

Συστήματα Πολυμέσων

Εργασία 2019

Παναγιώτης Ιωαννίδης Π16036

Αθανάσιος Παραβάντης Π16112

Περιεχόμενα

[2. Εισαγωγή 4](#_Toc10463291)

[3. Άσκηση 6.16 6](#_Toc10463292)

[3.1. Εκφώνηση 6](#_Toc10463293)

[3.2. Εκτέλεση 6](#_Toc10463294)

[3.3. Επεξήγηση 6](#_Toc10463295)

[4. Άσκηση 6.17 7](#_Toc10463296)

[4.1. Εκφώνηση 7](#_Toc10463297)

[4.2. Εκτέλεση 7](#_Toc10463298)

[4.3. Επεξήγηση 7](#_Toc10463299)

[5. Άσκηση 8.17 8](#_Toc10463300)

[5.1. Εκφώνηση 8](#_Toc10463301)

[5.2. Εκτέλεση 10](#_Toc10463302)

[5.3. Επεξήγηση 12](#_Toc10463303)

[6. Άσκηση 8.18 13](#_Toc10463304)

[6.1. Εκφώνηση 13](#_Toc10463305)

[6.2. Εκτέλεση 15](#_Toc10463306)

[6.3. Επεξήγηση 16](#_Toc10463307)

[7. Βοηθητικές συναρτήσεις 17](#_Toc10463308)

[7.1. frame\_to\_macroblocks(frame, window = 16) 17](#_Toc10463309)

[7.2. macroblocks\_to\_frame(macroblocks) 17](#_Toc10463310)

[7.3. fit\_size(x, window = 16) 18](#_Toc10463311)

[7.4. get\_sad(m\_prev, m\_next, window = 16): 18](#_Toc10463312)

[7.5. get\_best\_match(ms\_prev, next\_row, next\_col, m\_next, k = 16) 18](#_Toc10463313)

[8. Αναφορές 19](#_Toc10463314)

# Εισαγωγή

Στην εργασία του μαθήματος κάναμε την επιλογή να δουλέψουμε με τη γλώσσα προγραμματισμού Python 3.7 χρησιμοποιώντας μια σειρά από βιβλιοθήκες και δικές μας υλοποιήσεις συναρτήσεων. Για την εύρεση των εικόνων και βίντεο αναζητήσαμε στο διαδίκτυο κατάλληλο υλικό ώστε να βασίσουμε πάνω σε αυτό όλη την επεξεργασία που απαιτείται στις ασκήσεις. Επειδή η Python είναι μια δυνατή γλώσσα με πολλά εργαλεία και επαρκή τεκμηρίωση, θεωρήσαμε απαραίτητο να δομήσουμε την εργασία ως εξής:

* Ο φάκελος exercise-6.16 περιλαμβάνει τη λύση της άσκησης 6.16.
* Ο φάκελος exercise-6.17 περιλαμβάνει τη λύση της άσκησης 6.17.
* Ο φάκελος exercise-8.17 περιλαμβάνει τη λύση της άσκησης 8.17.
* Ο φάκελος exercise-8.18 περιλαμβάνει τη λύση της άσκησης 8.18.
* Ο φάκελος images περιέχει όλες τις φωτογραφίες που χρησιμοποιούνται στον κώδικα των ασκήσεων.
* Ο φάκελος videos περιέχει όλα τα βίντεο που χρησιμοποιούνται στον κώδικα των ασκήσεων.
* Ο φάκελος screenshots περιλαμβάνει φωτογραφίες από στιγμιότυπα εκτέλεσης των προγραμμάτων.
* Το αρχείο documentation.pdf είναι το παρόν αρχείο και παρέχει όλη την απαραίτητη τεκμηρίωση της εργασίας.

# Άσκηση 6.16

## Εκφώνηση

## Εκτέλεση

## Επεξήγηση

# Άσκηση 6.17

## Εκφώνηση

## Εκτέλεση

## Επεξήγηση

# Άσκηση 8.17

## Εκφώνηση

Στην άσκηση αυτή θα υλοποιήσετε την τεχνική της αντιστάθμισης κίνησης και θα μελετήσετε πώς επηρεάζει τα σφάλματα πρόβλεψης. Δείγματα αρχείων βίντεο, μαζί με τη σχετική πληροφορία διαμόρφωσης, παρέχονται στην ενότητα αρχείων προς λήψη του ιστοτόπου [www.cengage.com](http://www.cengage.com). Εκεί θα βρείτε και τον κώδικα για να διαβάζει και να προβάλλει πλαίσια βίντεο. Υποθέστε ότι το πρώτο πλαίσιο θα είναι πάντα ένα πλαίσιο I και ότι τα υπόλοιπα πλαίσια θα είναι τύπου P. Η υπόθεση αυτή ευσταθεί στην περίπτωση των σύντομων ακολουθιών βίντεο που επεξεργάζεστε, που έχουν μήκος το πολύ 100 πλαισίων.

* Στο πρώτο μέρος της άσκησης αυτής, υποθέστε ότι θέλετε να προβλέπετε ολόκληρα P πλαίσια και όχι κατά τμήματα. Η πρόβλεψη κάθε ολόκληρου πλαισίου γίνεται με βάση το προηγούμενο πλαίσιο. Υλοποιήστε μία διαδικασία που να δέχεται είσοδο δύο πλαίσια, υπολογίζει τη διαφορά τους και επιστρέφει ένα πλαίσιο σφαλμάτων. Δεν υπολογίζετε διάνυσμα κίνησης. Να προβάλετε τα πλαίσια σφαλμάτων. Σημειώστε ότι το πληροφοριακό περιεχόμενο του πλαισίου σφαλμάτων θα πρέπει να είναι μικρότερο, συγκρινόμενο με αυτό των πλαισίων.
* Στο δεύτερο βήμα θα υλοποιήσετε τεχνική πρόβλεψης κίνησης, η οποία θα υπολογίζει διανύσματα κίνησης ανά μπλοκ. Κάθε μπλοκ θα έχει το τυπικό MPEG μέγεθος 16 × 16. Υλοποιήστε μια συνάρτηση που θα δέχεται είσοδο δύο πλαίσια: ένα πλαίσιο αναφοράς, το οποίο θα χρησιμοποιηθεί κατά την αναζήτηση των διανυσμάτων κίνησης, και ένα πλαίσιο στόχος, το οποίο θα προβλεφθεί. Διαιρέστε το πλαίσιο-στόχο σε μακρομπλόκ μεγέθους 16 × 16. Εάν το πλάτος και ύψος του πλαισίου δεν είναι πολλαπλάσιο του 16, συμπληρώστε κατάλληλα το πλαίσιο με μαύρα εικονοστοιχεία. Για κάθε μπλοκ στο πλαίσιο-στόχο, ανατρέξτε στην αντίστοιχη θέση στο πλαίσιο αναφοράς και βρείτε την περιοχή που δίνει το καλύτερο ταίριασμα, όπως έχει εξηγηθεί στο κείμενο του κεφαλαίου. Χρησιμοποιήστε τη μετρική SAD σε περιοχή αναζήτησης που προκύπτει για k = 16, έτσι ώστε τα διανύσματα κίνησης να έχουν μέγεθος το πολύ 16 εικονοστοιχείων ως προς κάθε κατεύθυνση. Με βάση το μπλοκ πρόβλεψης, υπολογίστε το μπλοκ σφαλμάτων ως τη διαφορά μεταξύ του αρχικού μπλοκ και του προβλεφθέντος. Αφού αυτή η διαδικασία ολοκληρωθεί για όλα τα μπλοκ, θα προκύψει ένα πλαίσιο σφαλμάτων. Να γίνει η πρόβλεψη όλων των πλαισίων σφαλμάτων. Θα διαπιστώσετε ότι τα πλαίσια σφαλμάτων εμφανίζουν σημαντικά μικρότερη εντροπία σε σύγκριση με τη προηγούμενη περίπτωση, μολονότι απαιτείται περισσότερος χρόνος για τον υπολογισμό τους.

## Εκτέλεση

Η επίλυση της άσκησης βρίσκεται στον φάκελο exercise-8.17 και συγκεκριμένα στα αρχεία error\_frames.py και motion\_prediction.py.

Το πρώτο αρχείο error\_frames.py υλοποιεί τα πλαίσια σφαλμάτων μεταξύ των πλαισίων του βίντεο. Για να εκτελέσουμε τον κώδικα:

* python error\_frames.py

Παρατηρούμε ότι εμφανίζεται ένα παράθυρο που δείχνει για όλα τα πλαίσια B τη διαφορά τους με το προηγούμενο:

A screen shot of a computer

Description automatically generatedΜόλις προβληθούν όλα τα πλαίσια σφαλμάτων το πρόγραμμα τερματίζει.

Το δεύτερο αρχείο motion\_prediction.py υλοποιεί μια τεχνική πρόβλεψης κίνησης με τη βοήθεια της μετρικής SAD και λογαριθμικής αναζήτησης. Για να εκτελέσουμε τον κώδικα:

* python motion\_prediction.py

Αμέσως εμφανίζονται δυο παράθυρα, το πρώτο δείχνει τα δυο συγκρινόμενα πλαίσια και το δεύτερο τα μακρομπλόκ που συγκρίνονται και τη διαφορά τους:

A picture containing photo

Description automatically generatedΜόλις ο αλγόριθμος συγκρίνει όλα τα μακρομπλόκ του δεύτερου πλαισίου με το καλύτερο μακρομπλόκ του πρώτου, το πρόγραμμα τερματίζει.

## Επεξήγηση

Για την επίλυση της άσκησης χρησιμοποιούμε ένα διαφημιστικό βίντεο 5 δευτερολέπτων πάνω στο οποίο βασίζουμε όλη την επεξεργασία που γίνεται. Το αρχείο του βίντεο βρίσκεται στη τοποθεσία videos/video1.mp4. Στο πρώτο ερώτημα όσον αφορά τα πλαίσια σφαλμάτων, φορτώνουμε με τη βοήθεια της βιβλιοθήκης opencv-python όλα τα πλαίσια του βίντεο, αποθηκεύοντας πάντα το προηγούμενο (εκτός του πρώτου που είναι τύπου I). Η χρήσιμη συνάρτηση cv2.absdiff μας δίνει τη δυνατότητα να βρούμε τη διαφορά των δυο πλαισίων άμεσα και έπειτα γίνεται η προβολή τους.

Στο δεύτερο ερώτημα, παίρνουμε τα πρώτα δυο πλαίσια του βίντεο και τα χωρίζουμε σε μακρομπλόκ 16 × 16. Για να διαμορφωθούν όλα τα επιθυμητά μακρομπλόκ είναι αναγκαίο να μεταβάλλουμε το πλάτος και ύψος του βίντεο με τέτοιο τρόπο ώστε να διαιρούνται με το 16. Μόλις τελειώσει αυτή η διαδικασία, εφαρμόζουμε τον αλγόριθμο λογαριθμικής αναζήτησης με k = 16 για κάθε μακρομπλόκ του πλαισίου-στόχου ώστε να βρούμε το κατάλληλο πλαίσιο-αναφοράς χρησιμοποιώντας τη μετρική SAD. Τέλος, προβάλλουμε τα δυο πλαίσια σε ένα παράθυρο και όλα τα πλαίσια στόχων, αναφοράς και σφαλμάτων που προέκυψαν.

# Άσκηση 8.18

## Εκφώνηση

Σε αυτήν την άσκηση θα δείτε ότι η τεχνική της τμηματικής πρόβλεψης με βάση την αντιστάθμιση κίνησης, μπορεί επίσης να χρησιμοποιηθεί σε εφαρμογές εκτός συμπίεσης. Μία τέτοια ενδιαφέρουσα εφαρμογή είναι η απομάκρυνση αντικειμένων ή προσώπων από τη ροή του βίντεο. Για παράδειγμα, έστω ένα βίντεο στο οποίο η κάμερα δεν έχει κινηθεί και το παρασκήνιο είναι σχετικά στατικό, αλλά κινούνται ορισμένα αντικείμενα στο προσκήνιο. Στόχος σας είναι να προσεγγίσετε το αντικείμενο χρησιμοποιώντας μπλοκ και στη συνέχεια να αντικαταστήσετε αυτά τα μπλοκ με παρασκήνιο, σαν να μην ήταν ποτέ παρόν το αντικείμενο. Στη γενική περίπτωση, η λύση είναι πολύ δύσκολη, αλλά στο πλαίσιο αυτής της άσκησης θα επεξεργαστείτε ορισμένες απλούστερες ιδέες. Κατά την υλοποίηση, βεβαιωθείτε ότι μπορείτε να χειρίζεστε το μέγεθος του μπλοκ ως παράμετρο, προκειμένου να ελέγξετε πόσο καλά λειτουργεί ο αλγόριθμος απομάκρυνσης αντικειμένων για διάφορα μεγέθη μακρομπλόκ.

* Κατ’ αρχήν, φορτώστε ένα σύνολο πλαισίων βίντεο. Υποθέστε ότι το πρώτο πλαίσιο είναι αποκλειστικά πλαίσιο παρασκηνίου και δεν περιέχει αντικείμενα σε κίνηση. Δοθέντος ενός πλαισίου, n, προχωρήστε στη διαίρεσή του σε μπλοκ. Υπολογίστε ένα διάνυσμα κίνησης ανά μπλοκ με βάση το προηγούμενο πλαίσιο (αναφοράς). Για τα μπλοκ παρασκηνίου, θα πρέπει να προκύψουν μη μηδενικά διανύσματα κίνησης. Παρακολουθήστε όλα αυτά τα διανύσματα κίνησης. Μπορείτε ακόμη και να τα απεικονίσετε για κάθε μπλοκ, καθώς προβάλετε το βίντεο.
* Στη συνέχεια, βρείτε τα μπλοκ που αντιστοιχούν σε μη μηδενικά διανύσματα κίνησης. Αντικαταστήστε κάθε τέτοιο μπλοκ με το αντίστοιχο μπλοκ παρασκηνίου. Τα εικονοστοιχεία παρασκηνίου θα πρέπει να προέρχονται από προηγούμενο πλαίσιο, στο οποίο δεν υπήρχε αντικείμενο σε κίνηση. Η αντικατάσταση όλων αυτών των μπλόκ θα οδηγήσει στην απομάκρυνση των αντικειμένων, διότι τα αντικείμενα αντικαθίστανται από το παρασκήνιο. Επαναλάβετε τη διαδικασία, χρησιμοποιώντας μακρομπλόκ διαφορετικού μεγέθους.
* Είναι πιθανό να αντιμετωπίσετε ασυνέχειες και παρενέργειες στα όρια των μπλοκ που αντικαθίστανται. Πως θα ελαχιστοποιήσετε αυτά τα φαινόμενα;

Η προτεινόμενη λύση θα λειτουργήσει ικανοποιητικά, μόνο όταν υποθέστε ότι η κάμερα δεν κινείται, όταν τα αντικείμενα που κινούνται υπόκεινται σε λεία κίνηση και ότι δεν παρατηρούνται αλλαγές κατεύθυνσης. Όμως στη γενική περίπτωση, η κάμερα κινείται, τα αντικείμενα κινούνται με όχι αυστηρό τρόπο και επιπλέον, παρατηρούνται αλλαγές στον φωτισμό, καθώς η κάμερα κινείται.

* Αλλά ας υποθέσουμε ότι μπορείτε να εντοπίσετε όλα τα μακρομπλόκ ενός πλαισίου που αντιστοιχούν στο παρασκήνιο. Πως μπορείτε να αξιοποιήσετε αυτό το γεγονός, πλέον της χρήσης αντιστάθμισης κίνησης σε επίπεδο μακρομπλόκ;

## Εκτέλεση

Η επίλυση της άσκησης βρίσκεται στον φάκελο exercise-8.18 και συγκεκριμένα στο αρχείο object\_removal.py. Εκτελούμε τον κώδικα με την εντολή:

* python object\_removal.py

Παρατηρούμε αμέσως ότι εμφανίζονται τέσσερα παράθυρα:

* Το πρώτο παράθυρο απεικονίζει το αρχικό βίντεο με τους χαρακτήρες.
* Το δεύτερο παράθυρο απεικονίζει το βίντεο με τους εξαφανισμένους χαρακτήρες στο οποίο το μέγεθος μακρομπλόκ είναι ίσο με 16.
* Το τρίτο παράθυρο απεικονίζει το βίντεο με τους εξαφανισμένους χαρακτήρες στο οποίο το μέγεθος μακρομπλόκ είναι ίσο με 8.
* Το τέταρτο παράθυρο απεικονίζει το παράθυρο με τους εξαφανισμένους χαρακτήρες στο οποίο το μέγεθος μακρομπλόκ είναι ίσο με 4.

A flat screen tv

Description automatically generatedΜόλις προβληθούν όλα τα πλαίσια του βίντεο το πρόγραμμα τερματίζει.

## Επεξήγηση

Σε αυτή την άσκηση χρησιμοποιήσαμε ένα βίντεο με καρτούν 10 δευτερολέπτων. Στόχος μας είναι η απομάκρυνση των τριών χαρακτήρων που κινούνται στο προσκήνιο με μακρομπλόκ από το πρώτο πλαίσιο του βίντεο, πριν εμφανιστούν. Επιλέξαμε το πρώτο πλαίσιο επειδή τα διανύσματα κίνησης είναι σχεδόν μηδενικά λόγω του ότι το παρασκήνιο είναι στατικό. Αφού χωρίσαμε το βίντεο σε μακρομπλόκ μεγέθους 16, 8 και 4, το κάθε επόμενο πλαίσιο δανείζεται τα μακρομπλόκ του προηγούμενου από τη μέση και κάτω. Εφόσον το δεύτερο πλαίσιο παίρνει τα μακρομπλόκ παρασκηνίου του πρώτου, το τρίτο του δεύτερου κ.ο.κ. έχουμε ως αποτέλεσμα την αφαίρεση των κινούμενων χαρακτήρων από το βίντεο.

Αξίζει να σημειωθεί ότι η τεχνική αυτή έχει κάποιες παρενέργειες στο βίντεο, ένας προσεκτικός παρατηρητής μπορεί να διαπιστώσει ότι από τη μέση και κάτω υπάρχει μια απότομη αλλαγή στα εικονοστοιχεία του παρασκηνίου. Επίσης, εάν το παρασκήνιο είχε κάποια επιπλέον στατικά χαρακτηριστικά, αυτό θα απαιτούσε μια πιο προσεκτική αντικατάσταση των μακρομπλόκ. Τέλος, σε ένα πραγματικό παράδειγμα, μπορεί να υπήρχαν αλλαγές στη φωτεινότητα πράγμα που δεν μας απασχολεί σε αυτή τη περίπτωση.

Για να αντιμετωπίσουμε τις παρενέργειες στα όρια των μακρομπλόκ προτείνουμε τις εξής λύσεις:

* Έξυπνη αντικατάσταση από μακρομπλόκ του ίδιο και όχι προηγούμενου πλαισίου. Αυτό πρακτικά σημαίνει ότι εφόσον οι χαρακτήρες δεν καλύπτουν όλο το παρασκήνιο, μπορούμε να πάρουμε μέρους του από το ίδιο στιγμιότυπο.
* Συνδυασμός μακρομπλόκ από το ίδιο και προηγούμενα πλαίσια ώστε να λαμβάνεται υπ’ όψη ο φωτισμός και άλλα χαρακτηριστικά του παρασκηνίου που μπορεί να μην είναι διακριτά σε ένα διάστημα πλαισίων της κίνησης των χαρακτήρων.

# Βοηθητικές συναρτήσεις

Το αρχείο functions.py που βρίσκεται στους φακέλους exercise-8.17 και exercise-8.18 περιέχει χρήσιμες συναρτήσεις που χρησιμοποιούνται στον κώδικα των ασκήσεων. Ας δούμε λεπτομερώς πως λειτουργούν.

## frame\_to\_macroblocks(frame, window = 16)

Η συνάρτηση δέχεται ένα οποιοδήποτε πλαίσιο (frame) και επιστρέφει έναν πίνακα που περιέχει όλα τα μακρομπλόκ. Η προεπιλεγμένη τιμή των διαστάσεών τους είναι 16 (window).

Βήματα για την δημιουργία των μακρομπλόκ:

* Μεταβολή του πλάτους και ύψους του εκάστοτε πλαισίου ώστε να μπορέσουμε να διαιρέσουμε την εικόνα σε 16 × 16 μακρομπλόκ (προεπιλογή).
* Εάν το πλαίσιο που προκύψει έχει μεγαλύτερη διάσταση από το αρχικό, συμπληρώνεται με μαύρα εικονοστοιχεία.
* Διατρέχουμε όλες τις γραμμές και στήλες του νέου πλαισίου με βήμα ανάλογο με το μέγεθος του μακρομπλόκ.
* Αποθηκεύουμε το μακρομπλόκ στην εκάστοτε γραμμή και στήλη του πίνακα αποτελεσμάτων.

Τέλος επιστρέφουμε τον πίνακα αποτελεσμάτων της μορφής: (γραμμή, στήλη, μακρομπλόκ, 16, 16, 3).

## macroblocks\_to\_frame(macroblocks)

Η συνάρτηση δέχεται ένα αποτέλεσμα της συνάρτησης frame\_to\_macroblocks και επιστρέφει τον ανακατασκευασμένο πλαίσιο.

Διατρέχουμε τον πίνακα αποτελεσμάτων των μακρομπλόκ ανά γραμμή και στήλη και τοποθετούμε όλα τα pixel ξανά σε διαστάσεις πλαισίου.

Επιστρέφουμε το νέο πλαίσιο που μπορούμε έπειτα να το προβάλλουμε σαν εικόνα.

## fit\_size(x, window = 16)

Επιστρέφει έναν ακέραιο αριθμό y που διαιρείται από το μέγεθος του μακρομπλόκ (window) και είναι ο πλησιέστερος μεγαλύτερος αριθμός στον x.

## get\_sad(m\_prev, m\_next, window = 16):

Η συνάρτηση δέχεται δυο μακρομπλόκ, ενός πλαισίου αναφοράς (m\_prev) και ενός πλαισίου στόχου (m\_next) αντίστοιχα. Επιστρέφει την μετρική τιμή SAD για αυτά τα δυο μακρομπλόκ. Το προεπιλεγμένο μέγεθος των μακρομπλόκ είναι 16 (window).

Διατρέχουμε τους δυο πίνακες μακρομπλόκ ανά εικονοστοιχείο και προσθέτουμε στο σκορ την απόλυτη διαφορά ανά χρωματική συνιστώσα.

Το αποτέλεσμα που επιστρέφουμε είναι το άθροισμα των απόλυτων διαφορών από όλα τα εικονοστοιχεία.

## get\_best\_match(ms\_prev, next\_row, next\_col, m\_next, k = 16)

Εφαρμόζει λογαριθμική αναζήτηση για ένα μακρομπλόκ (m\_next) που βρίσκεται στη θέση (next\_row, next\_col) με βάση τα μακρομπλόκ του πλαισίου αναφοράς (ms\_prev).

Τα βήματα αντιστοίχισης μακρομπλόκ:

* Συλλέγουμε τις 8 γειτονικές θέσεις γύρω από το μακρομπλόκ του πλαισίου στόχου εφόσον υπάρχουν.
* Υπολογίζουμε τις μετρικές SAD για τα 8 γειτονικά μακρομπλόκ.
* Επαναπροσδιορίζουμε το κέντρο αναζήτησης στο γειτονικό μακρομπλόκ με το μικρότερο σκορ SAD.
* Επαναλαμβάνουμε τη διαδικασία με k = 16 (προεπιλογή) μειώνοντας κάθε φορά το k στο μισό, μέχρι το k να γίνει 1.

Τέλος, επιστρέφουμε το μακρομπλόκ που επιλέχθηκε μετά από αλλεπάλληλες αλλαγές του κέντρου αναζήτησης.

# Αναφορές

*Sesame Street - Mad*. (n.d.). Retrieved from YouTube: https://www.youtube.com/watch?v=D6oINRXLt-8

Tips, P. (n.d.). *PG tips Green Tea*. Retrieved from YouTube: https://www.youtube.com/watch?v=mEh5GiQ4\_Yk