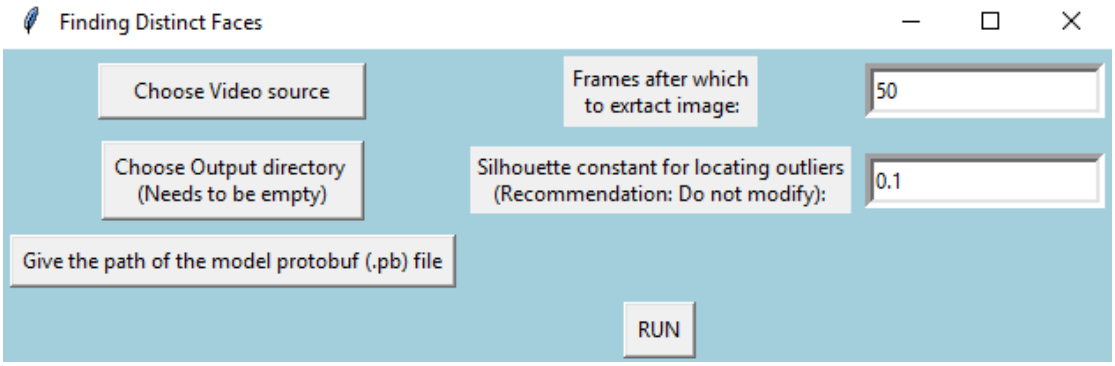
Γράφουμε την παρακάτω εντολή χωρίς παραμέτρους:

```
python src/compare.py
```

Δηλαδή, αποφεύγοντας να συμπεριλάβουμε παραμέτρους, το πρόγραμμα αναγνωρίζει την έλλειψη εισόδου και εμφανίζει ένα παράθυρο, το οποίο παραπέμπει τον χρήστη να

την ορίσει διαδραστικά μέσω μίας φόρμας εισαγωγής παραμέτρων. Το παράθυρο φαίνεται στην εικόνα 4.2.

Αφού εκτελεστεί ο αλγόριθμος εμφανίζεται ένα ακόμα παράθυρο το οποίο παρουσιάζει τα αποτελέσματα της ομαδοποίησης του φακέλου 'SUMMARY'. Επίσης, δίνει τη δυνατότητα εμφάνισης των αποτελεσμάτων πιο αναλυτικά και διαγραφής αυτών των αποτελεσμάτων. Παράδειγμα του τελικού παραθύρου στην εικόνα 4.3.



Εικόνα 4.2. Παράθυρο φόρμας εισαγωγής παραμέτρων.

Εικόνα 4.3. Παράθυρο εμφάνισης των αποτελεσμάτων. Στο συγκεκριμένο παράδειγμα υπάρχουν 2 ομάδες (clusters).

Κεφάλαιο 5. Πειραματική Αξιολόγηση

Για την αξιολόγηση του αλγορίθμου συλλέξαμε 20 βίντεο συνεντεύξεων ως είσοδο.

Για διευκόλυνση της παρουσίασης των αποτελεσμάτων θα ταξινομήσουμε τα παρεκκλίνοντα πρόσωπα (outliers) που περιγράφηκαν στην ενότητα 3.4.1 σε κατηγορίες.

1. **Έλλειψη δειγμάτων.** Πρόσωπο με λίγα δείγματα ώστε να σχηματίσει δική του ομάδα.
2. **Λάθος εντοπισμός.** Εντοπισμός προσώπου που δεν υπάρχει πρόσωπο (false detection).
3. **Χαμηλή ποιότητα.** Επικαλυπτόμενα πρόσωπα, εκτός εστίασης, κτλ.
4. **Προφίλ.** Στραμμένο πρόσωπο (λήψη από πλάγια θέση).

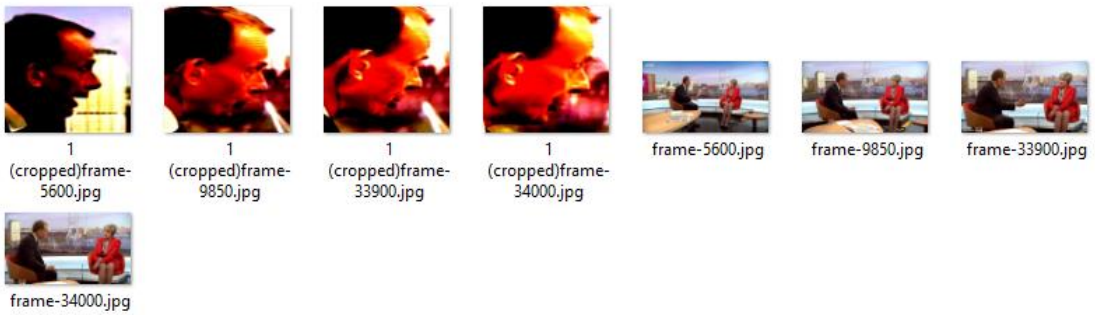
Πρέπει να σημειωθεί για την κατηγορία προφίλ, ότι αυτά τα πρόσωπα τα οποία είναι στραμμένα σε πολλές περιπτώσεις ανιχνεύονται ως ‘outliers’ ενώ κανονικά θα έπρεπε να ταξινομηθούν στην ίδια ομάδα με το ίδιο πρόσωπο εμπρόσθιας λήψης. Όμως, αυτό το γεγονός δεν θα το θεωρήσουμε ως λάθος της υλοποίησης επειδή ο τρόπος με τον οποίο εξάγει τα διανύσματα ο FaceNet είναι τέτοιος που αναθέτει διαφορετικές αναπαραστάσεις στα μεταξύ τους δείγματα και άρα τα πρόσωπα-προφίλ έχουν χαμηλό ‘silhouette score’.

Στη συνέχεια, εκτελούμε τον αλγόριθμο για τα 20 βίντεο, παρουσιάζοντας τα αποτελέσματα της ομαδοποίησης, καθώς και ορισμένες παρατηρήσεις σε ενδιαφέροντα σημεία.

5.1 Βίντεο 1

- Συνολικά δείγματα: 634
- Σύνολο προσώπων στο βίντεο: 2
- Αριθμός ομάδων που δημιουργήθηκαν: 2
- Αριθμός παρεκκλινόντων δειγμάτων που εντοπίστηκαν: 4
- Αριθμός παρεκκλινόντων δειγμάτων που δεν εντοπίστηκαν: 0
- Αριθμός δειγμάτων σε λάθος κατηγορία: 0
- Κατηγορία παρεκκλινόντων προσώπων: Προφίλ

Εικόνα 5.1.1. Παράθυρο εμφάνισης των αποτελεσμάτων.

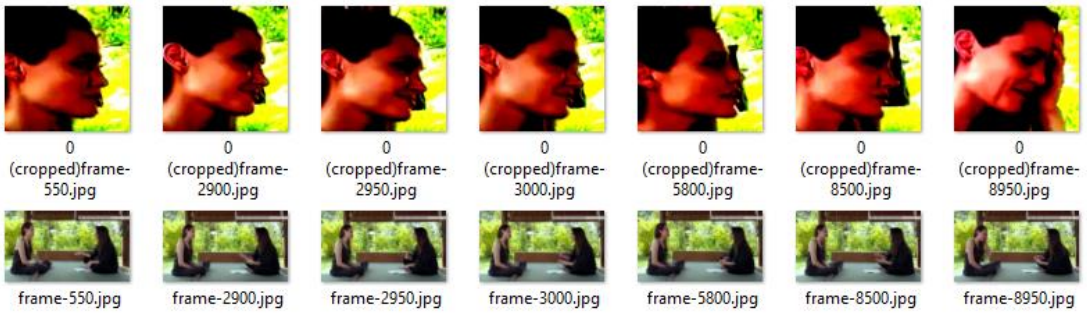


Εικόνα 5.1.2. Τα περιεχόμενα του φακέλου outliers.

5.2 Βίντεο 2

- Συνολικά δείγματα: 244
- Σύνολο προσώπων στο βίντεο: 2
- Αριθμός ομάδων που δημιουργήθηκαν: 2
- Αριθμός παρεκκλινόντων δειγμάτων που εντοπίστηκαν: 7
- Αριθμός παρεκκλινόντων δειγμάτων που δεν εντοπίστηκαν: 0
- Αριθμός δειγμάτων σε λάθος κατηγορία: 0
- Κατηγορία παρεκκλινόντων προσώπων: Προφίλ

Εικόνα 5.2.1. Παράθυρο εμφάνισης των αποτελεσμάτων.



Εικόνα 5.2.2. Τα περιεχόμενα του φακέλου outliers.

5.3 Βίντεο 3

- Συνολικά δείγματα: 109
- Σύνολο προσώπων στο βίντεο: 2
- Αριθμός ομάδων που δημιουργήθηκαν: 2
- Αριθμός παρεκκλινόντων δειγμάτων που εντοπίστηκαν: 0
- Αριθμός παρεκκλινόντων δειγμάτων που δεν εντοπίστηκαν: 0
- Αριθμός δειγμάτων σε λάθος κατηγορία: 0
- Κατηγορία παρεκκλινόντων προσώπων: -

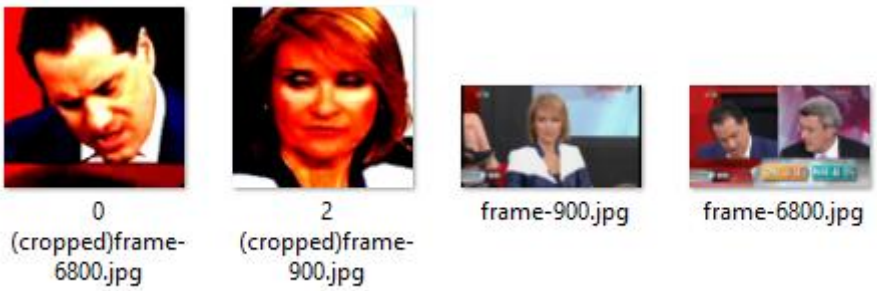
Εικόνα 5.3.1. Παράθυρο εμφάνισης των αποτελεσμάτων.

5.4 Βίντεο 4

- Συνολικά δείγματα: 183
- Σύνολο προσώπων στο βίντεο: 3
- Αριθμός ομάδων που δημιουργήθηκαν: 3
- Αριθμός παρεκκλινόντων δειγμάτων που εντοπίστηκαν: 2
- Αριθμός παρεκκλινόντων δειγμάτων που δεν εντοπίστηκαν: 0
- Αριθμός δειγμάτων σε λάθος κατηγορία: 0

- Κατηγορία παρεκκλινόντων προσώπων: Χαμηλής ποιότητας, Έλλειψη δειγμάτων

Εικόνα 5.4.1. Παράθυρο εμφάνισης των αποτελεσμάτων.



Εικόνα 5.4.2. Τα περιεχόμενα του φακέλου outliers. Το 1ο είναι χαμηλής ποιότητας. Το 2ο ανήκει στην κατηγορία έλλειψης δειγμάτων.

5.5 Βίντεο 5

- Συνολικά δείγματα: 277
- Σύνολο προσώπων στο βίντεο: 2
- Αριθμός ομάδων που δημιουργήθηκαν: 2
- Αριθμός παρεκκλινόντων δειγμάτων που εντοπίστηκαν: 0
- Αριθμός παρεκκλινόντων δειγμάτων που δεν εντοπίστηκαν: 0
- Αριθμός δειγμάτων σε λάθος κατηγορία: 0
- Κατηγορία παρεκκλινόντων προσώπων: -

Εικόνα 5.5.1. Παράθυρο εμφάνισης των αποτελεσμάτων.

5.6 Βίντεο 6

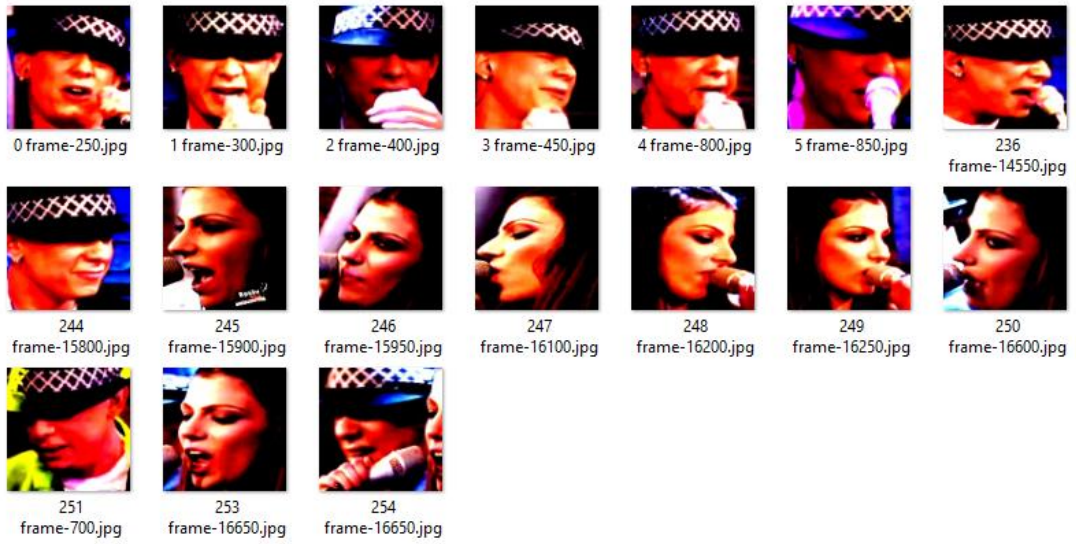
- Συνολικά δείγματα: 259
- Σύνολο προσώπων στο βίντεο: 4
- Αριθμός ομάδων που δημιουργήθηκαν: 3
- Αριθμός παρεκκλινόντων δειγμάτων που εντοπίστηκαν: 7
- Αριθμός παρεκκλινόντων δειγμάτων που δεν εντοπίστηκαν: 7
- Αριθμός δειγμάτων σε λάθος κατηγορία: 7

- Κατηγορία παρεκκλινόντων προσώπων: Χαμηλής ποιότητας, Λάθος εντοπισμός, Προφίλ

Αξίζει να σημειωθεί ότι, όπως φαίνεται και στην εικόνα 5.6.3, έπρεπε να εντοπιστούν 7 ακόμα ‘outliers’ και έτσι να χωριζόταν το ‘cluster’ σε δύο. Εκτιμούμε ότι ίσως για αυτό το λάθος ευθύνεται ο φωτισμός και το μικρόφωνο.

Εικόνα 5.6.1. Παράθυρο εμφάνισης των αποτελεσμάτων.

Εικόνα 5.6.2. Τα περιεχόμενα του φακέλου outliers. Τα πρώτα 5 δείγματα είναι χαμηλής ποιότητας. Το 6ο είναι λάθος εντοπισμός. Το 7ο είναι προφίλ.

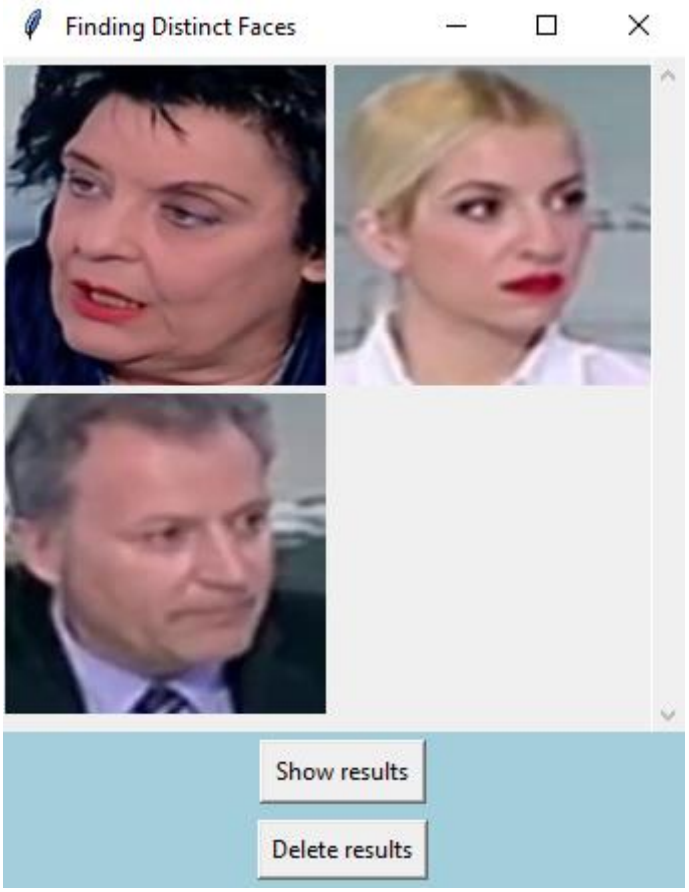


Εικόνα 5.6.3. Τα περιεχόμενα του φακέλου face cluster 2.

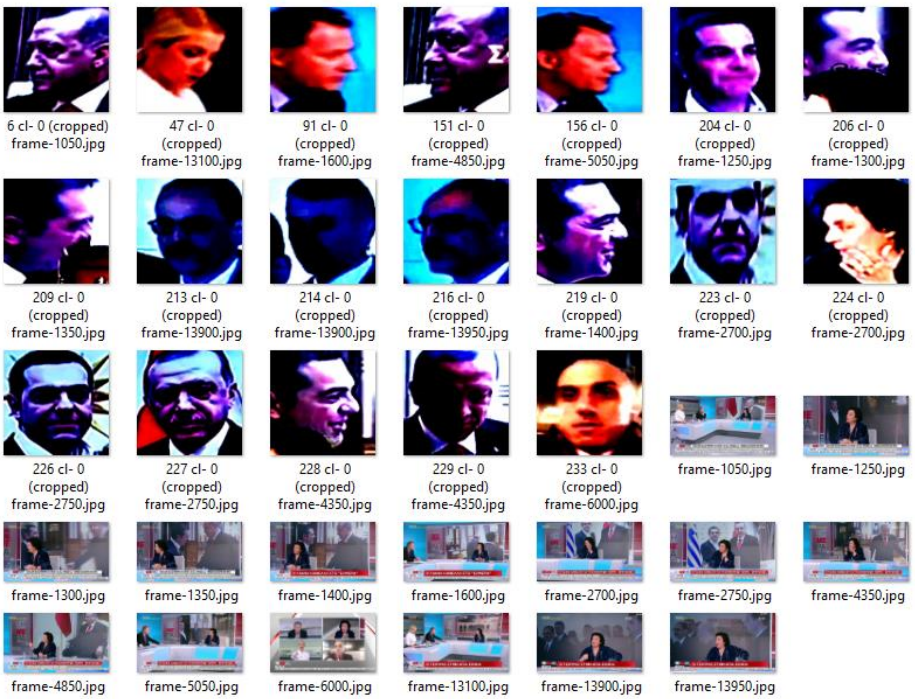
5.7 Βίντεο 7

- Συνολικά δείγματα: 233
- Σύνολο προσώπων στο βίντεο: 3
- Αριθμός ομάδων που δημιουργήθηκαν: 3
- Αριθμός παρεκκλινόντων δειγμάτων που εντοπίστηκαν: 19
- Αριθμός παρεκκλινόντων δειγμάτων που δεν εντοπίστηκαν: 7
- Αριθμός δειγμάτων σε λάθος κατηγορία: 0
- Κατηγορία παρεκκλινόντων προσώπων: Χαμηλής ποιότητας, Προφίλ

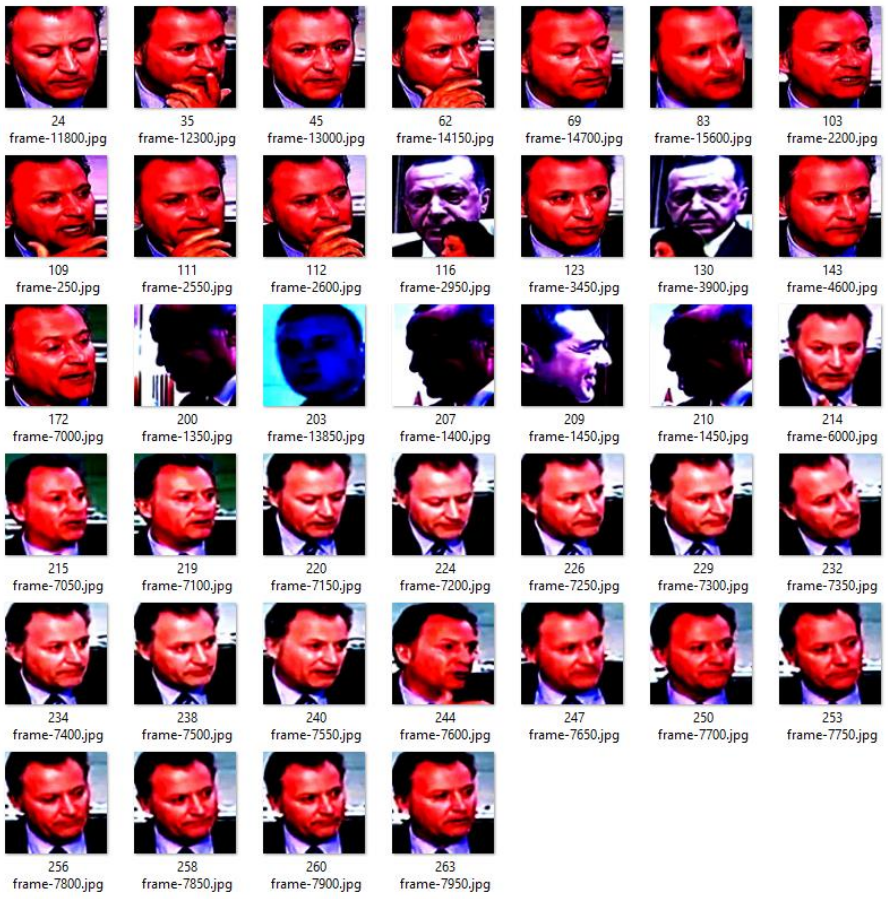
Στην εικόνα 5.7.3 δείχνουμε κάποια δείγματα τα οποία είναι ‘outliers’ αλλά δεν εντοπίστηκαν από τον έλεγχο.



Εικόνα 5.7.1. Παράθυρο εμφάνισης των αποτελεσμάτων.



Εικόνα 5.7.2. Τα περιεχόμενα του φακέλου outliers.



Εικόνα 5.7.3. Τα περιεχόμενα του φακέλου face cluster 2.

5.8 Βίντεο 8

- Συνολικά δείγματα: 91
- Σύνολο προσώπων στο βίντεο: 3
- Αριθμός ομάδων που δημιουργήθηκαν: 3
- Αριθμός παρεκκλινόντων δειγμάτων που εντοπίστηκαν: 6
- Αριθμός παρεκκλινόντων δειγμάτων που δεν εντοπίστηκαν: 0
- Αριθμός δειγμάτων σε λάθος κατηγορία: 0
- Κατηγορία παρεκκλινόντων προσώπων: Έλλειψη δειγμάτων

Εικόνα 5.8.1. Παράθυρο εμφάνισης των αποτελεσμάτων.

Εικόνα 5.8.2. Τα περιεχόμενα του φακέλου outliers.

5.9 Βίντεο 9

- Συνολικά δείγματα: 170
- Σύνολο προσώπων στο βίντεο: 7
- Αριθμός ομάδων που δημιουργήθηκαν: 7
- Αριθμός παρεκκλινόντων δειγμάτων που εντοπίστηκαν: 1
- Αριθμός παρεκκλινόντων δειγμάτων που δεν εντοπίστηκαν: 0
- Αριθμός δειγμάτων σε λάθος κατηγορία: 0
- Κατηγορία παρεκκλινόντων προσώπων: Χαμηλής ποιότητας

Όπως φαίνεται στην εικόνα 5.9.1 το πρόσωπο στην 2η ομάδα μοιάζει με αυτό της 5ης ομάδας. Στην εικόνα 5.9.3 φαίνεται ότι ο λόγος είναι ότι ο ηθοποιός παίζει διαφορετικούς χαρακτήρες στο σκετς και για αυτό έχει διαφορετική μύτη, χτένισμα και πηγούνι.

Εικόνα 5.9.1. Παράθυρο εμφάνισης των αποτελεσμάτων.

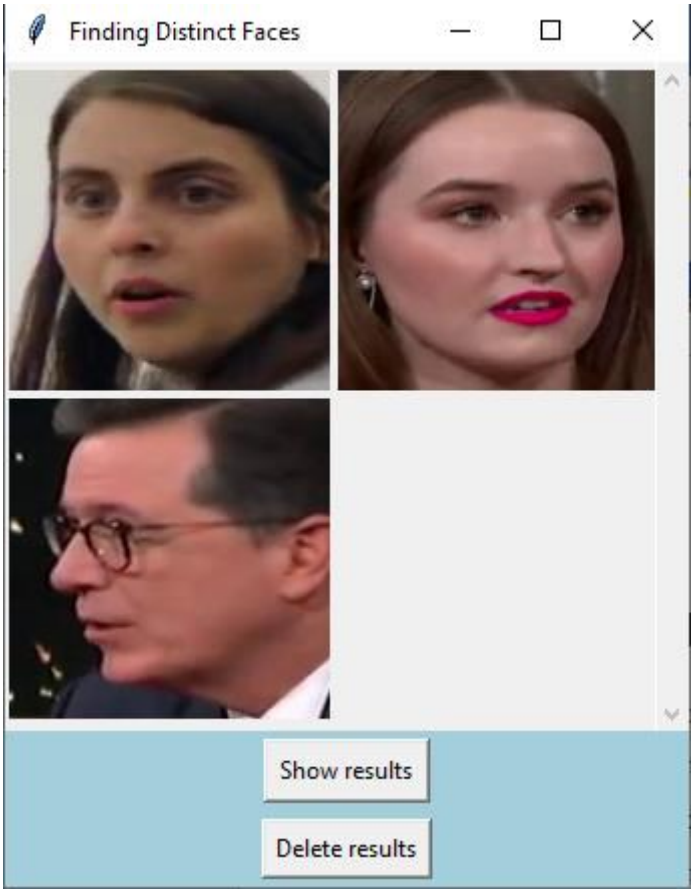
Εικόνα 5.9.2. Τα περιεχόμενα του φακέλου outliers.

Εικόνα 5.9.3. Σύγκριση των αντιπροσωπευτικών εικονοπλαισίων της 2ης και 5ης ομάδας.

5.10 Βίντεο 10

- Συνολικά δείγματα: 198
- Σύνολο προσώπων στο βίντεο: 3
- Αριθμός ομάδων που δημιουργήθηκαν: 3
- Αριθμός παρεκκλινόντων δειγμάτων που εντοπίστηκαν: 0
- Αριθμός παρεκκλινόντων δειγμάτων που δεν εντοπίστηκαν: 0
- Αριθμός δειγμάτων σε λάθος κατηγορία: 0
- Κατηγορία παρεκκλινόντων προσώπων: -

Αξίζει να σημειωθεί η ευρωστία που παρουσιάζει η μέθοδος στη διαφορά εμφάνισης της ηθοποιού, όπως φαίνεται στην εικόνα 5.10.2, όπου διαφορετικές εμφανίσεις ταξινομήθηκαν σωστά στην ίδια ομάδα.



Εικόνα 5.10.1. Παράθυρο εμφάνισης των αποτελεσμάτων.

Εικόνα 5.10.2. Δύο εικονοπλαίσια της 2ης ομάδας.

5.11 Βίντεο 11

- Συνολικά δείγματα: 303
- Σύνολο προσώπων στο βίντεο: 2
- Αριθμός ομάδων που δημιουργήθηκαν: 2
- Αριθμός παρεκκλινόντων δειγμάτων που εντοπίστηκαν: 2

- Αριθμός παρεκκλινόντων δειγμάτων που δεν εντοπίστηκαν: 0
- Αριθμός δειγμάτων σε λάθος κατηγορία: 0
- Κατηγορία παρεκκλινόντων προσώπων: Χαμηλής ποιότητας, Λάθος εντοπισμός

Εικόνα 5.11.1. Παράθυρο εμφάνισης των αποτελεσμάτων.

Εικόνα 5.11.2. Τα περιεχόμενα του φακέλου outliers.

5.12 Βίντεο 12

- Συνολικά δείγματα: 104
- Σύνολο προσώπων στο βίντεο: 8
- Αριθμός ομάδων που δημιουργήθηκαν: 8
- Αριθμός παρεκκλινόντων δειγμάτων που εντοπίστηκαν: 0
- Αριθμός παρεκκλινόντων δειγμάτων που δεν εντοπίστηκαν: 0
- Αριθμός δειγμάτων σε λάθος κατηγορία: 0

- Κατηγορία παρεκκλινόντων προσώπων: -

Εικόνα 5.12.1. Παράθυρο εμφάνισης των αποτελεσμάτων.

5.13 Βίντεο 13

- Συνολικά δείγματα: 45
- Σύνολο προσώπων στο βίντεο: 6
- Αριθμός ομάδων που δημιουργήθηκαν: 6

- Αριθμός παρεκκλινόντων δειγμάτων που εντοπίστηκαν: 0
- Αριθμός παρεκκλινόντων δειγμάτων που δεν εντοπίστηκαν: 0
- Αριθμός δειγμάτων σε λάθος κατηγορία: 0
- Κατηγορία παρεκκλινόντων προσώπων: -

Είναι αξιοσημείωτο το γεγονός ότι ομαδοποιήθηκε σωστά το πρόσωπο της φωτογραφίας με το πρόσωπο της ηθοποιού, ενώ έχει αρκετά διαφορετικά χαρακτηριστικά (βλ. εικόνα 5.13.2). Δηλαδή, ο αλγόριθμος παρουσίασε αντοχή στην μεγάλη διαφορά φωτεινότητας, χτενίσματος αλλά και ηλικίας του προσώπου.

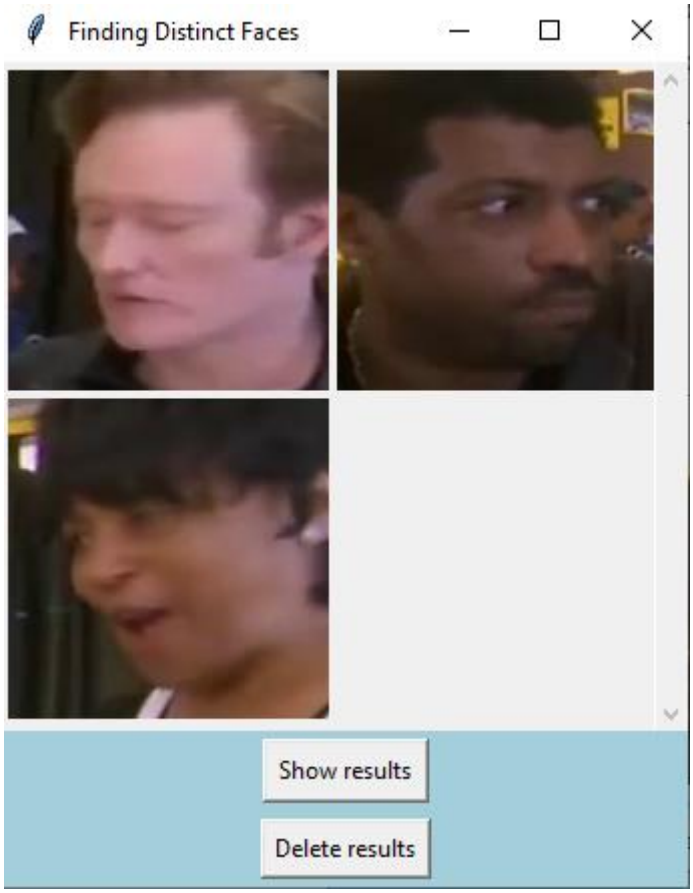
Εικόνα 5.13.1. Παράθυρο εμφάνισης των αποτελεσμάτων.

Εικόνα 5.13.2. Σύγκριση του αντιπροσωπευτικού προσώπου της 3ης ομάδας (αριστερά) με το εικονοπλαίσιο της φωτογραφίας (δεξιά). Το δεξί πρόσωπο της 2ης εικόνας ταξινομήθηκε στην 3η ομάδα.

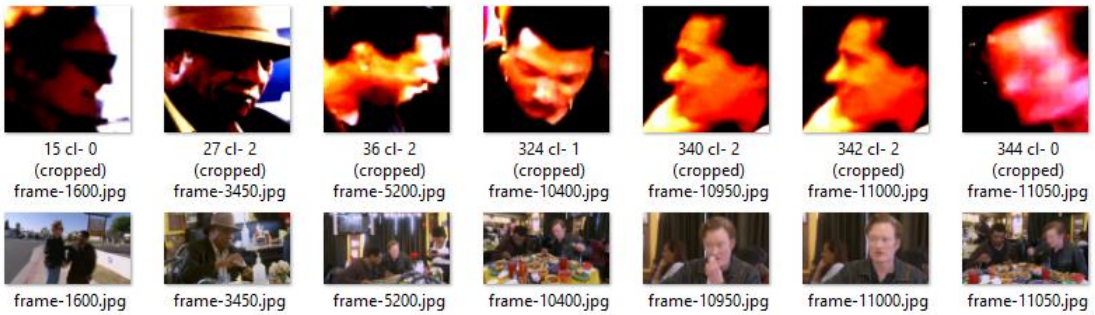
5.14 Βίντεο 14

- Συνολικά δείγματα: 232
- Σύνολο προσώπων στο βίντεο: 4
- Αριθμός ομάδων που δημιουργήθηκαν: 3
- Αριθμός παρεκκλινόντων δειγμάτων που εντοπίστηκαν: 7
- Αριθμός παρεκκλινόντων δειγμάτων που δεν εντοπίστηκαν: 0
- Αριθμός δειγμάτων σε λάθος κατηγορία: 7
- Κατηγορία παρεκκλινόντων προσώπων: Προφίλ, Χαμηλής ποιότητας, Έλλειψη δειγμάτων

Σημειώνουμε ότι υπάρχουν 2 σερβιτόρες στο βίντεο που συνενώθηκαν λανθασμένα σε μία ομάδα. Τα 7 δείγματα που είναι σε λάθος κατηγορία από τα 30 της 3ης ομάδας, αναφέρονται στην γυναίκα με τα γυαλιά της εικόνας 5.14.3.



Εικόνα 5.14.1. Παράθυρο εμφάνισης των αποτελεσμάτων.



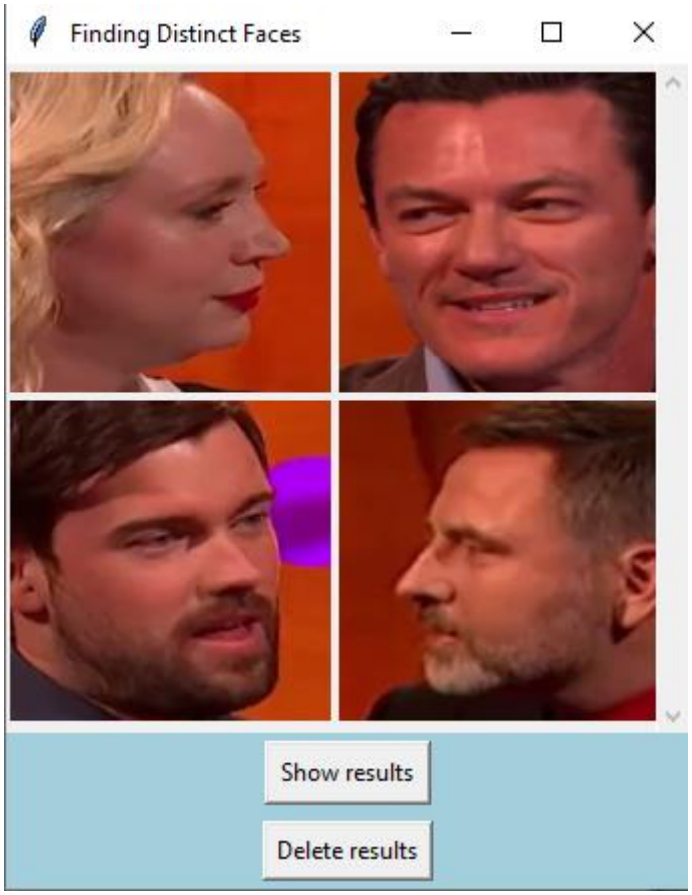
Εικόνα 5.14.2. Τα περιεχόμενα του φακέλου outliers. Η 1η είναι προφίλ, η 2η έλλειψη δειγμάτων και τα υπόλοιπα χαμηλής ποιότητας

Εικόνα 5.14.3. Λάθος ομαδοποίηση της γυναίκας με τα γυαλιά στην 3η ομάδα.

5.15 Βίντεο 15

- Συνολικά δείγματα: 31
- Σύνολο προσώπων στο βίντεο: 5
- Αριθμός ομάδων που δημιουργήθηκαν: 4
- Αριθμός παρεκκλινόντων δειγμάτων που εντοπίστηκαν: 0
- Αριθμός παρεκκλινόντων δειγμάτων που δεν εντοπίστηκαν: 0
- Αριθμός δειγμάτων σε λάθος κατηγορία: 0
- Κατηγορία παρεκκλινόντων προσώπων: -

Στο παρόν βίντεο υπάρχει ένα ακόμα πρόσωπο το οποίο δεν υπάρχει στα πρόσωπα που συλλέξαμε επειδή δεν εντοπίζεται από τον αλγόριθμο MTCNN (βλ. εικόνα 5.15.2). Αυτό ίσως συμβαίνει επειδή σε όλο το βίντεο εμφανίζεται στραμμένο υπό περίεργη γωνία.



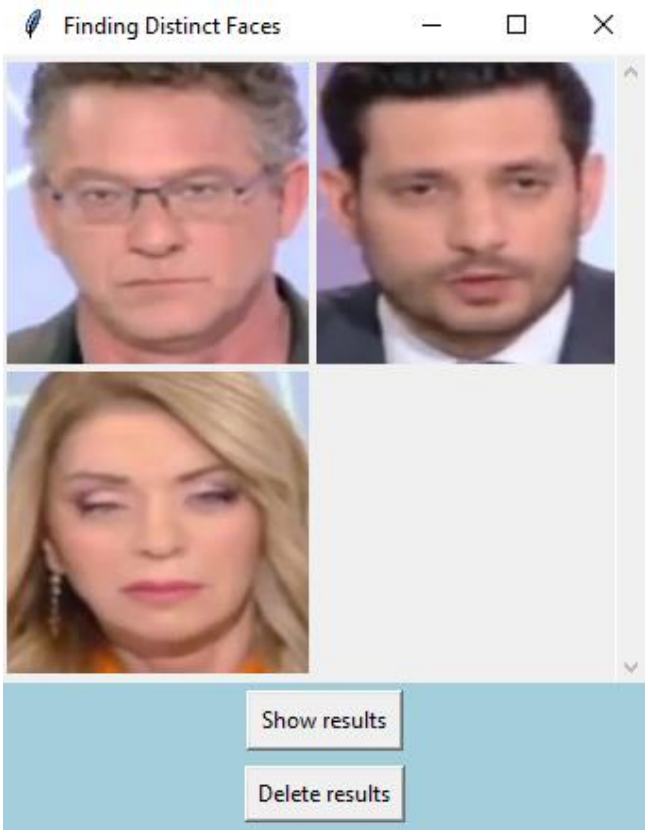
Εικόνα 5.15.1. Παράθυρο εμφάνισης των αποτελεσμάτων.

Εικόνα 5.15.2. Το πρόσωπο που δεν εντοπίζεται.

5.16 Βίντεο 16

- Συνολικά δείγματα: 140
- Σύνολο προσώπων στο βίντεο: 3
- Αριθμός ομάδων που δημιουργήθηκαν: 3
- Αριθμός παρεκκλινόντων δειγμάτων που εντοπίστηκαν: 0
- Αριθμός παρεκκλινόντων δειγμάτων που δεν εντοπίστηκαν: 0
- Αριθμός δειγμάτων σε λάθος κατηγορία: 0
- Κατηγορία παρεκκλινόντων προσώπων: -

Σε όλο το βίντεο τα άτομα εμφανίζονται σε πάνελ και άρα κάθε εικονοπλαίσιο έχει από 3 πρόσωπα (βλ. εικόνα 5.16.2).



Εικόνα 5.16.1. Παράθυρο εμφάνισης των αποτελεσμάτων.

Εικόνα 5.16.2. Ένα τυχαίο εικονοπλαίσιο του βίντεο.

5.17 Βίντεο 17

- Συνολικά δείγματα: 216
- Σύνολο προσώπων στο βίντεο: 2
- Αριθμός ομάδων που δημιουργήθηκαν: 2
- Αριθμός παρεκκλινόντων δειγμάτων που εντοπίστηκαν: 0
- Αριθμός παρεκκλινόντων δειγμάτων που δεν εντοπίστηκαν: 0
- Αριθμός δειγμάτων σε λάθος κατηγορία: 0
- Κατηγορία παρεκκλινόντων προσώπων: -

Εικόνα 5.17.1. Παράθυρο εμφάνισης των αποτελεσμάτων.

5.18 Βίντεο 18

- Συνολικά δείγματα: 254
- Σύνολο προσώπων στο βίντεο: 2
- Αριθμός ομάδων που δημιουργήθηκαν: 2
- Αριθμός παρεκκλινόντων δειγμάτων που εντοπίστηκαν: 3
- Αριθμός παρεκκλινόντων δειγμάτων που δεν εντοπίστηκαν: 0
- Αριθμός δειγμάτων σε λάθος κατηγορία: 0
- Κατηγορία παρεκκλινόντων προσώπων: Προφίλ, Έλλειψη δειγμάτων

Εικόνα 5.18.1. Παράθυρο εμφάνισης των αποτελεσμάτων.

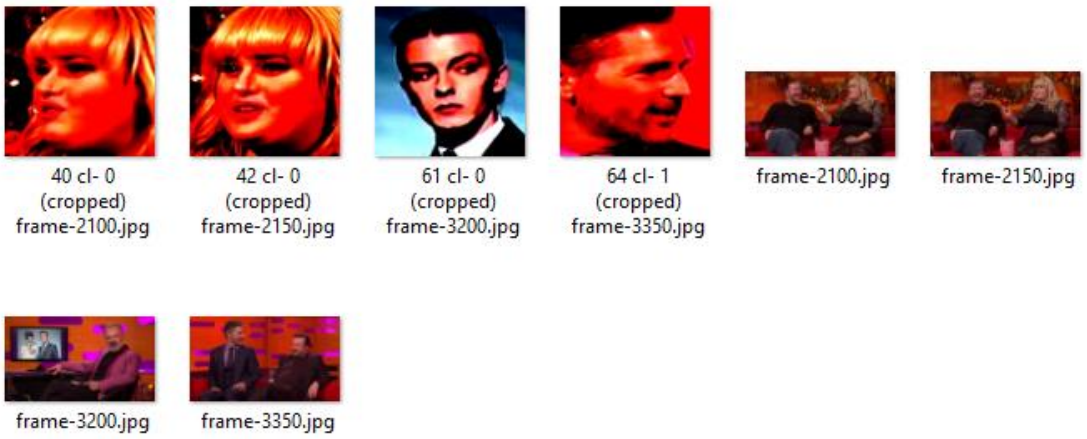
Εικόνα 5.18.2. Τα περιεχόμενα του φακέλου outliers.

5.19 Βίντεο 19

- Συνολικά δείγματα: 56
- Σύνολο προσώπων στο βίντεο: 4
- Αριθμός ομάδων που δημιουργήθηκαν: 2
- Αριθμός παρεκκλινόντων δειγμάτων που εντοπίστηκαν: 7
- Αριθμός παρεκκλινόντων δειγμάτων που δεν εντοπίστηκαν: 0
- Αριθμός δειγμάτων σε λάθος κατηγορία: 11
- Κατηγορία παρεκκλινόντων προσώπων: Προφίλ, Έλλειψης Δειγμάτων

Στην πρώτη ομάδα υπάρχουν δύο πρόσωπα τα οποία μοιάζουν μεταξύ τους, άλλα θα έπρεπε να είναι δύο ξεχωριστές ομάδες (βλ. εικόνα 5.19.3). Ένα ακόμα λάθος είναι ότι η γυναίκα της εικόνας 5.19.4, ενώ εμφανίζεται σε αρκετά καρέ, θεωρείται από τον αλγόριθμο ως 'outlier' λόγω έλλειψης δειγμάτων. Ο λόγος είναι ότι ο MTCNN την εντοπίζει μόνο σε δύο εικονοπλαίσια.

Εικόνα 5.19.1. Παράθυρο εμφάνισης των αποτελεσμάτων.



Εικόνα 5.19.2. Τα περιεχόμενα του φακέλου outliers.

Εικόνα 5.19.2. Τα δύο διαφορετικά πρόσωπα της πρώτης ομάδας

Εικόνα 5.19.3. Το δεξί πρόσωπο εντοπίζεται μόνο σε 2 εικονοπλαίσια από τα 7 στα οποία εμφανίζεται.

5.20 Βίντεο 20

- Συνολικά δείγματα: 129
- Σύνολο προσώπων στο βίντεο: 3
- Αριθμός ομάδων που δημιουργήθηκαν: 3
- Αριθμός παρεκκλινόντων δειγμάτων που εντοπίστηκαν: 5
- Αριθμός παρεκκλινόντων δειγμάτων που δεν εντοπίστηκαν: 0
- Αριθμός δειγμάτων σε λάθος κατηγορία: 0
- Κατηγορία παρεκκλινόντων προσώπων: Προφίλ, Χαμηλής ποιότητας

Σε αυτό το βίντεο δείχνουμε πως ανταποκρίνεται ο αλγόριθμος στο αποτέλεσμα ενός αλγόριθμου τύπου 'face-swap' που ονομάζεται DeepFake [11] ο οποίος αντικαθιστά το υπάρχον πρόσωπο με ένα άλλο. Στην περίπτωση αυτή, στο πρόσωπο του 3ου cluster αντικατέστησε το πρόσωπο γνωστού ηθοποιού και ο αλγόριθμός μας το κατέταξε στη 1η ομάδα.

Εικόνα 5.20.1. Παράθυρο εμφάνισης των αποτελεσμάτων.

Εικόνα 5.20.2. Τα περιεχόμενα του φακέλου outliers.

5.21 Συγκεντρωτικά Αποτελέσματα

Στον παρακάτω πίνακα παρουσιάζονται, για κάθε βίντεο που αξιολογήθηκε, τα στοιχεία που περιγράφηκαν.

Βίντεο:	Πραγματικός αρ. προσώπων (ground truth)	Αποτέλεσμα εκτίμησης προσώπων	Σύνολο Δειγμάτων	Εντοπι-σμένα outliers	Μη Εντοπι-σμένα outliers	Δείγμ. σε λάθος κατη-γορία
1.	2	2	634	4	0	0
2.	2	2	244	7	0	0
3.	2	2	109	0	0	0
4.	3	3	183	2	0	0
5.	2	2	277	0	0	0
6.	4	3	259	7	7	7
7.	3	3	233	19	7	0
8.	3	3	91	6	0	0
9.	7	7	170	1	0	0
10.	3	3	198	0	0	0
11.	2	2	303	2	0	0
12.	8	8	104	0	0	0
13.	6	6	45	0	0	0
14.	4	3	232	7	0	7
15.	5	4	31	0	0	0
16.	3	3	140	0	0	0
17.	2	2	216	0	0	0
18.	2	2	254	3	0	0
19.	4	2	56	7	0	11
20.	3	3	129	5	0	0

Πίνακας 5.1. Συγκεντρωτικά Αποτελέσματα

Κεφάλαιο 6. Επίλογος

6.1 Συμπεράσματα

Σε αυτή την εργασία ασχοληθήκαμε με την ομαδοποίηση όμοιων προσώπων και την εμφάνιση των διαφορετικών προσώπων ενός βίντεο μέσα από γραφικό περιβάλλον. Εκτιμήσαμε σωστά τον πραγματικό αριθμό προσώπων στα 16 από τα 20 βίντεο (κατά 80%). Καταφέραμε τον επιτυχή εντοπισμό των παρεκκλινόντων προσώπων στο 83,4% των περιπτώσεων, ενώ από το σύνολο των δειγμάτων μόλις 25 δείγματα ομαδοποιήθηκαν σε λάθος 'cluster'.

6.2 Μελλοντική Έρευνα

Έχοντας ως δεδομένο τα ενθαρρυντικά αποτελέσματα που προέκυψαν από τη πειραματική αξιολόγηση, υπάρχουν αρκετές ερευνητικές κατευθύνσεις που θα μπορούσαν να ακολουθηθούν ώστε να επεκταθεί και να βελτιωθεί η λειτουργία της μεθόδου.

Αρχικά, θα μπορούσε να γίνει τροποποίηση ώστε να αναγνωρίζει εάν υπάρχει μόνο ένα πρόσωπο στο βίντεο. Όπως είναι δομημένος τώρα ο αλγόριθμος αναγνωρίζει από 2 έως 12 όπως εξηγήθηκε στην ενότητα 3.3. Μια πρόταση για να υλοποιηθεί η αναγνώριση ενός προσώπου είναι στην περίπτωση που ανιχνευτούν 2 ομάδες, δηλαδή ο αριθμός silhouette είναι ο μεγαλύτερος για 2 ομάδες, αλλά είναι μικρός (πχ. 0.22) και ο αριθμός silhouette για 3 ομάδες είναι αρκετά μικρότερος (πχ 0.20) τότε υπάρχει ένδειξη ότι το βίντεο περιέχει μόνο ένα πρόσωπο.

Θα μπορούσε να προστεθεί ένα ενδιάμεσο βήμα μετά την προσπέλαση του βίντεο (εν. 3.1.1) και πριν τον εντοπισμό του προσώπου (εν. 3.1.2). Σε αυτό το βήμα θα επιλεχθούν

τα αντιπροσωπευτικά εικονοπλαίσια απορρίπτοντας εικονοπλαίσια μεγάλης ομοιότητας μεταξύ τους [12]. Με αυτόν τον τρόπο μειώνεται σημαντικά ο αποθηκευτικός χώρος και κυρίως γίνονται πιο γρήγορα οι υπόλοιποι υπολογισμοί του αλγορίθμου χωρίς κόστος στα αποτελέσματα της ομαδοποίησης.

Επίσης, μπορεί να επεκταθεί η λειτουργία του προγράμματος για την δημιουργία μίας βάσης δεδομένων προσώπων. Πιο συγκεκριμένα, για κάθε πρόσωπο, θα αποθηκεύεται στη βάση δεδομένων το όνομα του βίντεο από το οποίο εξάχθηκε, το λεπτό στο οποίο εντοπίστηκε και θα υπάρχει συσχέτιση με τις άλλες εικόνες, του ίδιου βίντεο ή και διαφορετικού, στις οποίες εμφανίζεται το ίδιο πρόσωπο. Έχοντας υλοποιήσει αυτή τη βάση δεδομένων, μένει να δημιουργηθεί και δομή ώστε να δέχεται νέα πρόσωπα και να τα συσχετίζει με αυτά που μοιάζουν αρκετά με τα υπάρχοντα της βάσης και αν δεν μοιάζουν με κανένα να δημιουργούνται νέες ομάδες. Με αυτή τη βάση δεδομένων είναι δυνατόν να αναζητήσουμε πρόσωπα μόνο με τα χαρακτηριστικά του προσώπου, δηλαδή, χωρίς την ανάγκη να γνωρίζουμε, για παράδειγμα, το όνομα του προσώπου όπως θα συνέβαινε σε μία συμβατική μηχανή αναζήτησης.

Βιβλιογραφία

[1] F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: A unified embedding for face recognition and clustering," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2015, vol. 07-12-June, pp. 815–823.

[2] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao, "Joint Face Detection and Alignment using Multi-task Cascaded Convolutional Networks," Apr. 2016.

[3] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, "Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning," Feb. 2016.

[4] Y. LeCun *et al.*, "Backpropagation Applied to Handwritten Zip Code Recognition," *Neural Comput.*, vol. 1, no. 4, pp. 541–551, Dec. 1989.

[5] J. C. Duchi, E. Hazan, and Y. Singer, *Adaptive Subgradient Methods for Online Learning and Stochastic Optimization*, vol. 12. 2011.

[6] J. A. Hartigan and M. A. Wong, "Algorithm AS 136: A K-Means Clustering Algorithm," *J. R. Stat. Soc. Ser. C (Applied Stat.*, vol. 28, no. 1, pp. 100–108, 1979.

[7] "TensorFlow." [Online]. Available: https://www.tensorflow.org/. [Accessed: 06-Apr-2019].

[8] "MS-Celeb-1M: Challenge of Recognizing One Million Celebrities in the Real World - Microsoft Research." [Online]. Available: https://www.microsoft.com/en-us/research/project/ms-celeb-1m-challenge-recognizing-one-million-celebrities-real-world/. [Accessed: 01-May-2019].

[9] "LFW." [Online]. Available: http://vis-www.cs.umass.edu/lfw/.

[10] Χ. Παπαδόπουλος, "Κώδικας για την εργασία: Εντοπισμός και Ομαδοποίηση προσώπων σε βίντεο." [Online]. Available: https://github.com/harpap/video-face-clusters.

[11] "DeepFaceLab." [Online]. Available:

https://github.com/iperov/DeepFaceLab.

[12] V. T. Chasanis, A. I. Ioannidis, and A. C. Likas, “Efficient Key-frame Extraction Based on Unimodality of Frame Sequences.”