

Τμηματοποίηση με χρήση μοντέλου Markov ή Gibbs

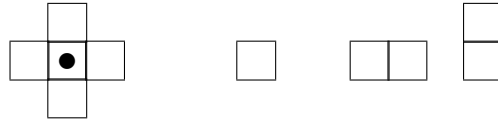
Γεώργιος Τζιρίτας
Τμήμα Επιστήμης Υπολογιστών
Πανεπιστήμιο Κρήτης

Μάρτιος 2019

Η ανάλυση με βάση ένα μοντέλο Markov για το πεδίο των αποφάσεων κατάταξης των σημείων σε κλάσεις, χρησιμοποιεί τις αλληλεξαρτήσεις που υπάρχουν τοπικά σε γειτονικά σημεία. Για κάθε μοντέλο Markov που δίνει τις τοπικές αλληλεξαρτήσεις ορίζεται ισοδύναμα μία κατανομή Gibbs που δίνει συνολικά όλες τις εξαρτήσεις του πεδίου των αποφάσεων. Ο ορισμός της κατανομής Gibbs δίδεται στη συνέχεια, αφού προηγουμένως δοθούν οι σχέσεις γειτνίασης που είναι αναγκαίες για την περιγραφή των εξαρτήσεων του πεδίου. Μια γενική επισκόπηση των τυχαίων πεδίων Markov δίδεται στο [8].

1 Πεδίο Gibbs και τμηματοποίηση

Ας είναι $L = \{(m, n) : 0 \leq m < M, 0 \leq n < N\}$ το σύνολο των σημείων της εικόνας. Η γειτονιά ενός σημείου (m, n) ορίζεται ως ένα υποσύνολο $\gamma(m, n)$ του L , που δεν περιλαμβάνει το σημείο (m, n) , με τον όρο ότι, εάν $(k, l) \in \gamma(m, n)$, τότε $(m, n) \in \gamma(k, l)$. Το σύστημα γειτνίασης στο L ορίζεται ως το σύνολο των γειτονιών όλων των σημείων του L . Με τη βοήθεια του ορισμού της γειτονιάς ορίζεται η παρέα, c , ως ένα σύνολο, είτε ενός μόνο σημείου, είτε περισσότερων του ενός σημείων, αρκεί για κάθε σημείο μιας παρέας τα υπόλοιπα σημεία της παρέας να ανήκουν στη γειτονιά του. Στο Σχήμα 1 δίδεται ο ορισμός μιας γειτονιάς 4 σημείων και οι αντίστοιχες σ' αυτή παρέες. Παρόμοια δίδονται οι παρέες για μια γειτονιά 8 σημείων στο Σχήμα 2.

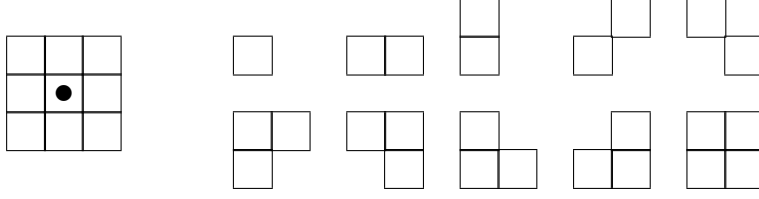


Σχήμα 1: Γειτονιά 4 σημείων και αντίστοιχες παρέες.

Ένα πεδίο Gibbs περιγράφει συνολικά το πεδίο των κατατάξεων σε όλη την εικόνα

$$Pr\{X(m, n) = x(m, n) : (m, n) \in L\} = \frac{e^{-U(x)}}{\Psi}, \quad (1)$$

όπου $U(x)$ ονομάζεται συνάρτηση ενέργειας, και είναι συνάρτηση MN μεταβλητών, όσα δηλαδή είναι τα σημεία της εικόνας. Ψ είναι μία σταθερά, $\Psi = \sum e^{-U(x)}$. Ιδιαίτερα ενδιαφέρουσα είναι η περίπτωση ενός πεδίου Gibbs, που είναι ισοδύναμο με πεδίο Markov, οπότε η συνάρτηση



Σχήμα 2: Γειτονιά 8 σημείων και αντίστοιχες παρέες.

ενέργειας ορίζεται με τη βοήθεια του συνόλου των παρεών, C , των σημείων της εικόνας

$$U(x) = \sum_{c \in C} \phi_c(x). \quad (2)$$

Το $\phi_c(x)$ ονομάζεται δυναμικό της παρέας c . Συχνά για τη γειτονιά 8 σημείων μόνο ζεύγη σημείων θεωρούνται ότι έχουν μη μηδενικό δυναμικό. Το δυναμικό ορίζεται ώστε οι παρέες με διαφορετικές κατατάξεις να είναι λιγότερο πιθανές από αυτές με ίδιες κατατάξεις. Ένα παράδειγμα δυναμικού για διμελείς παρέες που μπορεί να χρησιμοποιηθεί για την τμηματοποίηση των εικόνων ακολουθεί

$$\phi_c(x) = \begin{cases} 0 & \text{εάν τα δύο σημεία ανήκουν στην ίδια κλάση} \\ \zeta > 0 & \text{διαφορετικά} \end{cases} \quad (3)$$

Πρόκειται για το μοντέλο Ising, αν οι κλάσεις είναι δύο. Αν οι κλάσεις είναι περισσότερες και τα δυναμικά παίρνουν σταθερές τιμές για δοσμένο ζεύγος κλάσεων πρόκειται για το μοντέλο Potts.

Ας θεωρήσουμε τώρα το πρόβλημα της τμηματοποίησης της εικόνας κατατάσσοντας τα σημεία της σε K κλάσεις. Υποθέτουμε ότι η παρατηρούμενη εικόνα για κάθε σημείο εξαρτάται μόνο από την κλάση που ανήκει αυτό το σημείο, με μία παρενόχληση από αθροιστικό θόρυβο, που παρακάτω θα θεωρήσουμε ότι ακολουθεί την κανονική κατανομή με διασπορά σ^2

$$p(y(m, n) | X(k, l) = x(k, l), (k, l) \in L) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(y(m, n) - f(x(m, n)))^2}{2\sigma^2}}. \quad (4)$$

Η συνάρτηση $f(\cdot)$ δίδει τη μέση αντιπροσωπευτική τιμή για κάθε κλάση S_i

$$f(x(m, n)) = \mu_i, \quad \text{εάν } (m, n) \in S_i, \quad 0 \leq i \leq K - 1. \quad (5)$$

Η τμηματοποίηση της εικόνας μπορεί να γίνει μεγιστοποιώντας την ύστερη πιθανότητα του πεδίου κατάταξης των σημείων σε κλάσεις

$$\begin{aligned} & Pr\{X(m, n) = x(m, n) | Y(m, n) = y(m, n); (m, n) \in L\} \\ & \propto Pr\{X(m, n) = x(m, n) : (m, n) \in L\} \prod_{(m, n) \in L} p(y(m, n) | X(m, n) = x(m, n)). \end{aligned} \quad (6)$$

Αυτό ισοδυναμεί με την ελαχιστοποίηση της ακόλουθης ποσότητας ενέργειας

$$E(X) = \sum_{(m, n) \in L} \phi_c(x(m, n)) + \frac{1}{2\sigma^2} \sum_{i=0}^{K-1} \sum_{(m, n) \in S_i} (y(m, n) - \mu_i)^2. \quad (7)$$

Ωστόσο το κόστος εύρεσης της βέλτιστης λύσης με διεξοδική αναζήτηση είναι απολύτως απαγορευτικό, γιατί το πλήθος των δυνατών καταστάσεων είναι K^{MN} , ενώ υπάρχουν πολλά τοπικά

ελάχιστα. Η ελαχιστοποίηση μπορεί να επιτευχθεί με μία πιθανοκρατική αναζήτηση του ελάχιστου προσομοιώνοντας το πεδίο Gibbs [4]. Ωστόσο και σε αυτή την περίπτωση το υπολογιστικό κόστος είναι εξαιρετικά υψηλό για πολλές εφαρμογές.

Μια δυνατότητα γρήγορης εύρεσης ενός τοπικού ελαχίστου, εξαρτώμενου από την αρχικοποίηση του πεδίου, δίδεται από επαναληπτικό αλγόριθμο, που υπολογίζει σε κάθε βήμα το βέλτιστο τοπικά πεδίο με δοσμένη μια προηγούμενη εκτίμηση.

2 Μαρκοβιανό πεδίο υπό τοπικές συνθήκες

Στο προηγούμενο μοντέλο το δυναμικό ενός ζεύγους εικονοστοιχείων είναι ανεξάρτητο από τα δεδομένα. Ωστόσο βάσιμα μπορεί να υποτεθεί ότι η πιθανότητα διαφορετικής κλάσης σε γειτονικά εικονοστοιχεία είναι μεγαλύτερη όταν η διαφορά των χαρακτηριστικών είναι μεγαλύτερη. Βέβαια δεν είναι υποχρεωτικό ορισμένες σημαντικές τοπικές διαφορές να ευρίσκονται επί των συνόρων των τμημάτων.

Εφόσον ένα τέτοιο μοντέλο είναι αντιπροσωπευτικό, μπορεί να θεωρηθεί ότι η μαρκοβιανή ιδιότητα ισχύει για τις δεσμευμένες πιθανότητες των κατατάξεων ως προς τα δεδομένα. Τότε το δυναμικό για μια διμελή παρέα με διαφορετικές κλάσεις μπορεί να ορισθεί ανάλογο της ομοιότητας με βάση τα δεδομένα, όπως

$$a((m, n), (k, l)) = \zeta \exp(-\beta(y(m, n) - y(k, l))^2) + \lambda,$$

με την παράμετρο β να ρυθμίζει το ρόλο των τοπικών αντιθέσεων στα δεδομένα.

3 Αλγόριθμος επαναλαμβανόμενων τοπικά επικρατούσων τιμών

Η μέθοδος των επαναλαμβανόμενων επικρατούσων τιμών [2] είναι μια 'άπληστη' τεχνική που οδηγεί σε τοπικό ελάχιστο και μπορεί να χρησιμοποιηθεί για οποιονδήποτε αριθμό κλάσεων K . Η μέθοδος απαιτεί αρχικοποίηση που είναι καθοριστική για το τελικό αποτέλεσμα. Από μία αρχική κατάταξη, που συχνά είναι αυτή που προκύπτει από απλή ανάλυση του ιστογράμματος, γίνεται μία επανεκτίμηση, ώστε να μεγιστοποιείται η ακόλουθη πιθανότητα

$$Pr\{X(m, n) = x(m, n) | Y(m, n) = y(m, n), \hat{x}(k, l); (k, l) \in L - \{(m, n)\}\}.$$

Το αποτέλεσμα της μεγιστοποίησης είναι η επικρατούσα τιμή του πεδίου για δοσμένη παρατήρηση, και με δοσμένη την αρχική εκτίμηση. Αποδεικνύεται ότι αρκεί να ελαχιστοποιηθεί η ακόλουθη ποσότητα σε κάθε σημείο

$$\frac{1}{2\sigma^2}(y(m, n) - \mu_i)^2 + \sum \phi_c(x(m, n)),$$

όπου αθροίζονται μόνο τα δυναμικά των παρεών που περιλαμβάνουν το σημείο (m, n) , κι όπου οι τιμές του πεδίου για τα γειτονικά σημεία είναι αυτές της προηγούμενης εκτίμησης. Αν επομένως από τα 8 γειτονικά σημεία, κατά την προηγούμενη εκτίμηση κατατάχθηκαν $\alpha_i(m, n)$ στην κλάση i , τότε η τρέχουσα κατάταξη βασίζεται στην τοπική ελαχιστοποίηση του κριτηρίου

$$(y(m, n) - \mu_i)^2 - \sigma^2 \zeta \alpha_i(m, n).$$

Όταν ευρεθούν οι επικρατούσες τιμές σε όλα τα σημεία του πεδίου η διαδικασία επαναλαμβάνεται με τον ίδιο ακριβώς τρόπο, και σταματά όταν μειωθεί σημαντικά ο αριθμός των σημείων που από μία επανάληψη στην επόμενη αλλάζουν κλάση, ή ιδανικά όταν ουδένα σημείο αλλάζει κλάση. Ο αλγόριθμος επαναλαμβανόμενων τοπικά επικρατούσων τιμών δίδεται κατωτέρω σε ψευδοκώδικα.

Initialization:

```
for each pixel p
{
    current_label(p) := init_label(p)
    next_state(p) := active
}
change_pixels := number_of_pixels
```

Iterations:

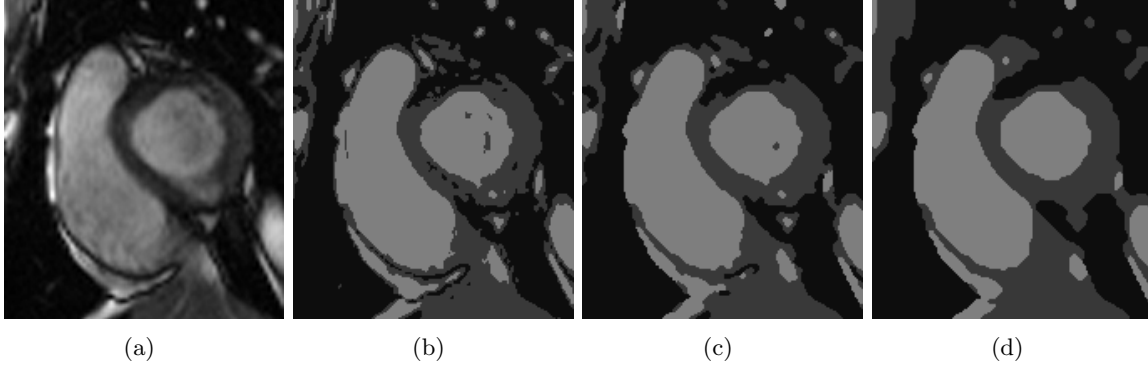
```
while (change_pixels != 0)
{
    change_pixels := 0
    for each pixel p
    {
        previous_label(p) := current_label(p)
        current_state(p) := next_state(p)
        next_state(p) := inactive
    }
    for each pixel p
    {
        if (current_state(p) == active)
        {
            current_label(p) := find_optimal_label(p)
            if (current_label(p) != previous_label(p))
            {
                change_pixels++
                next_state(neighbour(p)) := active
            }
        }
    }
}
```

Στο Σχήμα 3 δίδονται μια εικόνα καρδιάς, η αρχική κατάταξη με βάση τις πιθανοφάνειες των τριών κλάσεων, το αποτέλεσμα της τμηματοποίησης με τη μέθοδο των επαναλαμβανόμενων επικρατούσων τιμών και το αποτέλεσμα με μέθοδο που βασίζεται σε τομές γράφων, που περιγράφεται παρακάτω.

4 Τομές γράφων

Υπάρχει η δυνατότητα να σχεδιασθούν γρήγοροι αλγόριθμοι στην περίπτωση που το πεδίο Gibbs μπορεί να αναπαρασταθεί με ένα γράφο. Στην περίπτωση που οι παρέες του πεδίου είναι διμελείς, υπό όχι αυστηρές συνθήκες, το πεδίο αναπαρίσταται με γράφο και το πρόβλημα της κατάταξης ανάγεται σε πρόβλημα διαμέρισης του γράφου.

Στην περίπτωση όπου $K = 2$ υπάρχει αλγόριθμος που ευρίσκει τη βέλτιστη κατάταξη σε πολυωνυμικό χρόνο. Θεωρούμε ότι το πεδίο $X(m, n)$ λαμβάνει τιμές 0 με $f(0) = \mu_0$ και 1 με $f(1) = \mu_1$. Για το μοντέλο με τις κανονικές κατανομές η ποσότητα ενέργειας του πεδίου



Σχήμα 3: Τμηματοποίηση με τον αλγόριθμο επαναλαμβανόμενων επικρατούσων τιμών και με τομές γράφου.

γράφεται

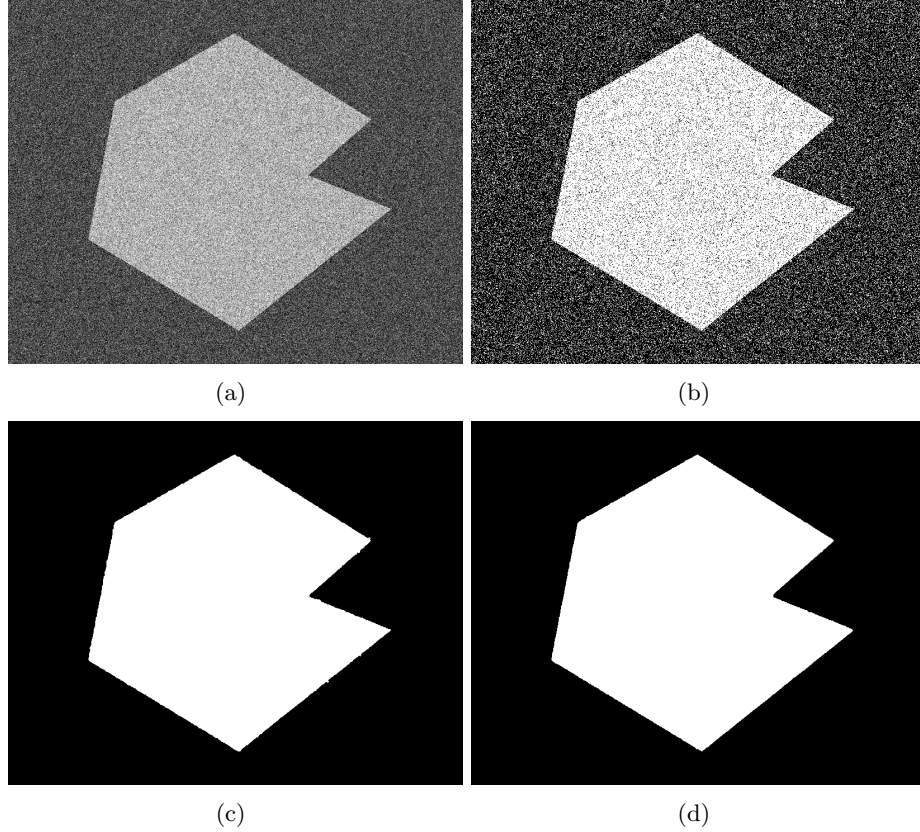
$$\begin{aligned}
 E(X) = & \frac{1}{\sigma^2} \sum_{(m,n) \in L} x(m,n)(y(m,n) - \mu_1)^2 + \frac{1}{\sigma^2} \sum_{(m,n) \in L} (1 - x(m,n))(y(m,n) - \mu_0)^2 \\
 & + \zeta \sum_{\substack{(m,n) \in L \\ (k,l) \in \gamma(m,n)}} (x(m,n) - x(k,l))^2.
 \end{aligned} \tag{8}$$

Επομένως το ανωτέρω πρόβλημα ελαχίστου ταυτίζεται με αυτό της εύρεσης της ελάχιστης τομής σε διμερή γράφο, και ισοδύναμα της επίλυσης του προβλήματος της μέγιστης ροής [5]. Συγκεκριμένα μπορούμε να θεωρήσουμε ένα γράφο με κόμβους τα σημεία της εικόνας και επιπλέον ένα κόμβο ‘πηγή’ (s) και ένα κόμβο ‘δεξαμενή’ (t). Τα σύνολα $C_1 = \{s\} \cup \{(m,n) : x(m,n) = 1\}$ και $C_0 = \{t\} \cup \{(m,n) : x(m,n) = 0\}$ ορίζουν μια διαμέριση του συνόλου των κόμβων του γράφου που θεωρείται ότι είναι δίκτυο με το βάρος των ακμών να σημαίνουν τη χωρητικότητα της σύνδεσης. Ως τομή ορίζεται το σύνολο των ακμών που συνδέουν τους κόμβους των δύο μερών. Η χωρητικότητα της τομής προκύπτει επομένως από το βάρος των ακμών, ή αλλιώς τη χωρητικότητα των συνδέσεων. Ένα οποιοδήποτε σημείο της εικόνας είναι συνδεδεμένο είτε με την ‘πηγή’ είτε με τη ‘δεξαμενή’. Ζητούμενο τελικά είναι η διαμέριση του συνόλου των σημείων της εικόνας, ώστε η τομή να έχει την ελάχιστη χωρητικότητα $E(X)$.

Ένα αποτέλεσμα δίδεται στο Σχήμα 4 για μια συνθετική εικόνα με θόρυβο που ακολουθεί κανονική κατανομή σύμφωνα με την υπόθεση που έγινε στην Εξίσωση 4. Δίδεται η κατάταξη μέγιστης πιθανοφάνειας (b), το αποτέλεσμα με τον αλγόριθμο των επαναλαμβανόμενων τοπικά επικρατούσων τιμών (c) και η βέλτιστη κατάταξη με τομή του γράφου (d).

Για την περίπτωση όπου ο αριθμός των κλάσεων είναι μεγαλύτερος του δύο ($K > 2$), έχουν προταθεί αλγόριθμοι που ανάγουν το πρόβλημα σε μια ακολουθία από τομές για διμερείς γράφους και δίδουν λύσεις κοντά στη βέλτιστη σε πολυωνυμικό χρόνο [3], ιδιαίτερα στην υλοποίηση που έχει προταθεί στο [7], όπου η λύση ευρίσκεται με γραμμικό προγραμματισμό. Μια σύγκριση μεγάλου αριθμού αλγορίθμων για την εκτίμηση του πεδίου σε διαφορετικά προβλήματα ανάλυσης εικόνων και υπολογιστικής όρασης δίδεται στο [6].

Ένα αποτέλεσμα δίδεται στο Σχήμα 5, όπου στην εικόνα (b) δίδεται η κατάταξη της εικόνας σε υπερ-εικονοστοιχεία με τον αλγόριθμο του [1]. Η ομαδοποίηση των υπερ-εικονοστοιχείων σε τρεις κλάσεις δίδεται στην εικόνα (d). Από την εικόνα αυτή γίνεται εκτίμηση των τριών κατανομών των χρωματικών συνιστωσών στο σύστημα *Lab* με την απλούστευση να είναι μεταξύ



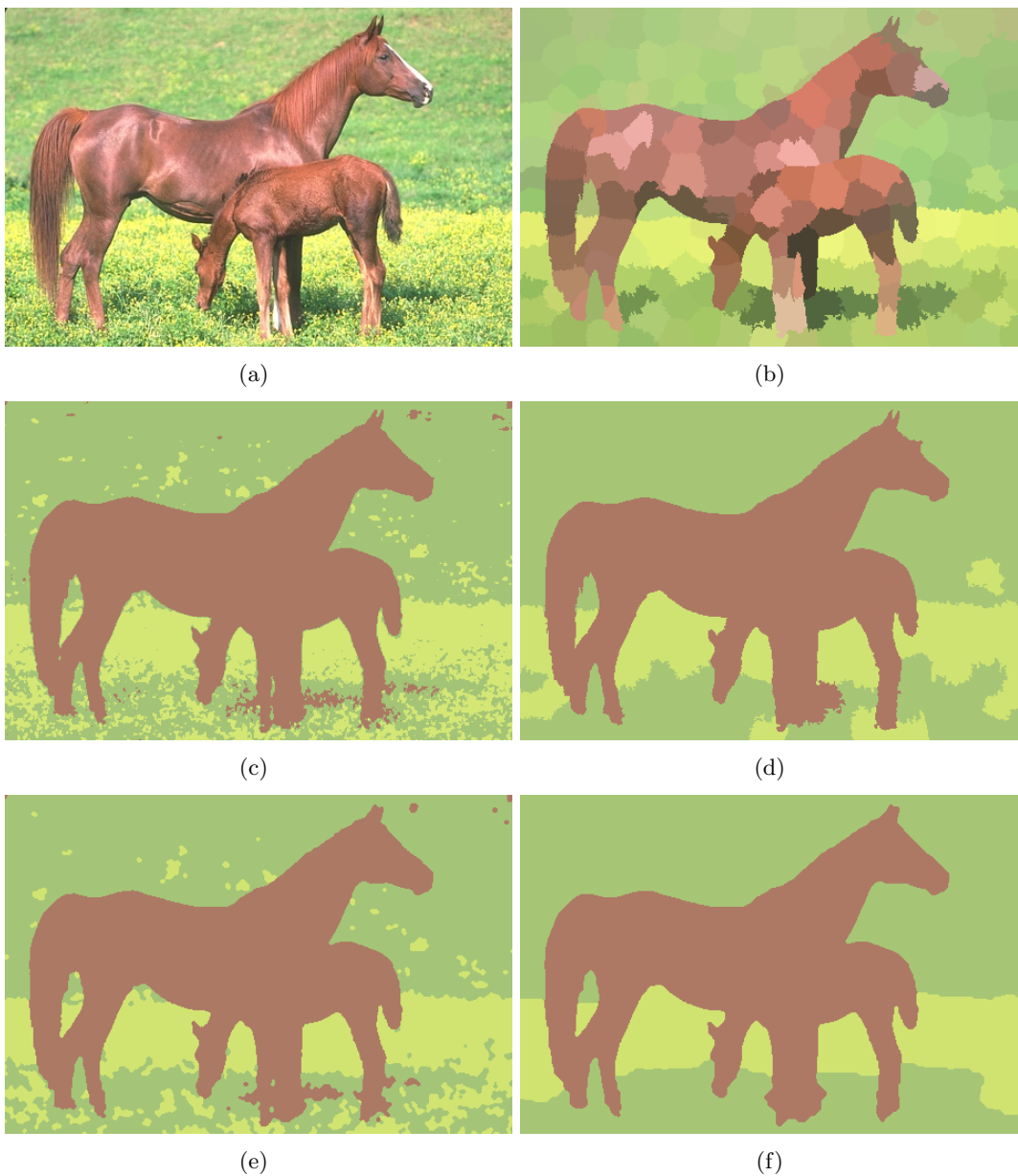
Σχήμα 4: Τμηματοποίηση με τον αλγόριθμο των επαναλαμβανόμενων τοπικά επικρατουσών τιμών (c) και με αλγόριθμο τομών γράφου για δύο κλάσεις.

τους ανεξάρτητες. Η κατάταξη με βάση την αρχή της μέγιστης πιθανοφάνειας δίδεται στην εικόνα (c). Στην εικόνα (e) δίδεται το αποτέλεσμα με τη μέθοδο των επαναλαμβανόμενων τοπικά επικρατουσών τιμών και στην εικόνα (f) δίδεται το αποτέλεσμα με τομές γράφων στην υλοποίηση του [7] για πεδίο με διμελείς παρέες σε συνδέσεις 8 σημείων.

5 Βιβλιογραφία

- [1] R. Achanta, *et al.*, SLIC superpixels compared to state-of-the-art superpixel methods, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 34, No. 11, pp. 2274–2282, 2012.
- [2] J. Besag, On the statistical analysis of dirty pictures (with discussion), *Journal of the Royal Statistical Society, Series B*, Vol. 48, No.3, pp. 259–302, 1986.
- [3] Y. Boykov, O. Veksler and R. Zabih, Fast approximate energy minimization via graph cuts, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 23, No. 11, 1222–1239, 2001.

- [4] S. Geman and D. Geman, Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 6, No. 6, 721–741, 1984.
- [5] D. Greig, B. Porteous and A. Seheult, Exact maximum a posteriori estimation for binary images, *Journal of the Royal Statistical Society, Series B*, Vol. 51, No. 2, pp. 271–279, 1989.
- [6] J. Kappes, *et al.*, A comparative study of modern inference techniques for structured discrete energy minimization problems, *International Journal of Computer Vision*, Vol. 115, No. 2, pp. 155–184, 2015.
- [7] N. Komodakis and G. Tziritas, Approximate labeling via graph cuts based on linear programming, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 29, No. 8, 1436–1453, 2007.
- [8] C. Wang, N. Komodakis and N. Paragios, Markov Random Field modeling, inference and learning in computer vision and image understanding: A survey, *Computer Vision and Image Understanding*, Vol. 117, No. 11, pp. 1610–1627, 2013.



Σχήμα 5: Τμηματοποίηση με κριτήριο τη μέγιστη πιθανοφάνεια (c), με ομαδοποίηση υπερεικονοστοιχείων (d), με τον αλγόριθμο επαναλαμβανόμενων τοπικά επικρατούσων τιμών (e) και με αλγόριθμο τομών γράφου (f) για τρεις κλάσεις με διμελείς παρέες σε συνδέσεις 8 σημείων για την εικόνα (a). Τα υπερεικονοστοιχεία φαίνονται στην εικόνα (b).