

ΕΡΓΑΣΤΗΡΙΑ 11 ΚΗ ΑΣΚΗΣΗ



Τεχνικές Κατωφλίωσης Εικόνας

1. Εισαγωγή

Ο όρος *μονοχρωματική εικόνα* ή απλά *εικόνα* αναφέρεται σε μια δισδιάστατη συνάρτηση έντασης φωτός $f(x, y)$, όπου x και y είναι οι χωρικές συντεταγμένες και η τιμή της f σε οποιοδήποτε σημείο (x, y) είναι ανάλογη της *φωτεινότητας* ή του *επιπέδου του γκρι* (gray level) της εικόνας σε εκείνο το σημείο.

Μια *ψηφιακή εικόνα* είναι μια εικόνα της οποίας τόσο οι χωρικές συντεταγμένες, όσο και το πεδίο τιμών της f έχουν διακριτοποιηθεί. Πιο συγκεκριμένα, έστω (x, y) οι χωρικές συντεταγμένες μιας ψηφιακής εικόνας και έστω $G = \{0, 1, \dots, L-1\}$ ένα σύνολο μη αρνητικών ακεραίων που αντιπροσωπεύουν τα επίπεδα φωτεινότητας, με το 0 να αντιστοιχεί στο μαύρο χρώμα και το $L-1$ στο λευκό. Η ψηφιακή εικόνα μπορεί να οριστεί τότε σαν μια απεικόνιση $f : \mathcal{X} \times \mathcal{Y} \rightarrow G$, όπου \mathcal{X} είναι το σύνολο των φυσικών αριθμών. Η φωτεινότητα ενός *εικονοστοιχείου* (picture element ή pixel) με συντεταγμένες (x, y) υποδηλώνεται σαν $f(x, y)$. Πρακτικά η ψηφιακή εικόνα μπορεί να θεωρηθεί σαν ένας δισδιάστατος πίνακας του οποίου οι δείκτες γραμμής και στήλης καθορίζουν ένα σημείο στην εικόνα και η τιμή του αντίστοιχου στοιχείου του πίνακα καθορίζει το επίπεδο φωτεινότητας της εικόνας στο εν λόγω σημείο.

Το *ιστόγραμμα* μιας ψηφιακής εικόνας είναι μια ακολουθία:

$$p_i = \frac{n_i}{n}, \quad 0 \leq i \leq L-1, \quad (1)$$

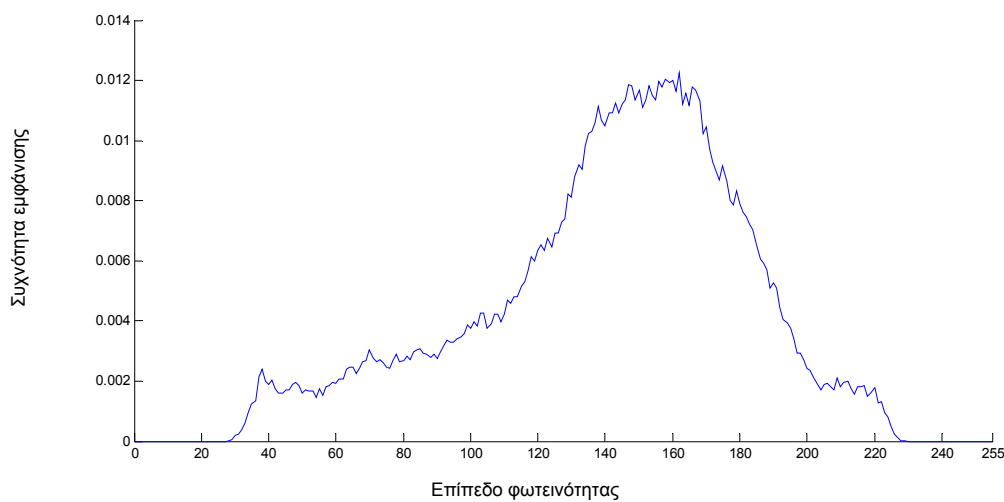
όπου n_i είναι το πλήθος των εικονοστοιχείων της εικόνας με επίπεδο φωτεινότητας i και n είναι το συνολικό πλήθος των εικονοστοιχείων. Η ποσότητα p_i δηλώνει τη συχνότητα εμφάνισης του

επίπεδο φωτεινότητας i στην εικόνα, ενώ το συνολικό ιστόγραμμα της εικόνας μπορεί να θεωρηθεί σαν μια εκτίμηση της σ.π.π. της φωτεινότητας, $p(z)$. Στο Σχήμα 1.α παρουσιάζεται μια ψηφιακή εικόνα με 256 επίπεδα φωτεινότητας ενώ στο Σχήμα 1.β έχει σχεδιαστεί το ιστόγραμμά της.

Ας θεωρήσουμε τώρα την περίπτωση όπου το περιεχόμενο μιας ψηφιακής εικόνας αποτελείται από «σκούρα» αντικείμενα και «φωτεινό» φόντο¹, έτσι ώστε το επίπεδο φωτεινότητας των εικονοστοιχείων που ανήκουν σε περιοχές αντικειμένων να κυμαίνεται γύρω από μία τιμή μ_0 , ενώ το επίπεδο φωτεινότητας των εικονοστοιχείων που ανήκουν σε περιοχές φόντου να κυμαίνεται γύρω από μία τιμή μ_1 , με $\mu_0 < \mu_1$, $\mu_0, \mu_1 \in G$. Το ιστόγραμμα επομένως μιας τέτοιας εικόνας αναμένεται να αποτελείται από δύο «λόφους», έναν με κορυφή στο μ_0 και έναν με κορυφή στο μ_1 . Μια κατηγορία εικόνων με τα παραπάνω χαρακτηριστικά είναι οι εικόνες κειμένου, όπως για παράδειγμα εκείνη του Σχήματος 2.α, το ιστόγραμμα της οποίας έχει σχεδιαστεί στο Σχήμα 2.β.



(α)

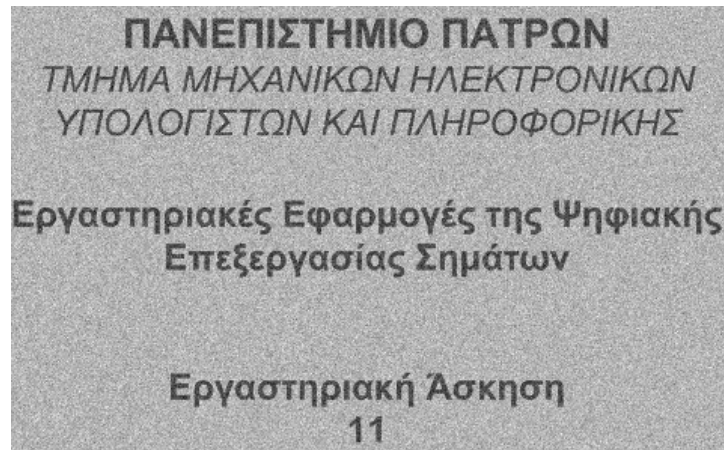


(β)

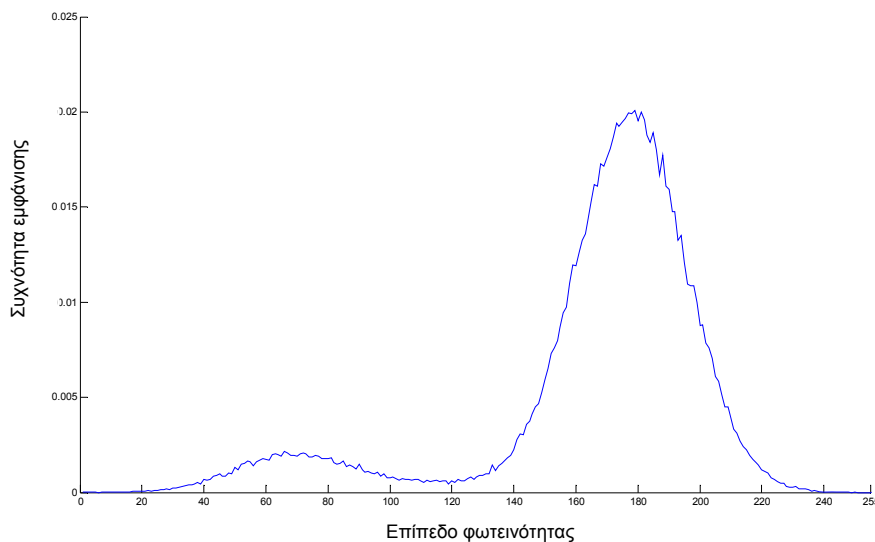
Σχήμα 1. (α): Μια μονοχρωματική ψηφιακή εικόνα. (β): Το ιστόγραμμά της.

¹ Η ίδια ανάλυση ακολουθείται και στην αντίθετη περίπτωση, όταν δηλαδή έχουμε «φωτεινά» αντικείμενα και «σκοτεινό» φόντο.

Σε τέτοιες περιπτώσεις, ένας προφανής τρόπος να διαχωρίσουμε τα αντικείμενα (για παράδειγμα τους χαρακτήρες στην παραπάνω εικόνα) από το φόντο είναι ο καθορισμός ενός κατωφλίου $t \in G$ έτσι ώστε κάθε σημείο (x, y) να χαρακτηρίζεται ως σημείο αντικειμένου εάν $f(x, y) \leq t$ και ως σημείο φόντου, εάν $f(x, y) > t$. Στην κατωφλιωμένη εικόνα, το επίπεδο φωτεινότητας των εικονοστοιχείων των αντικειμένων τίθεται ίσο με μια τιμή g_α , ενώ το επίπεδο φωτεινότητας των εικονοστοιχείων του φόντου τίθεται ίσο με μια τιμή g_ϕ .



(α)



(β)

Σχήμα 2. (α): Μια ενθόρυβη εικόνα κειμένου. (β): Το ιστόγραμμα της.

Για τον προσδιορισμό της «βέλτιστης» τιμής κατωφλίου έχει προταθεί στη βιβλιογραφία ένας πολύ μεγάλος αριθμός μεθόδων κατωφλίωσης βασισμένων σε διαφορετικά κριτήρια. Στα πλαίσια αυτής της άσκησης θα γνωρίσουμε μερικές από τις πιο αντιπροσωπευτικές μεθόδους κατωφλίωσης εικόνας αρχίζοντας από τις πιο απλές.

2. Η Μέθοδος της Κατωφλίωσης με Ποσόστωση (Percentile Method)

Η κατωφλίωση με ποσόστωση αποτελεί μια από τις πιο πρώιμες και απλές μεθόδους κατωφλίωσης. Στη συγκεκριμένη μέθοδο θεωρούμε ότι η εικόνα αποτελείται από σκοτεινά αντικείμενα σε φωτεινό φόντο (ή το αντίστροφο) και ότι το ποσοστό της συνολικής έκτασης της εικόνας που καταλαμβάνεται από τα αντικείμενα είναι γνωστό και ίσο με $p\%$. Ως κατώφλι επιλέγεται τότε η ελάχιστη (μέγιστη) τιμή επιπέδου φωτεινότητας που κατανέμει τουλάχιστον το $p\%$ των εικονοστοιχείων της κατωφλιωμένης εικόνας σε περιοχές που καταλαμβάνονται από αντικείμενα.

Ας υποθέσουμε για παράδειγμα ότι έχουμε ψηφιοποιήσει μια σελίδα ενός βιβλίου με χρήση ενός σαρωτή και ότι γνωρίζουμε ότι το 40% της έκτασης της σελίδας καταλαμβάνεται από μαύρους χαρακτήρες ενώ το υπόλοιπο είναι λευκό. Καθότι το πιθανότερο είναι η εικόνα που θα λάβουμε να μην αποτελείται αποκλειστικά από μαύρα και λευκά εικονοστοιχεία, αλλά από «σκοτεινά» και «φωτεινά», μπορούμε να εκμεταλλευτούμε το γεγονός ότι γνωρίζουμε το ποσοστό των εικονοστοιχείων που ανήκουν σε χαρακτήρες και να μετατρέψουμε την εικόνα σε δυαδική. Στη συγκεκριμένη περίπτωση, η κατωφλίωση επιτυγχάνεται απεικονίζοντας το σκοτεινότερο 40% της εικόνας με μαύρο και το υπόλοιπο με λευκό.

3. Η Μέθοδος Hill–Clustering

Η συγκεκριμένη προσέγγιση βασίζεται στην υπόθεση ότι το ιστόγραμμα της εικόνας αποτελείται από μια σειρά «λόφων», κάθε ένας από τους οποίους αντιστοιχεί στα βασικά επίπεδα φωτεινότητας που υπάρχουν στην εικόνα. Η μέθοδος υποθέτει επιπλέον ότι μεταξύ δύο γειτονικών λόφων παρεμβάλλεται μια «κοιλιάδα» και θέτει ως στόχο τον εντοπισμό του σημείου κάθε κοιλιάδας που διαχωρίζει με το βέλτιστο τρόπο τους παρακείμενους λόφους.

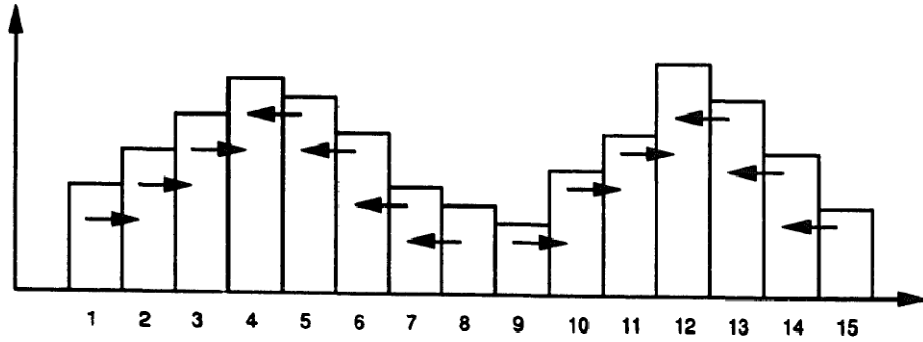
Εάν συμβολίσουμε με p_i την τιμή του ιστογράμματος στο κελί i , τότε μπορούμε να διατυπώσουμε τη βασική ιδέα της μεθόδου ως εξής:

όρισε ένα βέλος από το κελί i προς το γειτονικό κελί με τη μεγαλύτερη τιμή, εφόσον αυτή η τιμή είναι μεγαλύτερη από p_i , αλλιώς μην κάνεις τίποτα.

Με βάση το προκύπτον διάγραμμα, οι κορυφές των λόφων εντοπίζονται σε κελιά στα οποία καταλήγουν αντικρουόμενα βέλη ενώ τα βαθύτερα σημεία των κοιλιάδων (τα αναζητούμενα κατώφλια) εντοπίζονται σε γειτονικά κελιά από τα οποία ξεκινούν βέλη με αντίθετη κατεύθυνση. Στο Σχήμα 2 παρουσιάζεται το διάγραμμα που προκύπτει από την εφαρμογή της μεθόδου σε ένα απλοποιημένο παράδειγμα ιστογράμματος με δύο λόφους. Στη συγκεκριμένη περίπτωση, και με βάση τα όσα είπαμε παραπάνω, οι κορυφές των λόφων εντοπίζονται στα κελιά 4 και 12 αντίστοιχα, ενώ το βαθύτερο σημείο της κοιλιάδας βρίσκεται ανάμεσα στα κελιά 8 και 9.

Για να χρησιμοποιήσει κανείς τη συγκεκριμένη μέθοδο θα πρέπει εκ των προτέρων να καθορίσει το επιθυμητό (ή αναμενόμενο) πλήθος κορυφών. Σε περίπτωση που η μέθοδος αναγνωρίσει μεγαλύτερο πλήθος κορυφών από αυτό που έχει καθοριστεί, το αρχικό ιστόγραμμα εξωμαλύνεται διπλασιάζοντας το μέγεθος του κελιού ή αλλιώς υποδιπλασιάζοντας το πλήθος των κελιών. Συγκεκριμένα τιμή p'_i στο νέο ιστόγραμμα προκύπτει από την ακόλουθη σχέση

$$p'_i = \frac{P_{2i} + P_{2i+1}}{2}, \quad i = 0, 1, \dots$$



Σχήμα 3. Εφαρμογή της μεθόδου hill clustering σε ένα ιστόγραμμα με δύο λόφους

Η διαδικασία αυτή επαναλαμβάνεται μέχρι το πλήθος των κορυφών που αναγνωρίζεται από τον αλγόριθμο να είναι μικρότερο ή ίσο με την επιθυμητή τιμή.

Εάν υποθέσουμε ότι το κατώτατο σημείο της κοιλάδας έχει εντοπιστεί ανάμεσα στα κελιά a και b , τότε ως τιμή κατωφλίου λαμβάνεται η $a_i + b_r / 2$, όπου a_i είναι το αριστερό άκρο του διαστήματος τιμών που αντιπροσωπεύει το κελί a και b_r είναι το δεξί άκρο του διαστήματος που αντιπροσωπεύει το κελί b . Στην περίπτωση που μέσω της διαδικασίας εξομάλυνσης το εύρος του κελιού έχει γίνει πολύ μεγάλο τότε, η παραπάνω τιμή κατωφλίου θα είναι προσεγγιστική, οπότε μια καλύτερη εκτίμηση μπορεί να ληφθεί εάν εφαρμοστεί εξαρχής ολόκληρη η διαδικασία του hill clustering, αλλά μόνο για τα επίπεδα φωτεινότητας μεταξύ a_i και b_r .

4. Μέθοδος Διατήρησης των Ροπών (Moments Preservation Method)

Με βάση αυτή τη μέθοδο η τιμή κατωφλίου υπολογίζεται από την απαίτησή μας οι ροπές της εικόνας πριν και μετά την κατωφλίωση να διατηρούνται ίδιες. Για τον σκοπό αυτό ας υποθέσουμε ότι είναι γνωστές οι ροπές (moments) m_i , $i=1, 2, 3, \dots$ της προς κατωφλίωση *μονοχρωματικής* εικόνας (η εφαρμογή της μεθόδου στην περίπτωση *ψηφιακής* εικόνας, θα δωθεί παρακάτω). Αν g_a, g_φ τα επίπεδα φωτεινότητας του αντικειμένου και του φόντου αντίστοιχα στη εικόνα που προκύπτει μετά την κατωφλίωση, και p_a, p_φ οι αντίστοιχες συχνότητες εμφάνισής τους, τότε για την ικανοποίηση των απαιτήσεών μας θα πρέπει να ισχύουν οι ακόλουθες εξισώσεις:

$$\begin{aligned} p_a g_a + p_\varphi g_\varphi &= m_1 \\ p_a g_a^2 + p_\varphi g_\varphi^2 &= m_2 \\ p_a g_a^3 + p_\varphi g_\varphi^3 &= m_3 \end{aligned} \quad (2)$$

με

$$p_a + p_\varphi = 1. \quad (3)$$

Ερώτημα 1.

Αποδείξτε ότι η λύση του παραπάνω συστήματος των (προφανώς) μη-γραμμικών εξισώσεων, οδηγεί στην ακόλουθη λύση:

$$\begin{aligned} p_a &= \frac{m_1 - g_\varphi}{g_a - g_\varphi} \\ g_a &= \frac{k_0}{2} - \frac{\sqrt{k_0^2 - 4k_1}}{2} \end{aligned} \quad (4)$$

$$g_\varphi = \frac{k_0}{2} + \frac{\sqrt{k_0^2 - 4k_1}}{2},$$

όπου

$$k_0 = \frac{m_3 - m_1 m_2}{m_2 - m_1^2} \quad (5)$$

και

$$k_1 = \frac{m_1 m_3 - m_2^2}{m_2 - m_1^2}. \quad (6)$$

Ερώτημα 2.

Υποθέστε ότι η συνάρτηση π.π. της μονοχρωματικής εικόνας δίνεται από το ακόλουθο μίγμα Γκαουσιανών (Gaussian mixture) συναρτήσεων (δες Σχήμα 2.β)

$$f(z) = p_0 G(z; \mu_0, \sigma_0^2) + p_1 G(z; \mu_1, \sigma_1^2), \quad (7)$$

όπου

$$G(z; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(z-\mu)^2}{2\sigma^2}} \quad (8)$$

και

$$p_0 + p_1 = 1. \quad (9)$$

Για την παραπάνω περίπτωση αποδείξτε ότι οι ροπές m_i , $i=1, 2, 3$ της τ.μ. z δίνονται από τις ακόλουθες σχέσεις:

$$\begin{aligned} m_1 &= p_0 \mu_0 + p_1 \mu_1 \\ m_2 &= p_0 (\mu_0^2 + \sigma_0^2) + p_1 (\mu_1^2 + \sigma_1^2) \\ m_3 &= p_0 \mu_0 (\mu_0^2 + 3\sigma_0^2) + p_1 \mu_1 (\mu_1^2 + 3\sigma_1^2). \end{aligned} \quad (10)$$

Ερώτημα 3.

Υποθέστε ότι η συνάρτηση π.π. της μονοχρωματικής εικόνας δίνεται από την Σχέση (7) με

$$\sigma_0^2 = \sigma_1^2 = \sigma^2. \quad (11)$$

Σχεδιάστε τη σ.π.π. και υπολογίστε τις ποσότητες που ορίστηκαν στις Σχέσεις (4)-(6). Σχολιάστε τα αποτελέσματά σας. Ποιό είναι το βέλτιστο κατώφλι (χρησιμοποιείστε τη μέθοδο της ποσόστωσης);

Ερώτημα 4.

Υποθέστε ότι η συνάρτηση π.π. της μονοχρωματικής εικόνας δίνεται από τη Σχέση (7) με

$$p_0 = p_1 \quad (12)$$

και

$$\sigma_0^2 = \sigma_1^2 = \sigma^2. \quad (13)$$

Σχεδιάστε το μίγμα Γκαουσιανών συναρτήσεων και υπολογίστε τις ποσότητες που ορίστηκαν στις Σχέσεις (4)-(6). Σχολιάστε τα αποτελέσματά σας. Ποιό είναι το βέλτιστο κατώφλι αυτή τη φορά;

Ερώτημα 5.

Υποθέστε τώρα ότι θέλουμε να εφαρμόσουμε την παραπάνω μέθοδο σε μία ψηφιακή εικόνα. Γράψτε μία συνάρτηση στη Matlab που θα δέχεται ως όρισμα εισόδου την ψηφιακή εικόνα και θα σας επιστρέφει την κατωφλιωμένη εικόνα.

Αν $p_i, i=0, 1, \dots, L-1$ το ιστόγραμμα της ψηφιακής εικόνας, όπως αυτό ορίστηκε στη Σχέση (1), τότε χρησιμοποιείτε την ακόλουθη σχέση

$$m_n = \sum_{i=0}^{L-1} i^n p_i, n = 1, 2, 3 \quad (14)$$

για τον υπολογισμό των αναγκαίων ροπών.

5. Η Κατωφλίωση Εικόνας ως ένα Πρόβλημα Ελέγχου Υποθέσεων

Το πρόβλημα της κατωφλίωσης εικόνας μπορεί να ειπωθεί ως ένα πρόβλημα ελέγχου δυαδικών υποθέσεων. Για το σκοπό αυτό ας υποθέσουμε ότι το επίπεδο φωτεινότητας κάθε εικονοστοιχείου της εικόνας προκύπτει από την πραγματοποίηση μια τυχαίας μεταβλητής, z , η οποία ακολουθεί κατανομή γνωστής συναρτησιακής μορφής $f(z; \theta_0)$ όταν το εικονοστοιχείο ανήκει σε περιοχή αντικειμένου και την κατανομή $f(z; \theta_1)$ όταν το εικονοστοιχείο ανήκει σε περιοχή φόντου. Εάν αυτό το μοντέλο είναι σωστό, τότε το πρόβλημα του χαρακτηρισμού του εικονοστοιχείου (x, y) σαν αντικείμενο ή φόντο (δηλαδή το πρόβλημα της εύρεσης του κατωφλίου t), αποτελεί στην ουσία ένα πρόβλημα ελέγχου των παρακάτω δύο υποθέσεων:

$$\mathcal{H}_0: z \sim f(z; \theta_0)$$

$$\mathcal{H}_1: z \sim f(z; \theta_1)$$

με βάση την πραγματοποίηση της z , $z = f(x, y)$. Υποθέτοντας ότι τόσο οι παράμετροι των δύο κατανομών, όσο και οι εκ των προτέρων πιθανότητες εμφάνισης των υποθέσεων, $P(\mathcal{H}_0)$ και $P(\mathcal{H}_1)$ είναι γνωστές, (τι εκφράζουν αυτές οι πιθανότητες στο συγκεκριμένο πρόβλημα;) τότε, ακολουθώντας τη Μπεϋσιανή προσέγγιση (βλέπε Εργαστηριακή Άσκηση 10), η τιμή του κατωφλίου που ελαχιστοποιεί τη συνολική πιθανότητα σφάλματος θα δίνεται από την επίλυση της παρακάτω εξίσωσης:

$$\mathcal{L}(z) = \frac{P(\mathcal{H}_0)}{P(\mathcal{H}_1)}, \quad (15)$$

όπου $\mathcal{L}(z)$ η συνάρτηση λόγου πιθανοφάνειας (δες Σχέση (19) στην Εργαστηριακή Άσκηση 10).

Ερώτημα 6.

Υποθέστε ότι η συνάρτηση π.π. της μονοχρωματικής εικόνας δίνεται από την Σχέση (7) (είναι δηλαδή ένα μίγμα Γκαουσιανών συναρτήσεων).

Διατυπώστε το πρόβλημα ελέγχου υποθέσεων στην περίπτωση αυτή και αποδείξτε ότι η Σχέση (15) οδηγεί στην επίλυση της παρακάτω εξίσωσης δευτέρου βαθμού:

$$Az^2 + Bz + C = 0, \quad (16)$$

όπου:

$$\begin{aligned}
A &= \sigma_0^2 - \sigma_1^2 \\
B &= 2(\mu_0\sigma_1^2 - \mu_1\sigma_0^2) \\
C &= \mu_1^2\sigma_0^2 - \mu_0^2\sigma_1^2 + 2\sigma_0^2\sigma_1^2 \ln(\sigma_1 P(H_0)/\sigma_0 P(H_1))
\end{aligned} \tag{17}$$

Ερώτημα 7.

Υιοθετώντας τις υποθέσεις των Ερωτήσεων 3 και 4 και χρησιμοποιώντας τις Σχέσεις (17) υπολογίστε τις ποσότητες A , B και C και προσδιορίστε σε κάθε περίπτωση την τιμή του βέλτιστου κατωφλιού. Σχολιάστε τα αποτελέσματά σας.

Σε πραγματικά προβλήματα βεβαίως, ακόμη και εάν η υπόθεσή μας ότι η εικόνα έχει προέλθει από την πραγματοποίηση δύο κανονικών κατανομών είναι ορθή, τόσο οι παράμετροι των κατανομών, όσο και οι δύο εκ των προτέρων πιθανότητες, είναι στη γενική περίπτωση άγνωστες. Η εκτίμηση αυτών των ποσοτήτων μπορεί να γίνει μέσω του ιστογράμματος της εικόνας με μεθόδους προσαρμογής καμπύλης, όμως η προσέγγιση αυτή οδηγεί σε μη γραμμικά προβλήματα, τα οποία είναι πολύ δύσκολο να επιλυθούν. Για αυτόν τον λόγο, στην πράξη χρησιμοποιούνται κυρίως μέθοδοι κατωφλίωσης, που βασίζονται στην πλειοψηφία τους στη μορφή και τις στατιστικές ποσότητες του ιστογράμματος. Τρεις μέθοδοι που ανήκουν στην κατηγορία αυτή περιγράφονται στην συνέχεια.

6. Η Μέθοδος του Otsu

Έχουμε ήδη αναφέρει ότι κάθε υπογήφιο κατώφλι t διαμερίζει τα εικονοστοιχεία της εικόνας σε δύο κλάσεις, $C_0(t)$ και $C_1(t)$, με την πρώτη να περιέχει τα εικονοστοιχεία με επίπεδο φωτεινότητας στο σύνολο $\{0,1,\dots,t\}$ και τη δεύτερη τα εικονοστοιχεία με επίπεδο φωτεινότητας στο σύνολο $\{t+1,t+2,\dots,L-1\}$. Εάν τώρα θεωρήσουμε ότι τα επίπεδα φωτεινότητας των αντικειμένων ακολουθούν την κατανομή F_0 (με σ.π.π. $f_0(z)$) ενώ τα επίπεδα φωτεινότητας του φόντου ακολουθούν την κατανομή F_1 (με σ.π.π. $f_1(z)$) τότε το ιστόγραμμα των εικονοστοιχείων της κλάσης $C_0(t)$ αποτελεί προσέγγιση της $f_0(z)$, ενώ αντίστοιχα, το ιστόγραμμα των εικονοστοιχείων της κλάσης $C_1(t)$ αποτελεί προσέγγιση της $f_1(z)$. Προφανώς η βέλτιστη προσέγγιση θα επιτυγχάνεται για $t=t^*$, όπου t^* το αναζητούμενο βέλτιστο κατώφλι. Εάν τώρα συμβολίσουμε με $p_i^0(t)$, $i=0,1,\dots,t$ το ιστόγραμμα των εικονοστοιχείων της κλάσης $C_0(t)$ και με $p_i^1(t)$, $i=t+1,\dots,L-1$ το ιστόγραμμα των εικονοστοιχείων της κλάσης $C_1(t)$, τότε είναι εύκολο να δει κανείς ότι:

$$p_i^0(t) = \frac{P_i}{\omega_0(t)}, \quad i=0,1,\dots,t \tag{18}$$

και

$$p_i^1(t) = \frac{P_i}{\omega_1(t)}, \quad i=t+1,\dots,L-1, \tag{19}$$

όπου $p_i, i=0, 1, \dots, L-1$ το ιστόγραμμα της ψηφιακής εικόνας όπως αυτό ορίστηκε στη Σχέση (1), ενώ $\omega_0(t)$ και $\omega_1(t)$ είναι οι πιθανότητες ένα εικονοστοιχείο να ανήκει στις κλάσεις $C_0(t)$ και $C_1(t)$ αντίστοιχα, και ορίζονται από τις παρακάτω σχέσεις:

$$\omega_0(t) = \sum_{i=0}^t p_i$$

και

$$\omega_1(t) = \sum_{i=t+1}^{L-1} p_i, \quad (20)$$

και προφανώς, $\omega_0(t) + \omega_1(t) = 1$. Με άλλα λόγια, το ιστόγραμμα των εικονοστοιχείων της κλάσης $C_0(t)$ ισούται με το τμήμα του ολικού ιστογράμματος της εικόνας που βρίσκεται *αριστερά* από το κατώφλι t , κανονικοποιημένο με την πιθανότητα $\omega_0(t)$, ενώ το ιστόγραμμα των εικονοστοιχείων της κλάσης $C_1(t)$ ισούται με το τμήμα του ολικού ιστογράμματος της εικόνας που βρίσκεται *δεξιά* από το κατώφλι, κανονικοποιημένο με την πιθανότητα $\omega_1(t)$.

Ερώτημα 8:

Με βάση τον ορισμό του ιστογράμματος που δίνεται στη Σχέση (1), καθώς και τον παραπάνω ορισμό των κλάσεων $C_0(t)$ και $C_1(t)$, αποδείξτε τις Σχέσεις (18) και (19).

Χωρίς να κάνει κάποια υπόθεση για τη συναρτησιακή μορφή των $f_0(z)$ και $f_1(z)$ (για παράδειγμα εάν είναι κανονικές ή όχι), η μέθοδος του Otsu βασίζεται στις στατιστικές ιδιότητες των δύο κλάσεων που σχηματίζονται από κάθε υποψήφιο κατώφλι και προσπαθεί να εντοπίσει το κατώφλι που οδηγεί στο σχηματισμό των κλάσεων με τη μεγαλύτερη μεταξύ τους διαχωριστικότητα (separation).

Συγκεκριμένα, η μέση τιμή και η διασπορά των επιπέδων φωτεινότητας σε κάθε κλάση ξεχωριστά θα δίνονται από τις ακόλουθες σχέσεις:

$$\mu_0(t) = \sum_{i=0}^t i p_i^0(t) = \frac{\sum_{i=0}^t i p_i}{\omega_0(t)} \quad (21)$$

$$\mu_1(t) = \sum_{i=t+1}^{L-1} i p_i^1(t) = \frac{\sum_{i=t+1}^{L-1} i p_i}{\omega_1(t)} \quad (22)$$

$$\sigma_0^2(t) = \sum_{i=0}^t (i - \mu_0(t))^2 p_i^0(t) = \frac{\sum_{i=0}^t (i - \mu_0(t))^2 p_i}{\omega_0(t)} \quad (23)$$

$$\sigma_1^2(t) = \sum_{i=t+1}^{L-1} (i - \mu_1(t))^2 p_i^1(t) = \frac{\sum_{i=t+1}^{L-1} (i - \mu_1(t))^2 p_i}{\omega_1(t)}, \quad (24)$$

ενώ η ολική μέση τιμή και διασπορά των επιπέδων φωτεινότητας της εικόνας θα δίνονται από τις σχέσεις:

$$\mu_T = \sum_{i=0}^{L-1} i p_i = \omega_0(t) \mu_0(t) + \omega_1(t) \mu_1(t) \quad (25)$$

$$\sigma_T^2 = \sum_{i=0}^{L-1} (i - \mu_T)^2 p_i . \quad (26)$$

Για την ολική διασπορά σ_T^2 έχουμε διαδοχικά:

$$\begin{aligned} \sigma_T^2 &= \sum_{i=0}^{L-1} (i - \mu_T)^2 p_i \\ &= \sum_{i=0}^{L-1} i^2 p_i + \mu_T^2 \sum_{i=0}^{L-1} p_i - 2\mu_T \sum_{i=0}^{L-1} i p_i = \sum_{i=0}^{L-1} i^2 p_i - \mu_T^2 \\ &= \sum_{i=0}^t i^2 p_i + \sum_{i=t+1}^{L-1} i^2 p_i - \mu_T^2 \\ &= \sum_{i=0}^t (i - \mu_0(t) + \mu_0(t))^2 p_i + \sum_{i=t+1}^{L-1} (i - \mu_1(t) + \mu_1(t))^2 p_i - \mu_T^2 \\ &= \sum_{i=0}^t (i - \mu_0(t))^2 p_i + \mu_0^2(t) \sum_{i=0}^t p_i + 2\mu_0(t) \underbrace{\sum_{i=0}^t (i - \mu_0(t)) p_i}_0 + \\ &\quad + \sum_{i=t+1}^{L-1} (i - \mu_1(t))^2 p_i + \mu_1^2(t) \sum_{i=t+1}^{L-1} p_i + 2\mu_1(t) \underbrace{\sum_{i=t+1}^{L-1} (i - \mu_1(t)) p_i}_0 - \mu_T^2 \\ &= \omega_0(t) \sigma_0^2(t) + \omega_0(t) \mu_0^2(t) + \omega_1(t) \sigma_1^2(t) + \omega_1(t) \mu_1^2(t) - \mu_T^2 \\ &= \omega_0(t) \sigma_0^2(t) + \omega_1(t) \sigma_1^2(t) + \omega_0(t) \mu_0^2(t) + \omega_1(t) \mu_1^2(t) - (\omega_0(t) + \omega_1(t)) \mu_T^2 \\ &= \omega_0(t) \sigma_0^2(t) + \omega_1(t) \sigma_1^2(t) + \omega_0(t) (\mu_0^2(t) - \mu_T^2) + \omega_1(t) (\mu_1^2(t) - \mu_T^2) \end{aligned} \quad (27)$$

Ορίζοντας τώρα ως *εσωτερική στις κλάσεις* διασπορά (within - class variance) την ποσότητα:

$$\sigma_w^2(t) = \omega_0(t) \sigma_0^2(t) + \omega_1(t) \sigma_1^2(t) \quad (28)$$

και ως *μεταξύ των κλάσεων* διασπορά (between - class variance) την ποσότητα:

$$\begin{aligned} \sigma_B^2(t) &= \omega_0(t) (\mu_0^2(t) - \mu_T^2) + \omega_1(t) (\mu_1^2(t) - \mu_T^2) \\ &= \omega_0(t) (\mu_0(t) - \mu_T)^2 + \omega_1(t) (\mu_1(t) - \mu_T)^2 \\ &= \omega_0(t) \omega_1(t) (\mu_0(t) - \mu_1(t))^2 . \end{aligned} \quad (29)$$

Από τη Σχέση (27) τώρα έχουμε:

$$\frac{\sigma_W^2(t)}{\sigma_T^2} + \frac{\sigma_B^2(t)}{\sigma_T^2} = 1 \quad (30)$$

Λαμβάνοντας την τιμή της $\sigma_B^2(t)$ σαν ένα μέτρο της διαχωριστικότητας μεταξύ των δύο κλάσεων, ο Otsu πρότεινε ο εντοπισμός του βέλτιστου κατωφλίου να γίνεται μέσω της μεγιστοποίησης της ακόλουθης συναρτήσης:

$$\eta(t) = \frac{\sigma_B^2(t)}{\sigma_T^2}, \quad (31)$$

δηλαδή, η βέλτιστη τιμή του κατωφλίου προκύπτει από την λύση του ακόλουθου προβλήματος βελτιστοποίησης:

$$t^* = \arg \max_t (\sigma_B^2(t)). \quad (32)$$

Ερώτημα 9:

Αποδείξτε ότι η λύση του προβλήματος βελτιστοποίησης της Σχέσης (32) μπορεί να προκύψει και από την μεγιστοποίηση των ακόλουθων συναρτήσεων κόστους:

$$\lambda(t) = \frac{\sigma_B^2(t)}{\sigma_W^2(t)}, \quad (33)$$

και

$$\kappa(t) = \frac{\sigma_T^2}{\sigma_W^2(t)}. \quad (34)$$

Με άλλα λόγια αποδείξτε ότι οι παραπάνω συναρτήσεις κόστους είναι ισοδύναμες.

Ερώτημα 10:

Γιατί επιλέγουμε να μεγιστοποιήσουμε την συνάρτηση κόστους της Σχέσης (31) για τον προσδιορισμό της βέλτιστης τιμής του κατωφλίου;

Ερώτημα 11:

Υλοποιείτε στην Matlab την μέθοδο κατωφλίωσης εικόνας του Otsu.

Κλείνοντας την ενότητα αυτή θα πρέπει να σημειώσουμε ότι η μέθοδος του Otsu είναι μια από τις πιο αποτελεσματικές και περισσότερο χρησιμοποιούμενες στην πράξη μεθόδους κατωφλίωσης.

7. Η Μέθοδος των Kapur, Sahoo και Wong

Η συγκεκριμένη μέθοδος κινείται στην ίδια κατεύθυνση με την προηγούμενη, αλλά προσδιορίζει το βέλτιστο κατώφλι βασιζόμενη στην έννοια της *εντροπίας* των δύο κλάσεων $C_0(t)$ και $C_1(t)$, η οποία ορίζεται ως εξής:

$$H_0(t) = -\sum_{i=0}^t p_i^0(t) \ln p_i^0(t) \quad (35)$$

$$H_1(t) = -\sum_{i=t+1}^{L-1} p_i^1(t) \ln p_i^1(t), \quad (36)$$

όπου τα $p_i^0(t)$ και $p_i^1(t)$ ορίζονται στις Σχέσεις (18) και (19), αντίστοιχα. Επομένως, κατά τα γνωστά το βέλτιστο κατώφλι προκύπτει τότε από την επίλυση του παρακάτω προβλήματος μεγιστοποίησης:

$$t^* = \arg \max_t (H_0(t) + H_1(t)). \quad (37)$$

8. Κατωφλίωση με τη μέθοδο k -means

Η μέθοδος k -means αποτελεί τον πιο γνωστό εκπρόσωπο της οικογένειας των μεθόδων κατηγοριοποίησης με χρήση τεχνικών βελτιστοποίησης. Στόχος των μεθόδων αυτών είναι κατηγοριοποίηση ενός συνόλου αντικειμένων σε ένα προκαθορισμένο αριθμό ομάδων μέσω της μεγιστοποίησης ή της ελαχιστοποίησης μιας συνάρτησης κόστους. Η χρήση των μεθόδων αυτών προϋποθέτει την ύπαρξη εκ των προτέρων εκτίμησης για το αναμενόμενο πλήθος των διαφορετικών ομάδων του δοσμένου συνόλου.

Συγκεκριμένα ο k -means κατηγοριοποιεί ένα σύνολο N αντικειμένων σε k ομάδες, G_1, \dots, G_k με N_1, \dots, N_k αντικείμενα αντίστοιχα, όπου $\sum_{i=1}^k N_i = N$, μέσω της ελαχιστοποίησης της παρακάτω συνάρτησης:

$$J = \sum_{i=1}^k \sum_{n=1}^{N_i} \|x_{i,n} - c_i\|^2, \quad (38)$$

όπου $x_{i,n}$ είναι το n -οστό μέλος τις ομάδας G_i και $c_i = \frac{1}{N_i} \sum_{n=1}^{N_i} x_{i,n}$ είναι ο αριθμητικός μέσος της.

Η προφανής λύση του παραπάνω προβλήματος ελαχιστοποίησης, δηλαδή η εύρεση όλων των δυνατών κατηγοριοποιήσεων και η επιλογή εκείνης που ελαχιστοποιεί την συνάρτηση κόστους J που ορίστηκε στη Σχέση (38) είναι πρακτικά αδύνατη, αφού ο συνολικός αριθμός των κατηγοριοποιήσεων N αντικειμένων σε k ομάδες είναι:

$$C(N, k) = \frac{1}{k!} \sum_{m=1}^k (-1)^{k-m} \binom{k}{m} m^N. \quad (39)$$

Η τιμή αυτή είναι απαγορευτικά μεγάλη ακόμα και για πολύ μικρές τιμές των N και k , για παράδειγμα $C(100, 5) \approx 6.6 \times 10^{67}$. Για αυτόν το λόγο η λύση επιτυγχάνεται με χρήση του παρακάτω επαναληπτικού αλγορίθμου, ο οποίος ανήκει στους λεγόμενους αλγορίθμους *ανάβασης λόφου* (hill climbing algorithms):

1. Διάλεξε τυχαία k αντικείμενα από το δοσμένο σύνολο. Τα αντικείμενα αυτά αποτελούν την αρχική επιλογή των k αριθμητικών μέσων c_1, \dots, c_k .
2. Κατέταξε το αντικείμενο x_j , $j = 1, \dots, N$ στην ομάδα G_i , $i = 1, \dots, k$, όπου:

$$i = \arg \min_{n \in \{1, \dots, k\}} \|x - c_n\|^2.$$

3. Υπολόγισε τις νέες τιμές των αριθμητικών μέσων c_i των k ομάδων.
4. Επανάλαβε τα βήματα 2 και 3 έως ότου καμία μετακίνηση αντικειμένου να μην λαμβάνει χώρα (η σύνθεση των ομάδων παραμένει αμετάκλητη).

Στην κατωφλίωση εικόνας, η μέθοδος k -means μπορεί να χρησιμοποιηθεί κατά τέτοιο τρόπο ώστε να χωριστούν τα εικονοστοιχεία της αρχικής εικόνας σε k ομάδες διαφορετικού επιπέδου φωτεινότητας και έπειτα να αποδοθεί στα εικονοστοιχεία κάθε ομάδας μια προκαθορισμένη τιμή ώστε να προκύψει η κατωφλιωμένη εικόνα. Για παράδειγμα, στην περίπτωση εικόνων κειμένου, μπορούμε να χωρίσουμε τα εικονοστοιχεία σε «φωτεινά» και «σκοτεινά» και έπειτα να απεικονίσουμε στην κατωφλιωμένη εικόνα τα εικονοστοιχεία της πρώτης ομάδας με λευκό χρώμα και εκείνα της δεύτερης με μαύρο.

Ερώτημα 12.

12.α: Φορτώστε στη MATLAB την εικόνα text.jpg που σας δίνεται χρησιμοποιώντας τη συνάρτηση imread(). Υπολογίστε και σχεδιάστε το ιστόγραμμα της χρησιμοποιώντας τη συνάρτηση imhist(). Τι παρατηρείτε;

12.β: Προσθέστε στην εικόνα λευκό θόρυβο Gauss με μέση τιμή μηδέν και διασπορά σ_0^2 για τα εικονοστοιχεία των γραμμάτων και σ_1^2 για τα εικονοστοιχεία του φόντου και για 5 τουλάχιστον διαφορετικούς συνδυασμούς τιμών για τα σ_0^2 και σ_1^2 :

1. Κατωφλιώστε την ενθόρυβη εικόνα χρησιμοποιώντας το βέλτιστο θεωρητικό κατώφλι σύμφωνα με τη μέθοδο που περιγράφεται στην Ενότητα 5. Ποιές είναι οι εκ των προτέρων πιθανότητες των υποθέσεων και πώς τις υπολογίσατε;
2. Κατωφλιώστε την ενθόρυβη εικόνα χρησιμοποιώντας τις μεθόδους των Ενοτήτων 2 – 8 (εκτός της 5).
3. Αξιολογήστε τις μεθόδους των Ενοτήτων 2 – 8 (εκτός της 5) συγκρίνοντας την τιμή κατωφλίου που επιστρέφουν με τη βέλτιστη τιμή του υποερωτήματος 12.β.1.
4. Συγκρίνετε τα αποτελέσματα της κατωφλίωσης των χρησιμοποιούμενων μεθόδων.

Βιβλιογραφία

- [1] W. Doyle, "Operation useful for similarity – invariant pattern recognition," *J. Assoc. Comput. Mach.*, Vol. 9, pp. 259 – 267, 1962.
- [2] N. Otsu, "A threshold selection method from gray – level histogram," *IEEE Trans. Systems Man. Cybernetics.*, SMC 8, pp. 62-66, 1978.
- [3] J. N. Kapur, P. K. Sahoo and A. K. C. Wong, "A new method for gray –level picture thresholding using the entropy of the histogram," *Comput. Vision Graphics Image Process*, Vol. 29, pp. 273-285, 1985.
- [4] W. Tsai, "Moment – preserving thresholding : A new approach," *Comput. Vision Graphics Image Process.*, Vol. 29, pp. 377-393, 1985.
- [5] R. C. Gonzalez & R. E. Woods, "Digital Image Processing," *Addison Wesley*, 1991.
- [6] D. M. Tsai and Y. H. Chen, "A fast histogram-clustering approach for multi-level thresholding," *Pattern Recognition Letters*, Vol. 13, pp. 245-252, 1992.
- [7] B. S. Everitt, S. Landau, M. Leese, "Cluster Analysis," *Arnold*, London, 2001.